70th Annual EAAP Meeting
26/08/19, Ghent

Optimizing the creation of base populations for breeding programs using allelic information

Diego Bersabé[1], Beatriz Villanueva[1], Miguel Ángel Toro[2] and Jesús Fernández[1]

[1]Departamento de Mejora Genética Animal, INIA, Madrid, Spain
[2]Departamento de Producción Animal, ETSI Agrónomos, UPM, Madrid, Spain

bersabe.diego@inia.es

Base populations

- **Starting point of a breeding program**
- **Plenty of genetic variability**
- **Created from different strains**
  - Scarce or no available information
    - Same number of individuals from each strain
  - Phenotypic records and genomic-wide measures of genetic diversity within and between strains
    - Optimal proportion of individuals from each strain (Fernández *et al*., 2014)

Molecular measures of genetic diversity

- **Expected heterozygosity**
  - Most widely used (as in Fernández *et al*., 2014)
  - Correlated with short-term response to selection
- **Allelic diversity**
  - Good indicator of past population bottlenecks
  - Provides information about exclusive genetic variants (private alleles)
  - Correlated with long-term response to selection (selection limits)

Objective

- **To compare the outcomes of using either allelic diversity or expected heterozygosity as the criterion to maximize genetic diversity when optimizing the creation of base populations**

"What to conserve?"

Materials and methods

- Computer simulations
- Three steps:
  1. **Population in mutation-drift equilibrium**
  2. **Creation of strains**
  3. **Foundation of the base population**

Materials and methods

Step 1: Population in mutation-drift equilibrium

- 20 chromosomes
  - 3000 non-markers
  - 600 SNPs

1 Morgan

N = 1000

Random mating

5000 generations

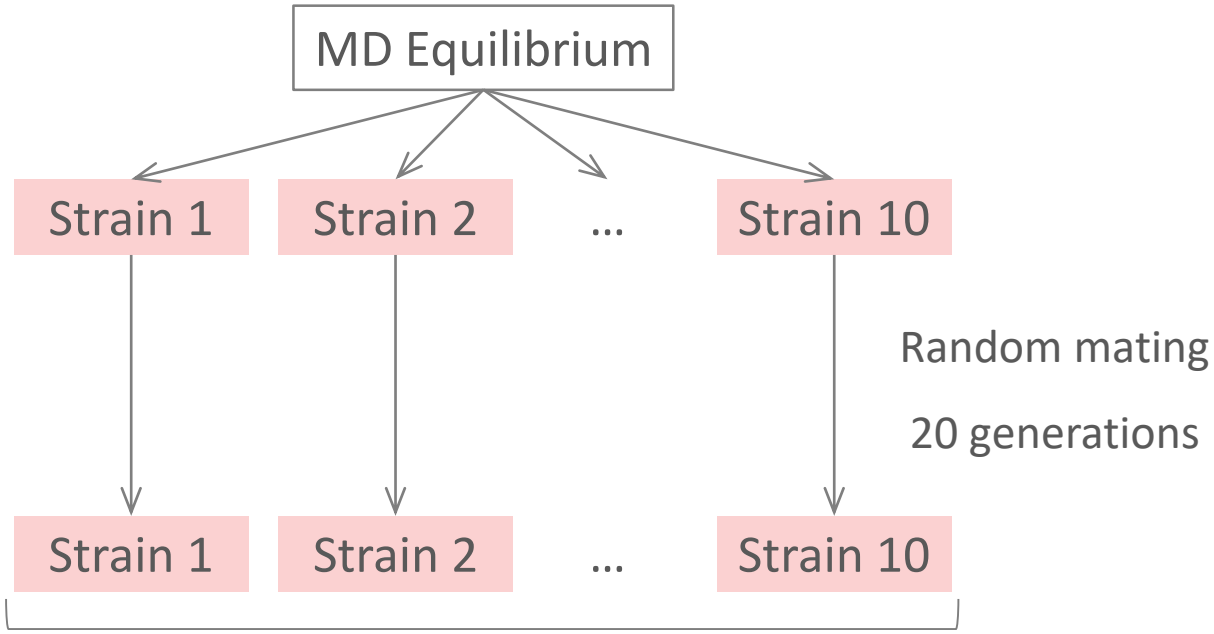
N = 1000

MD Equilibrium

Materials and methods

Step 2: Creation of strains

- Additive trait
  - 1000 total loci
  - $M_P = 100$
  - $V_P = 30$
  - $h^2 = 0.4$
- 100 replicates

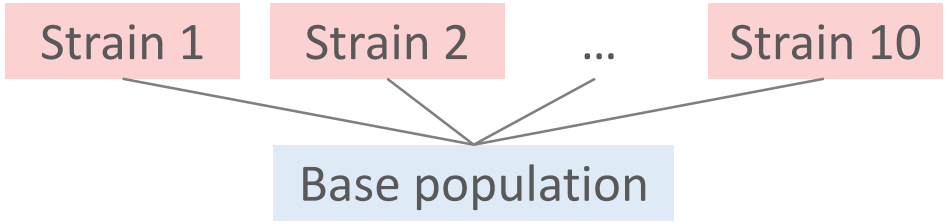


Differences in size:

- N = 10 (x3)
- N = 20 (x3)
- N = 40 (x4)

Step 3: Foundation of the base population

- Select a total of
  100 females and 100 males
  - How?



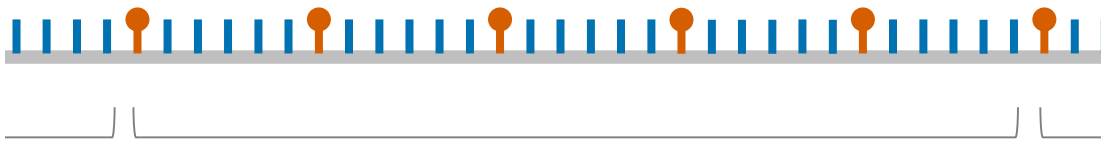
Strategy	Goal
EC	Equal number from each strain
MP-H	Maximize phenotypic value (expected heterozygosity ≥ EC)
MP-A	Maximize phenotypic value (number of alleles ≥ EC)
MH	Maximize expected heterozygosity
MA	Maximize number of alleles

Each strategy in two ways:

1. Individual values and relationships
2. Average strain values and relationships within and between strains

Step 3: Foundation of the base population

- Measuring genetic diversity at SNPs
  - Expected heterozygosity: not a problem
  - Number of alleles: not that meaningful for biallelic markers
- SNP haplotypes
  - Group several SNPs and use them as “alleles”
  - Groups of 5 SNPs



Haplotype of 5 SNPs

Results

Contributions of strains to the base population

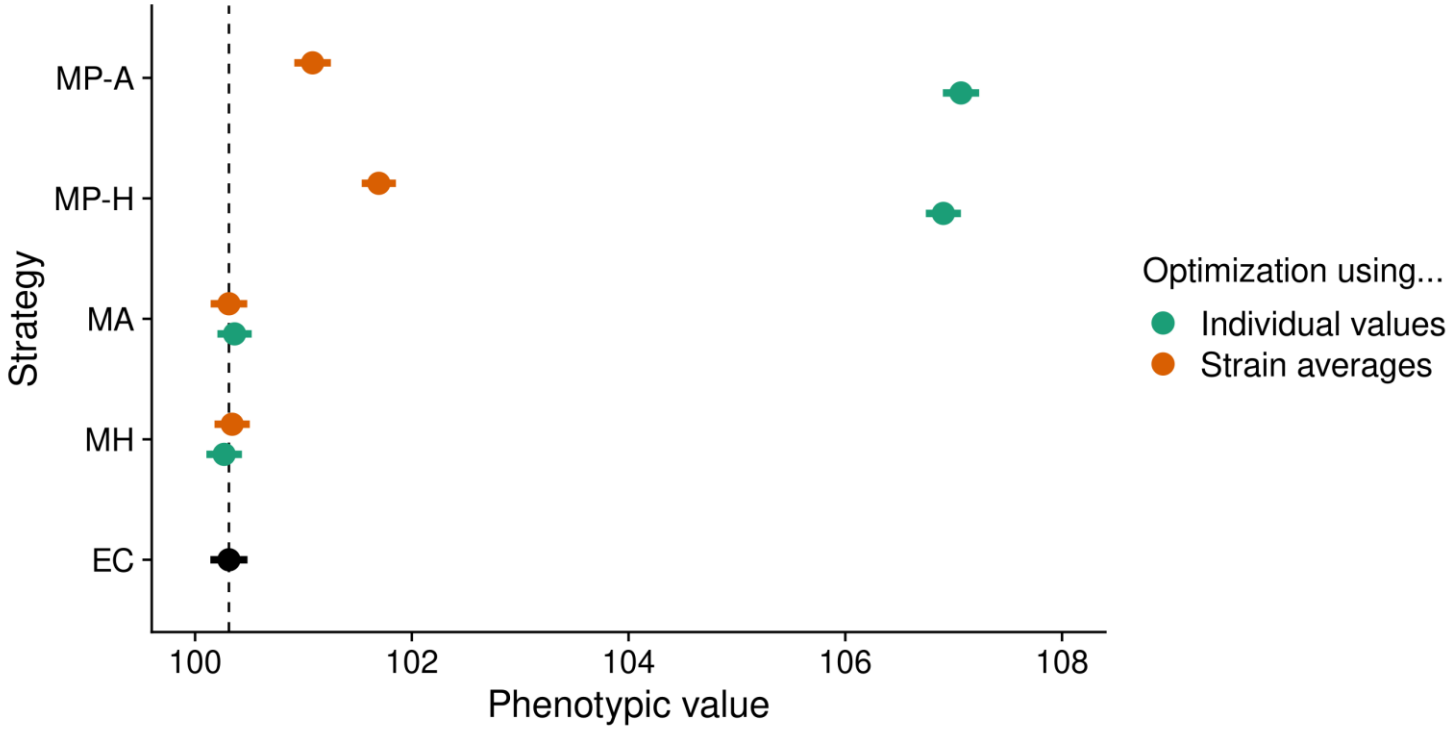
Strategy	*N* = 10		*N* = 20		*N* = 40		Variance	
EC	0.30		0.30		0.40		0.0	
MP-H	0.19	0.16	0.28	0.24	0.53	0.60	9.4	15.5
MP-A	0.22	0.14	0.28	0.24	0.50	0.62	4.5	18.5
MH	0.15	0.17	0.25	0.24	0.60	0.59	17.6	15.3
MA	0.12	0.14	0.24	0.24	0.64	0.62	25.5	21.6

Individual values
Strain averages

- Contributions are proportional to strain size
- MP-H and MP-A detect individuals with high phenotypes in small strains when data from all individuals is available
- MA tends to rely on small strains only to capture exclusive variants

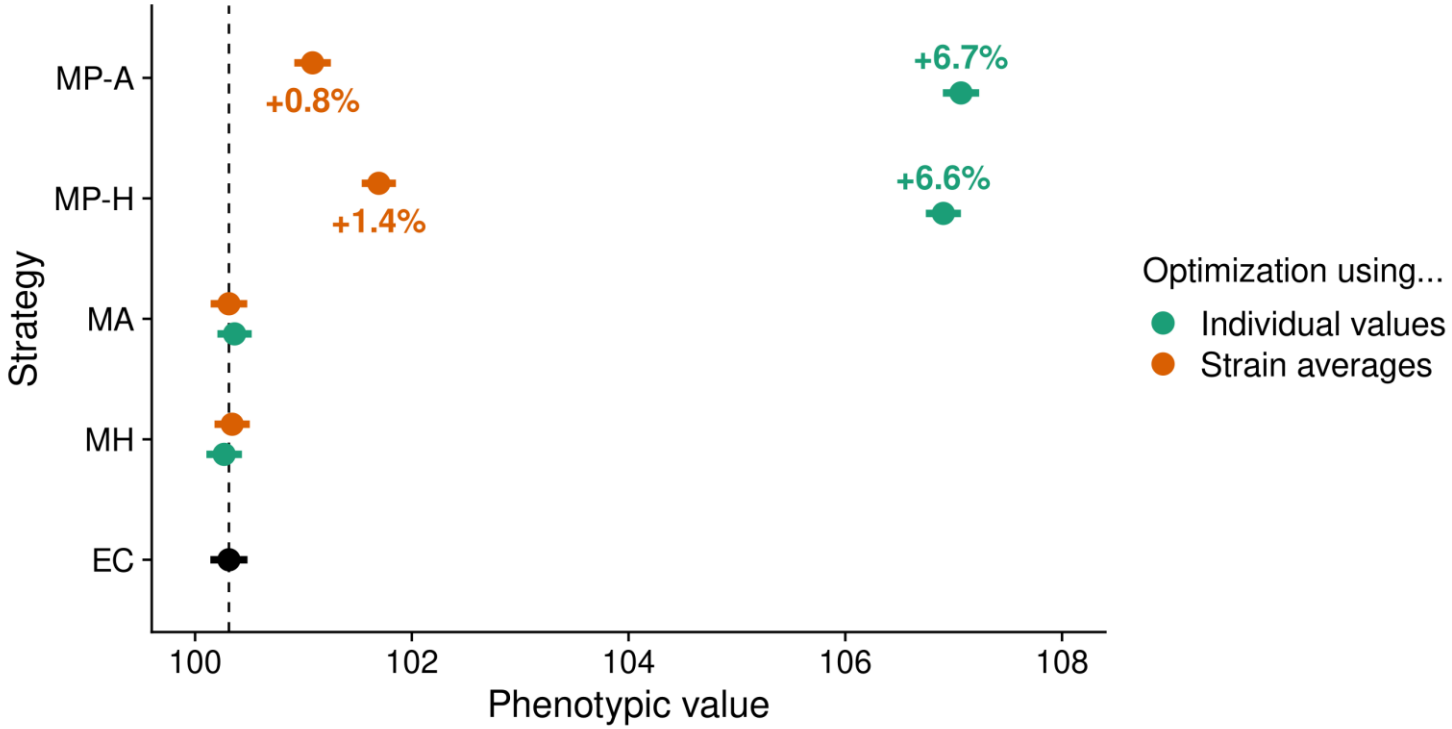
Results

Base population: average phenotypic value



Results

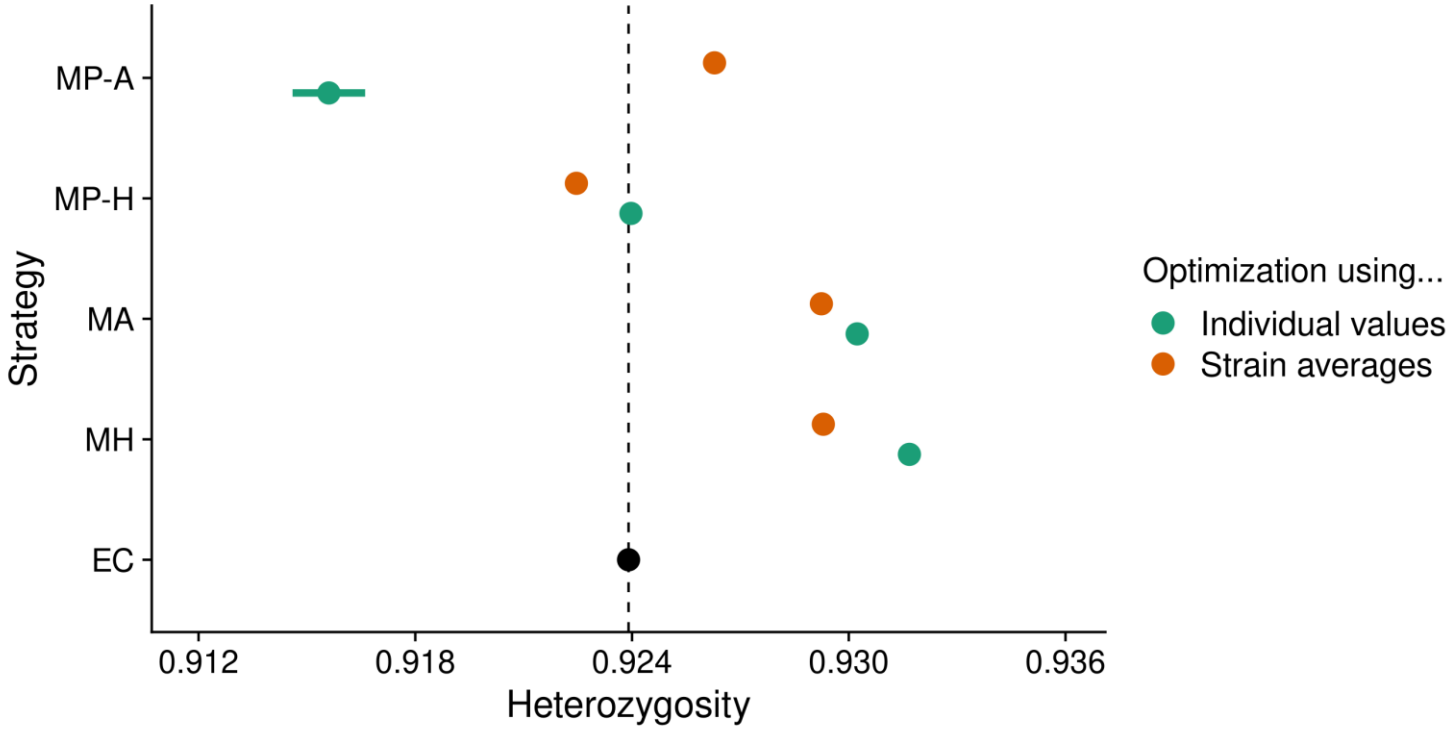
Base population: average phenotypic value



- MP-A and MP-H yield greater phenotypic levels than EC
- Performance is worse when using strain averages, especially for MP-A

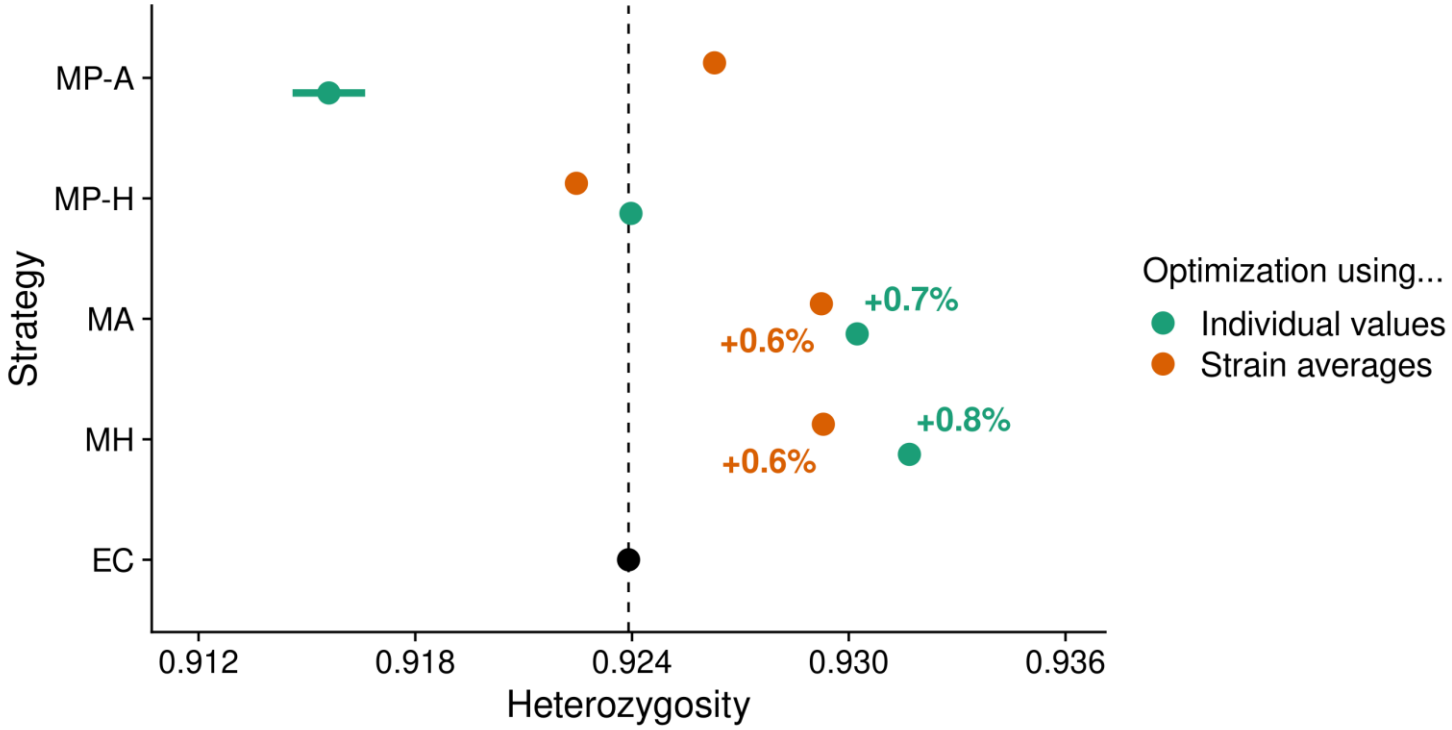
Results

Base population: average expected heterozygosity



Results

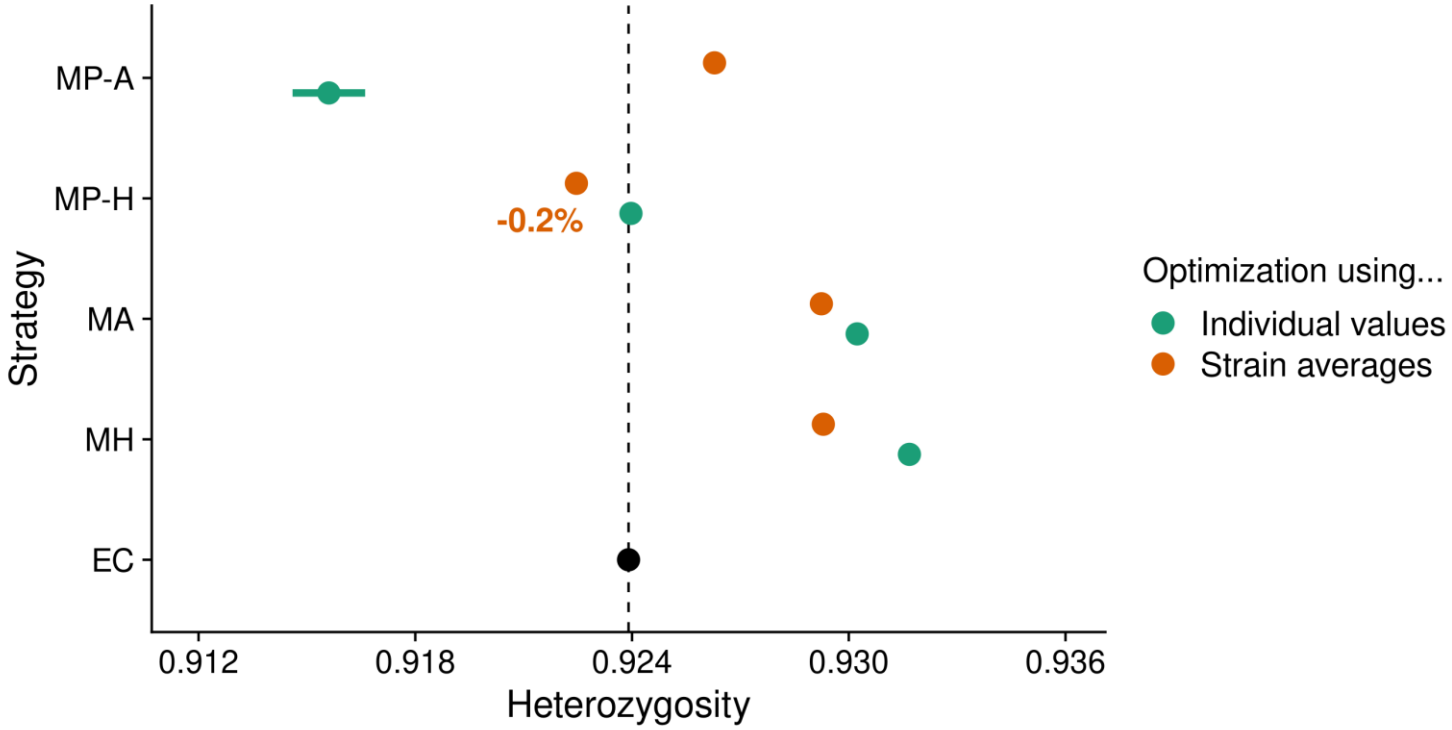
Base population: average expected heterozygosity



- MH and MA capture roughly the same levels of expected heterozygosity
  - Improvements are very small (less than 1%)

Results

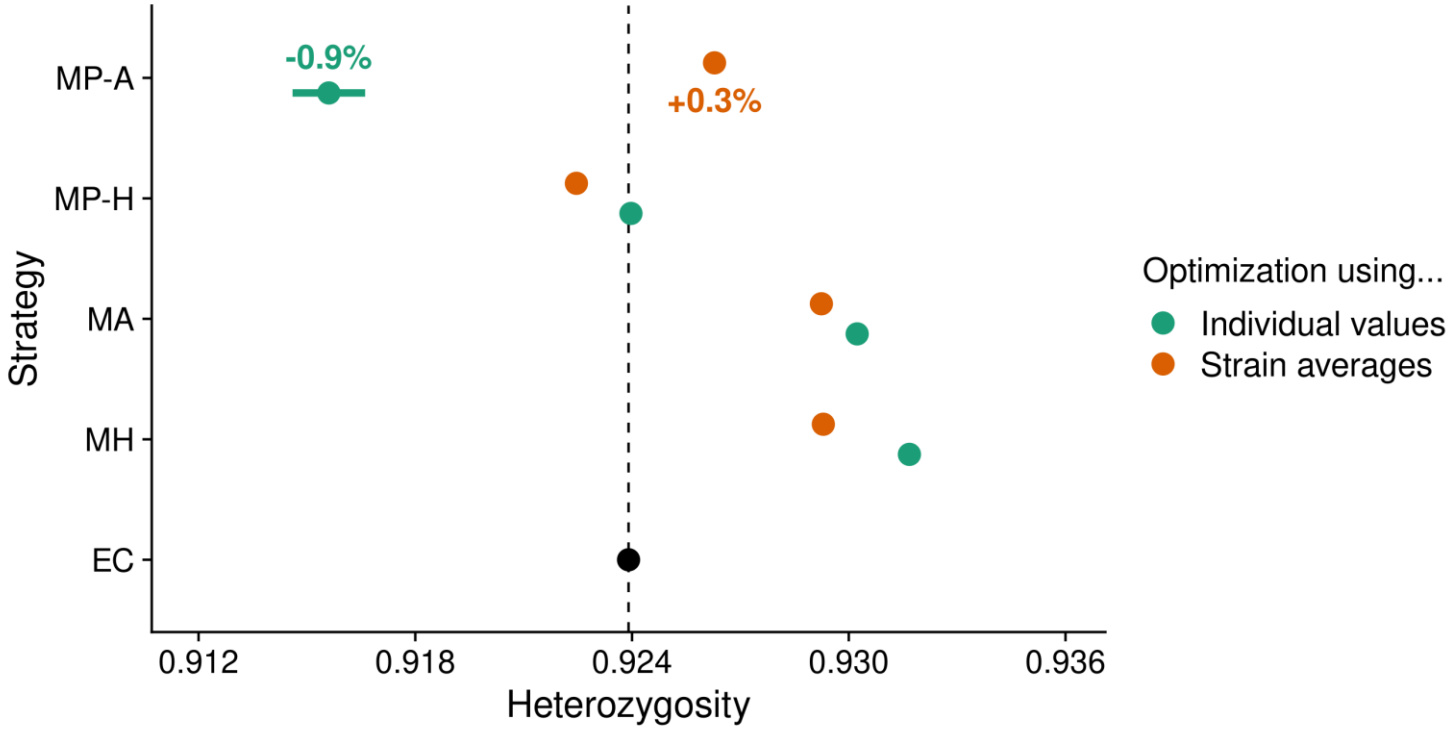
Base population: average expected heterozygosity



- Heterozygosities using MP-H and strain averages may drop below EC levels
  - Probably due to the random sampling of individuals within strains

Results

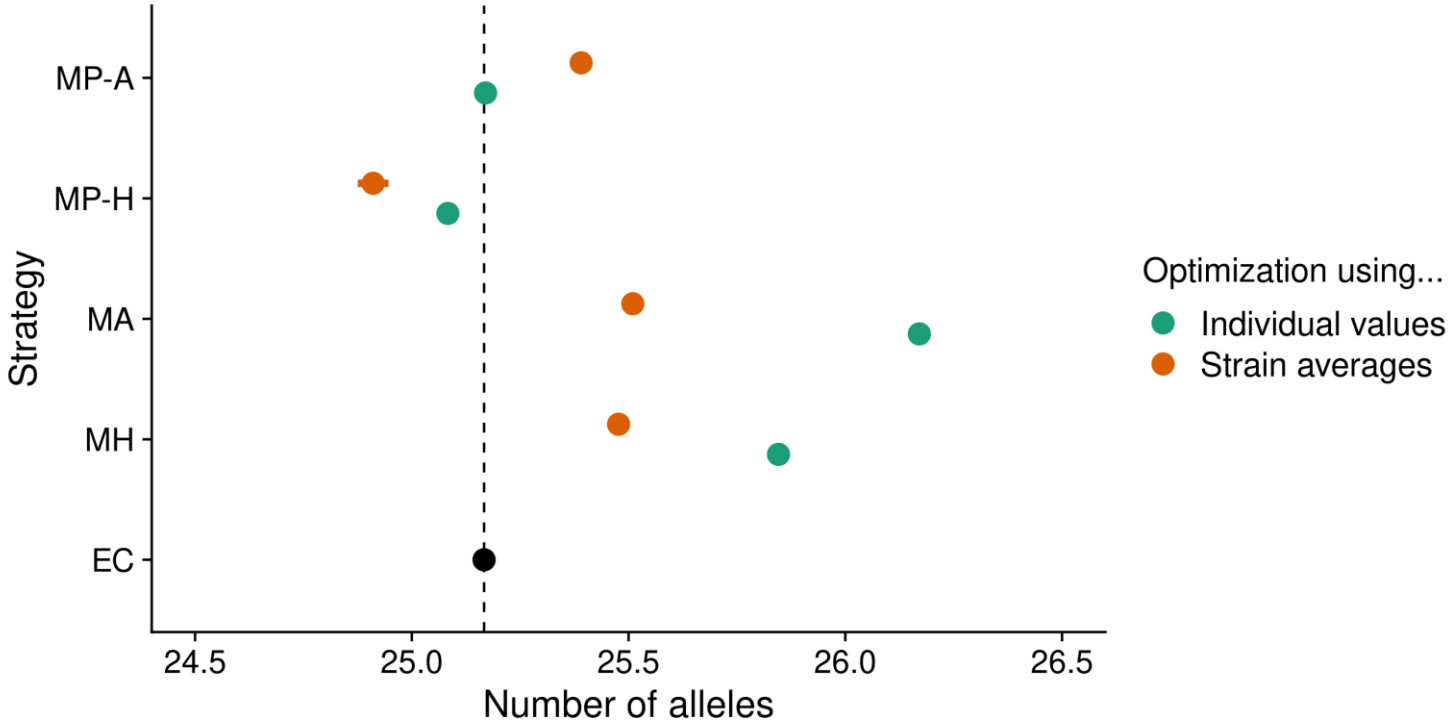
Base population: average expected heterozygosity



- MP-A strategies are not restricted for expected heterozygosity
  - A minimum heterozygosity value is not ensured

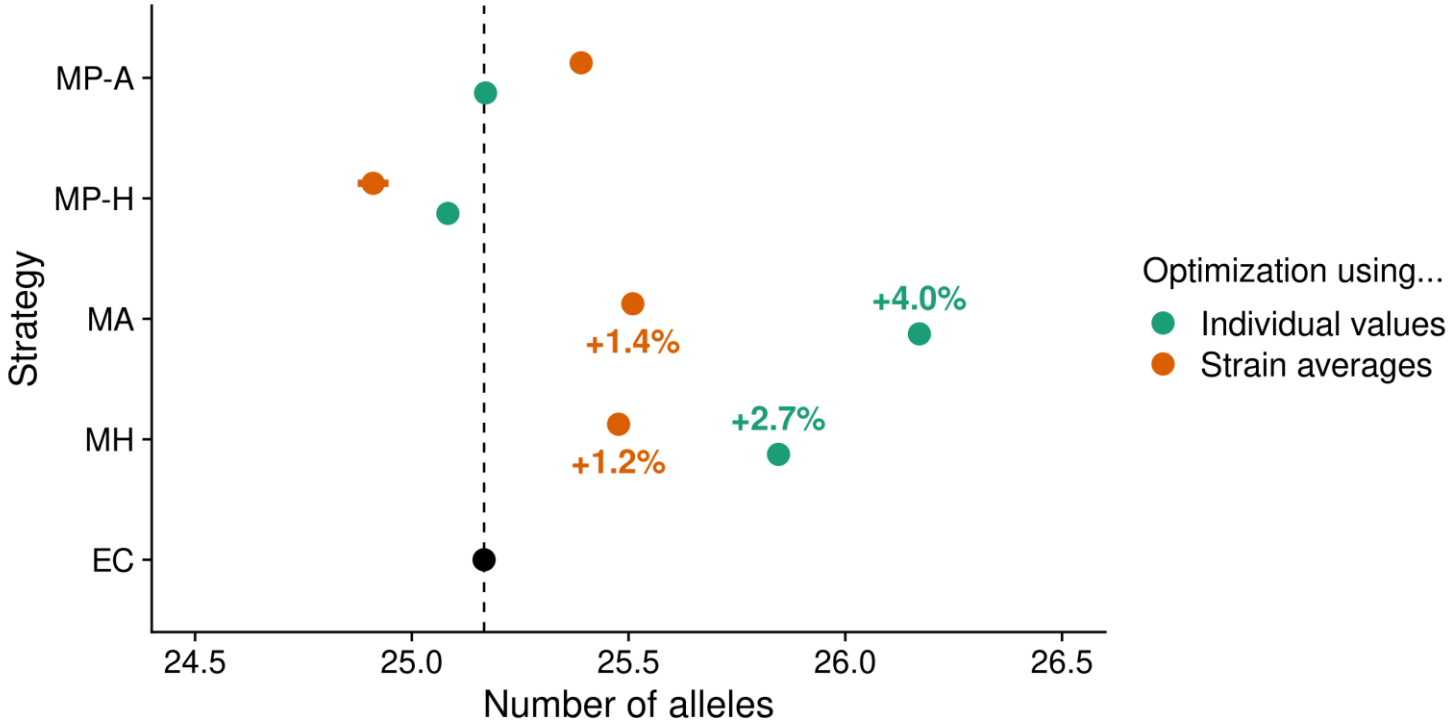
Results

Base population: average number of alleles



Results

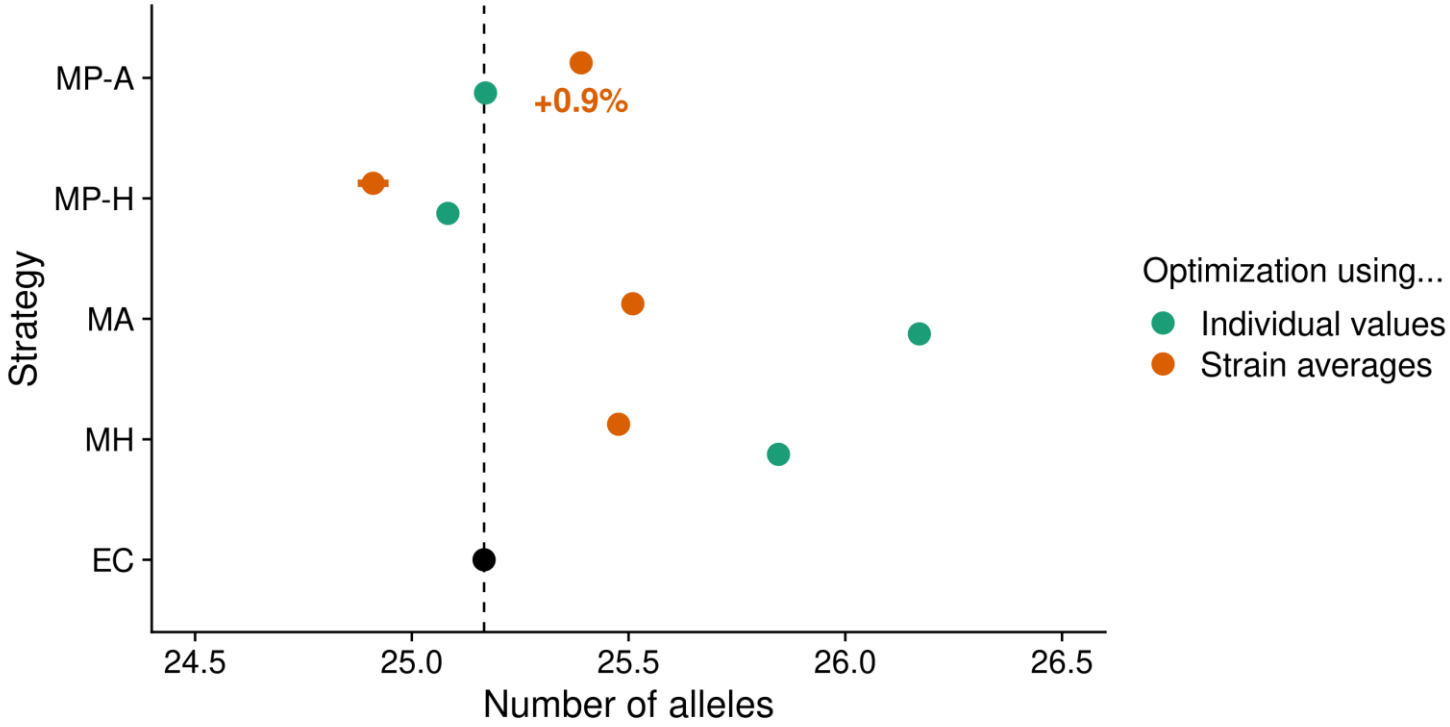
Base population: average number of alleles



- MA produces higher levels of allelic diversity than MH
  - Still, MH captures more allelic diversity than EC

Results

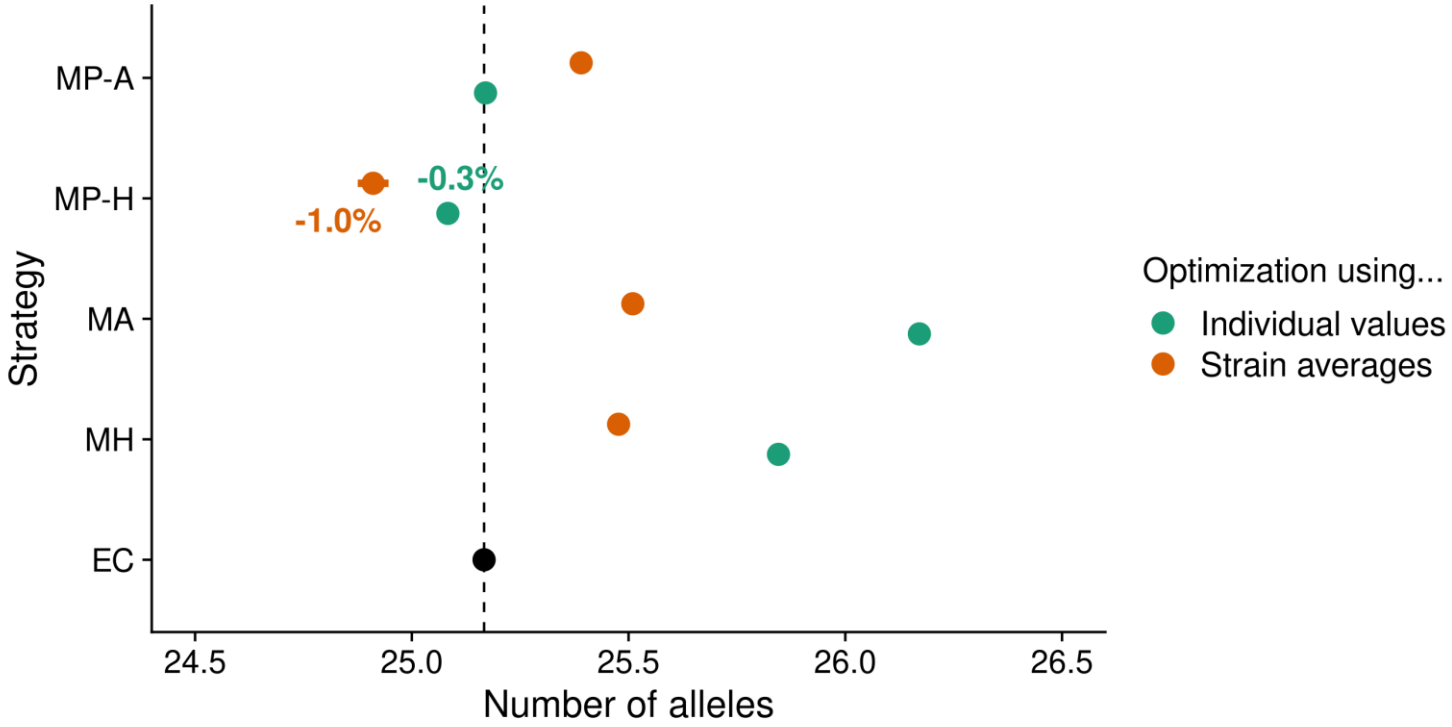
Base population: average number of alleles



- Allelic diversity may go above restriction levels using MP-A and strain averages

Results

Base population: average number of alleles



- As before, MP-H strategies are not restricted for the number of alleles

Conclusions

- **Higher phenotypic values can be achieved at similar levels of genetic diversity, either by using expected heterozygosity or allelic diversity**
  - Smaller increases are produced using strain averages, especially when restriction is on allelic diversity

- **Maximizing either expected heterozygosity or allelic diversity improves both measures of genetic diversity**
  - Slight overall advantage for maximizing allelic diversity

Future

- **Consequences of each strategy in a breeding program**
  - Trait performance and genetic diversity

- **Different simulation scenarios**
  - Preselected (commercial) strains
  - Two or more traits, especially those with negative correlations
  - Dominance

Acknowledgements

Funding from the Spanish Government (grant ref. CGL2016-75904-C2-2-P)

Acknowledgements

Thank you for your attention!