Genome wide diversity:
Variation over a genome

Krista Engelsma

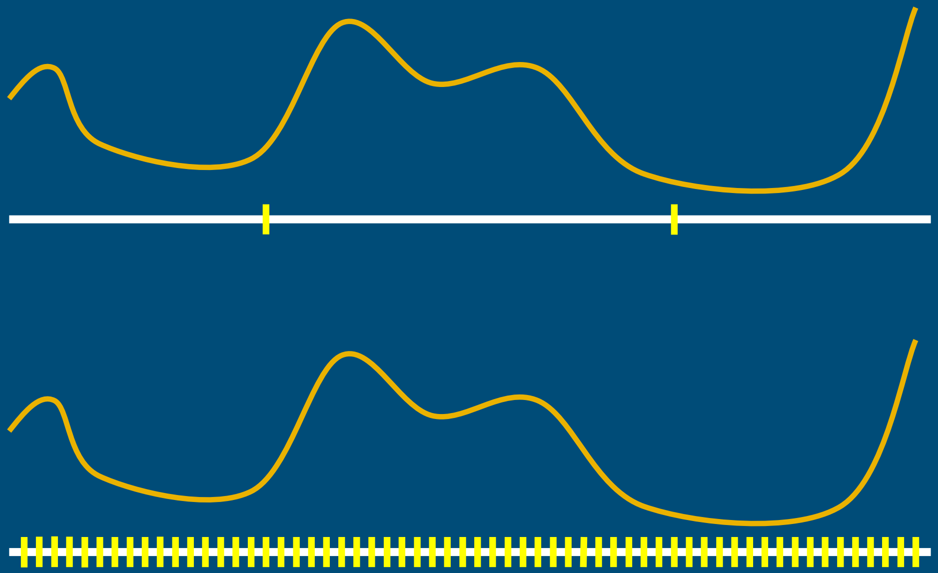
Animal Breeding & Genomics Centre
Centre for Genetic Resources, The Netherlands

Genetic conservation

- Currently based on
  - Pedigree
    - Kinships
    - → genetic diversity
  - Molecular markers
    - Marker variation
    - Microsatellites

More opportunities with dense marker maps

- >10.000 SNPs
- More information
- Patterns across the genome



Microsatellites (30)

SNPs (10,000)

Objectives

- Development method to determine distribution of genetic diversity over a genome with dense marker maps
- Test with simulation data
- Comparison with heterozygosity
- Influence of sampling on the genetic diversity

Genetic diversity – variation

- Genetic diversity = additive genetic variation
- Marker variation → genetic diversity

- Animal breeding
  - Kinships
  - Kinship ↑ variation ↓
  - Genetic diversity = 1 - kinship

Relationships between markers

- Identical by descent (IBD)
  - Relationships markers
  - IBD-matrix per locus
  - Kinships ⟶ IBD
- Set of 10 SNPs = haplotype

	1	2	3	4
1	**1**	0.9	0.9	0.1
2	0.9	**1**	0.8	0.1
3	0.9	0.8	**1**	0.1
4	0.1	0.1	0.1	**1**

IBD-matrix 1 locus

Method

- Relatedness for each locus in the population

  r = c‘IBD c

  r = relatedness

  c = haplotype contributions

  IBD = IBD-matrix

Genetic diversity = 1 – r

Example

- 50 animals = 100 Haplotypes

  Haplotype 1: 60x

  Haplotype 2: 10x

  Haplotype 3: 10x

  Haplotype 4: 20x

$c = [0.6 \quad 0.1 \quad 0.1 \quad 0.2]$

- IBD-matrix

	1	**2**	**3**	**4**
1	**1**	0.9	0.9	0.1
2	0.9	**1**	0.8	0.1
3	0.9	0.8	**1**	0.1
4	0.1	0.1	0.1	**1**

- $r = c'\mathrm{IBD}\,c$

$r = 0.68$

- Current situation in the population
- What will happen when animals are selected?

- Sampling animals for gene bank
  → Change in distribution of alleles

Alternative marker contributions

- Equal
- Optimal

	1	**2**	**3**	**4**
1	**1**	0.9	0.9	0.1
2	0.9	**1**	0.8	0.1
3	0.9	0.8	**1**	0.1
4	0.1	0.1	0.1	**1**

Real c = [0.6 0.1 0.1 0.2]

r = 0.68

Equal c = [0.25 0.25 0.25 0.25]

r = 0.61

Optimal c = [0.00 0.265 0.265 0.47]

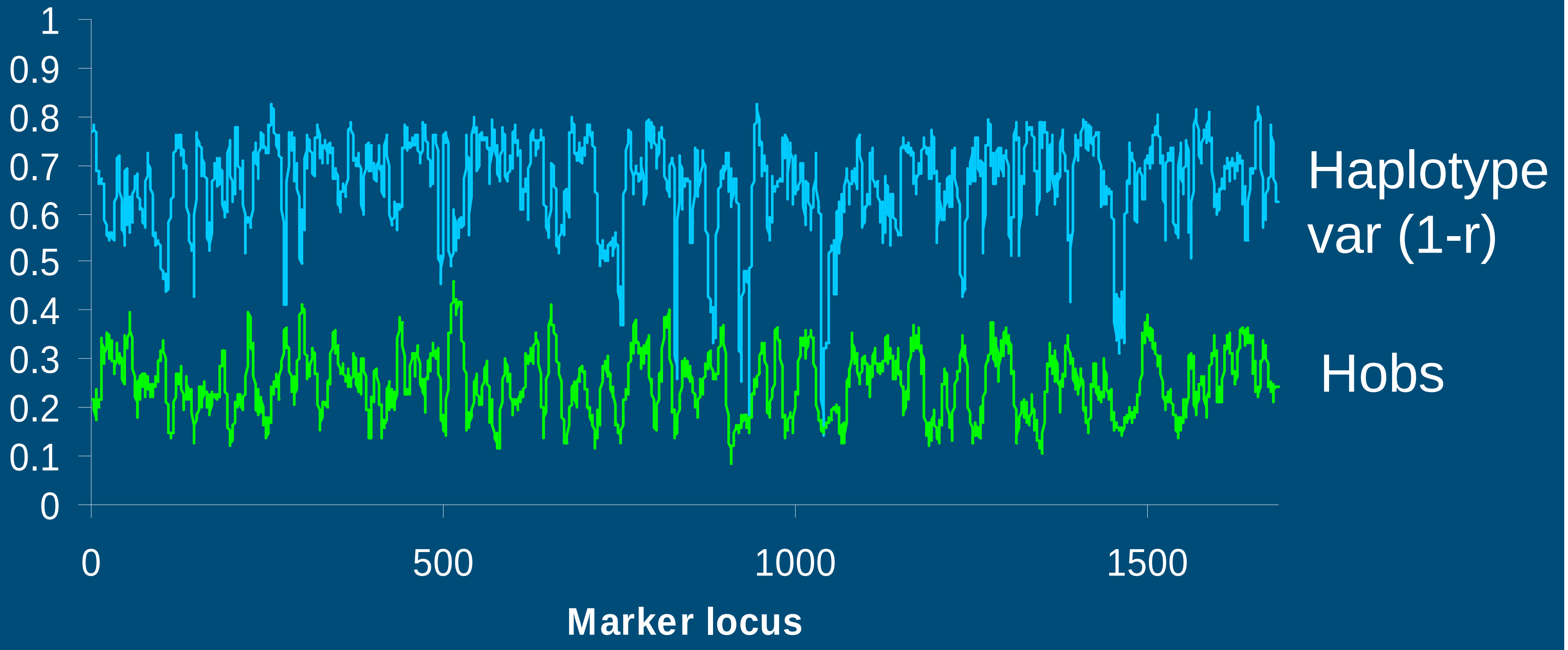
r = 0.52

r = c'IBD c

Results

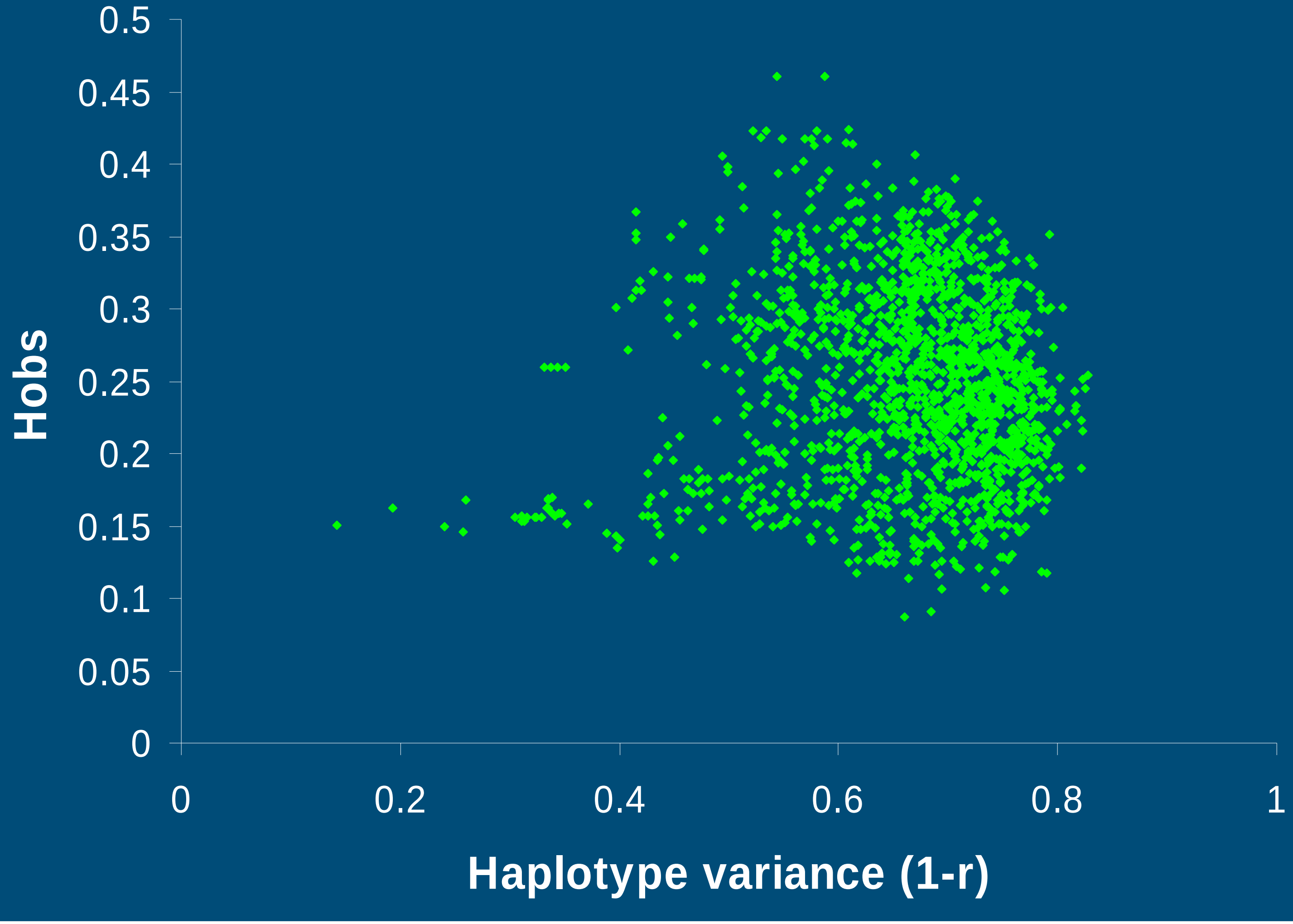
- Data simulated
- 1000 generations
- 200,000 SNP markers, 1,700 SNP with variation

Results

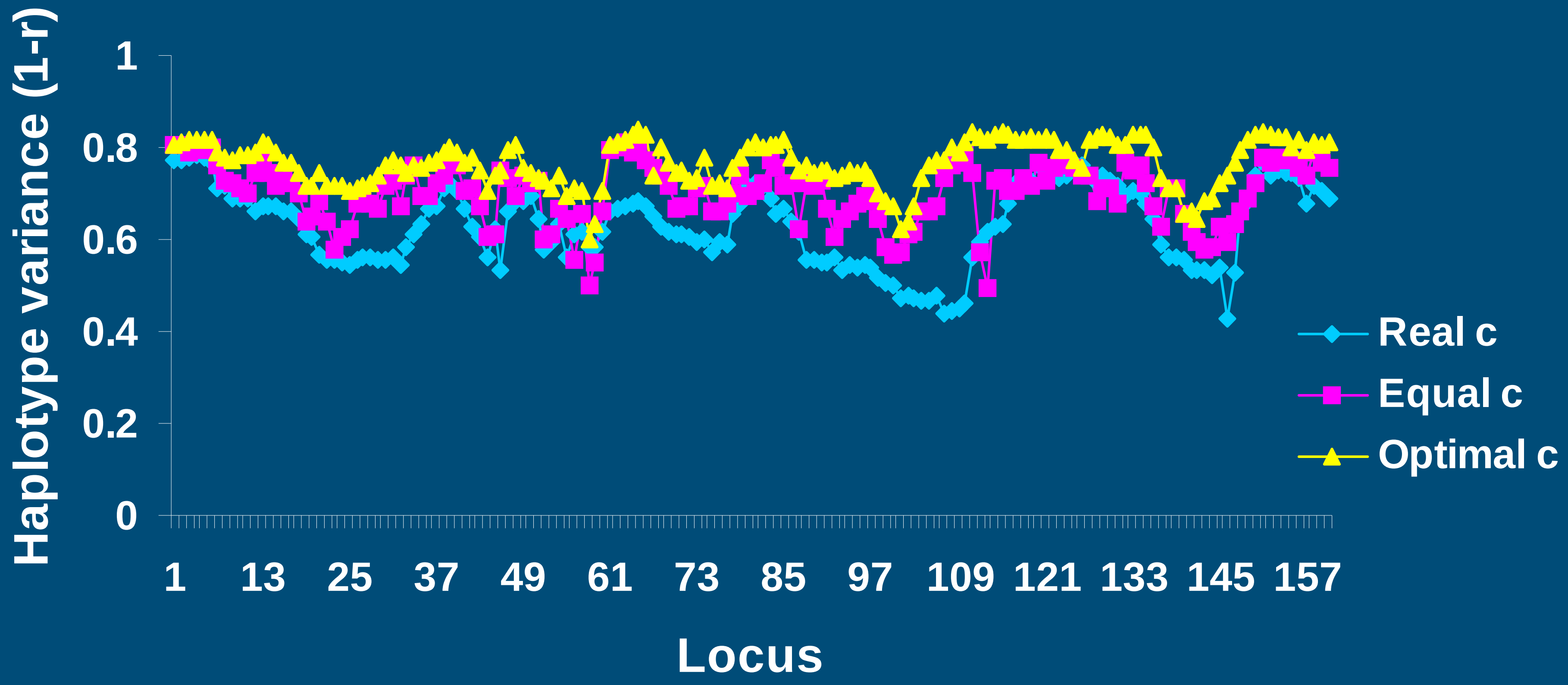


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
500
1000
1500
Marker locus
Haplotype var (1-r)
Hobs

Results


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Hobs
0
0.2
0.4
0.6
0.8
1
Haplotype variance (1-r)

Results


Haplotype variance (1-r)
1
0.8
0.6
0.4
0.2
0
1
13
25
37
49
61
73
85
97
109
121
133
145
157
Locus
Real c
Equal c
Optimal c

Discussion

- Heterozygosity – genetic diversity
  - Low correlation
- Optimal and equal contributions
  - Ideal situation
  - Selection of animals
  - Selection of alleles???

Conclusions

- Variation over the genome
  - Different over loci
  - Can be determined by developed method
- Heterozygosity – genetic diversity
  - Different measures
- Sampling
  - Sampling animals → preference of rare alleles
  - Equal contributions: increases genetic diversity
  - Optimal contributions: highest genetic diversity

Future perspective

- Variation of a specific trait over the genome

Overall GD
Specific trait GD } Conservation breeds in gene bank

Questions?