Use of molecular information in gene banks

Jack Windig, Krista Engelsma

Animal Breeding & Genomics Centre
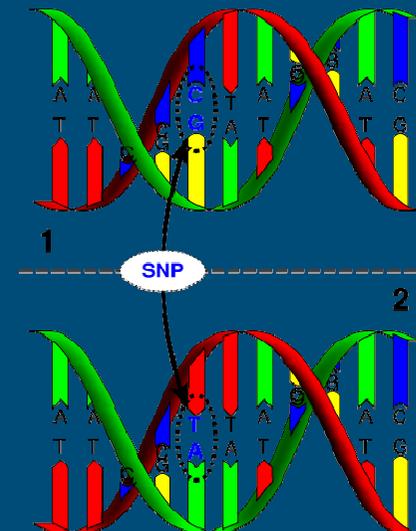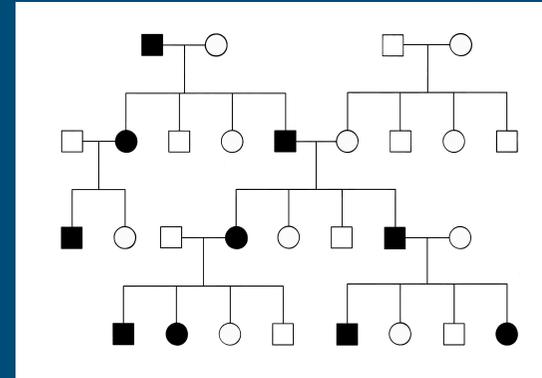

ANIMAL SCIENCES GROUP
WAGENINGEN UR
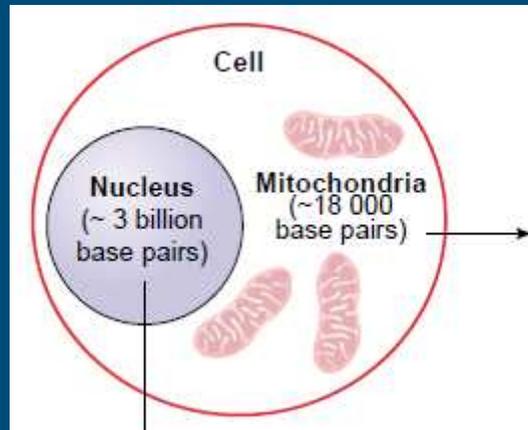
**Animal Breeding &
Genomics Centre**

Genetic diversity in animals

■ **Genetic diversity based on pedigree information**

- Relatedness / inbreeding coefficient
- %DNA Identical by descent
- Overall diversity

■ **Genetic diversity based on markers**

- Detailed information
- DNA same or different
- Diversity at specific genome regions

Markers

Nuclear DNA

- Biparently inherited
- Different chromosomes
- Recombination
  - Provides information on mixture and diversity between and within breeds
  - Provides information on (past) selection events

Mitochondrial DNA

- Maternally inherited
  - Does not pick up use of different sires
- No recombination
- More suitable as molecular clock
  - Provides information on domestication events
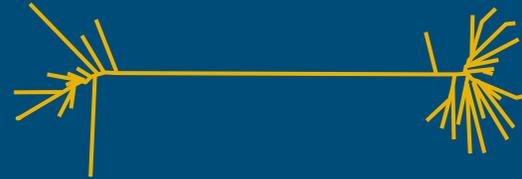 - And mixture afterwards

Mt-DNA and Domestication history

- Two main events for

Cattle

Sheep

Pig

Buffalo

- Multiple events for Horse

ANIMAL SCIENCES GROUP
WAGENINGEN UR

Animal Breeding & Genomics Centre
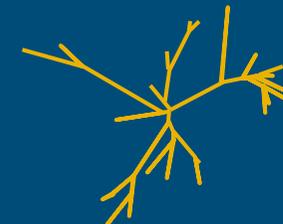
Nuclear DNA-markers

- **Cut**
  - DNA fragments of different length
- **and Paste**
  - Attach to specific sequences
- **Often scored on a (electrophoresis) gel**

- **Examples**

RFLPs	Allozymes
Minisatellites	ISSR
RAPDs	AFLPs
Microsatellites	SNPs
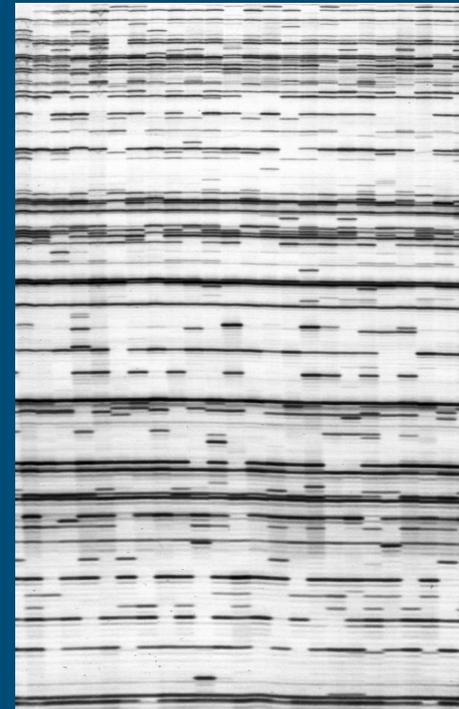
Micro-satellites

- DNA fragments where combinations of 2-5 nucleotides are repeated several times
  - E.g. specific sequence - $(GT)_N$ - specific sequence
- Polymorphic
  - 2-20 alleles per locus
- Standardized sets for livestock species available
  - Different labs -> same results
- Typically 20 – 30 microsatellites per animal

Microsatelittes

Animal	Genotype
31	133 139
32	119 119
33	119 119
34	133 137
35	119 119
36	119 133
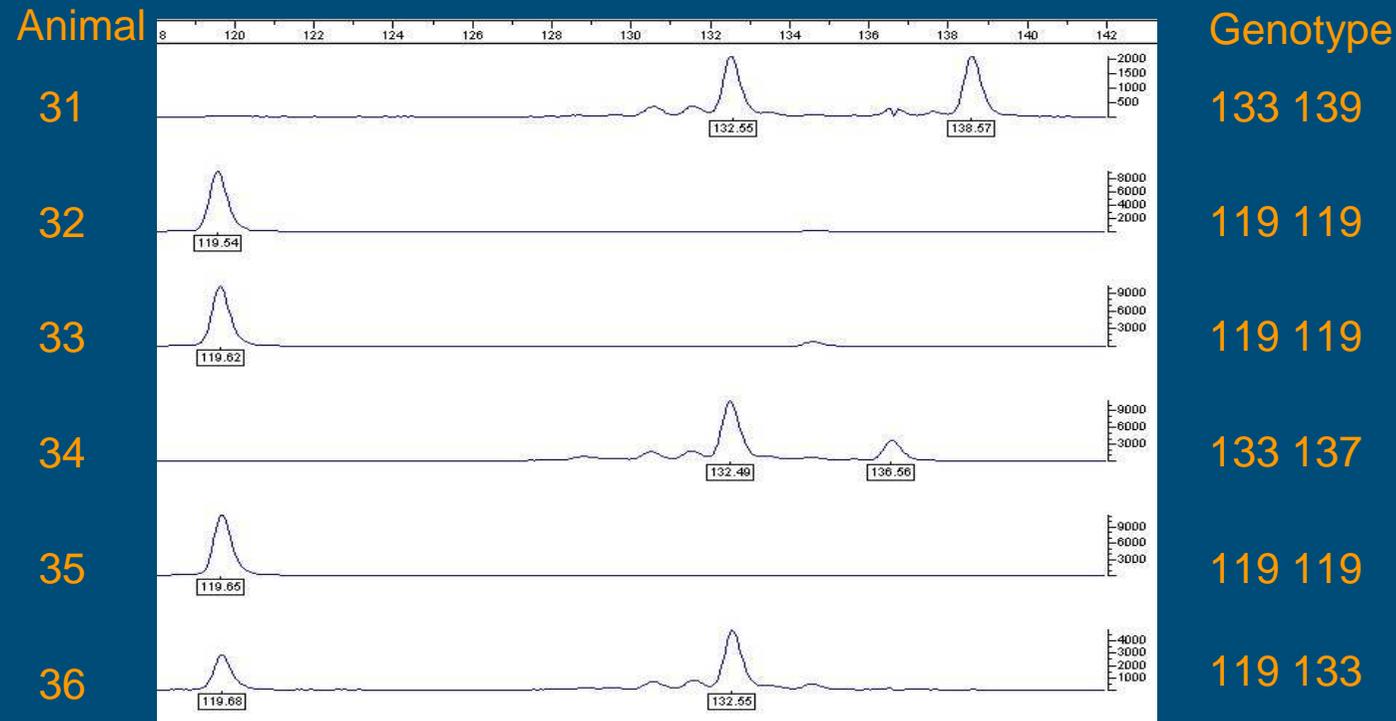
- Scored as number of base pairs
- Sometimes homo- instead of heterozygote
- Sometimes tow base pairs too high or too low

ANIMAL SCIENCES GROUP
WAGENINGEN UR

**Animal Breeding &
Genomics Centre**

Example data file microsatellites

| Regnr.	BM1818a	BM1818b	BM1824a	BM1824b	BM2113a	BM2113b	CSRM60a	CSRM60b	CSSM66a	CSSM66b	ETH10a	ETH10b	ETH225a	ETH225b	ETH3a	ETH3b	HAUT27a	HAUT27b	ILSTS006a	ILSTS006b	INRA23a	INRA23b	SPS115a	SPS115b
NL 2389	262	262	180	182	133	137	98	100	185	197	221	223	140	144	119	127	148	148	296	296	206	214	248	254
NL 2390	262	266	182	182	133	139	100	102	197	197	221	223	144	146	109	119	148	148	288	288	214	216	248	254
NL 2391	262	266	180	182	137	139	98	100	183	185	221	221	144	150	119	127	148	148	288	296	206	216	248	248
NL 2392	262	262	182	182	127	133	98	100	185	185	221	223	146	150	119	119	148	148	288	296	206	216	248	248
NL 2393	262	264	182	182	127	137	102	102	197	197	217	221	144	146	119	127	150	150	288	288	198	198	248	248
NL 2394	262	264	180	182	127	133	98	100	179	197	221	223	144	144	117	127	148	148	292	292	206	214	248	248
NL 2395	262	264	178	188	127	139	100	100	185	197	221	221	144	150	117	127	148	148	292	292	198	206	248	248
NL 2396	262	266	182	182	127	137	98	100	185	197	221	223	144	146	119	119	148	148	288	296	206	216	248	248
NL 2397	262	266	180	180	137	137	98	100	179	185	221	221	144	144	109	119	148	150	296	296	198	198	248	254
NL 2398	262	262	180	182	133	139	100	102	185	197	223	223	144	144	109	119	148	148	288	292	206	214	248	248
NL 2399	264	266	182	182	135	137	100	102	179	185	217	221	140	144	109	127	148	150	288	296	206	214	248	254
NL 2400	262	266	180	180	127	133	100	102	179	197	221	223	144	144	109	119	148	148	288	292	198	206	248	248
NL 2401	262	266	182	182	127	135	98	102	185	197	217	221	144	146	119	127	148	148	288	296	198	208	248	248
NL 2402	262	266	182	182	137	139	98	102	179	185	217	221	144	146	117	127	148	150	288	288	198	216	248	248
NL 9833	262	266	182	182	127	137	98	100	179	197	217	223	144	146	117	127	150	150	288	296	198	214	248	248
NL 9834	266	266	182	182	127	137	102	102	185	197	217	221	146	146	119	119	148	148	288	292	198	208	248	248
NL 9835	262	266	182	182	137	139	98	100	185	197	217	221	144	144	117	117	148	150	292	292	198	214	248	248
NL 9836	262	266	182	182	137	139	100	100	179	197	217	221	144	146	119	119	148	150	292	292	198	216	248	248
NL 9837	264	266	182	182	133	137	98	100	185	185	217	223	144	144	109	127	148	148	288	296	206	206	248	254
NL 9838	262	266	182	182	127	137	100	102	197	197	221	223	144	144	117	119	148	150	296	296	198	198	248	248
NL 9839	262	262	180	182	133	137	98	98	179	179	217	221	144	146	119	127	148	150	292	296	206	216	248	248

ANIMAL SCIENCES GROUP
WAGENINGEN UR

Animal Breeding & Genomics Centre

Single Nucleotide Polymorphism (SNP)

- Variation for a single base pair
  in a DNA sequence
- High genomic abundance
  - 3 000 – 3 000 000 SNPs per animal
  - Standard for cattle now 50 000
  - In practice after data edits 42 500 left
- Platform-independent

**Animal Breeding &
Genomics Centre**
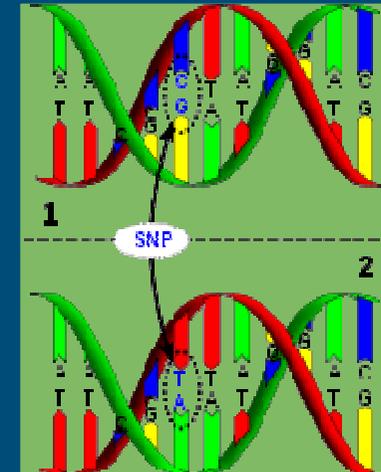
SNP data

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10
animal 1	2	1	1	2	1	2	2	2	2	1
animal 1	2	2	1	2	1	1	2	1	2	1
animal 2	2	1	1	2	1	2	2	2	2	1
animal 2	2	2	2	2	1	1	1	2	2	1
animal 3	2	2	2	2	1	1	1	2	2	1
animal 3	2	1	1	2	2	2	2	2	2	1
animal 4	2	2	1	2	2	2	2	2	1	1
animal 4	2	1	1	2	1	1	2	2	2	1
animal 5	2	1	1	2	1	2	2	2	2	1
animal 5	2	2	1	2	1	2	2	2	1	1

SNP file example

Genetic markers and relatedness

- **Similarity (per locus)**
  - Compare alleles of two individuals
- **Average similarity is estimate of relatedness**
- **Correction needed for alleles alike in state (AIS) but not identical by descent (IBD)**
- **Programs available to estimate relatedness from markers**

ANIMAL SCIENCES GROUP
WAGENINGEN **UR**

Animal Breeding & Genomics Centre

Estimation of r and F from markers

- **If all markers AIS are IBD**
  - $S_{xy} = [I_{11} + I_{12} + I_{21} + I_{22}]$ with $I_{ij} = 1$ or 0
  - S_{xy} is 1, ½ or ¼
  - S_{xy} averaged over multiple loci is estimate of ½r
- **Correction needed for alleles AIS but not IBD**
  - $S_{xy,l} = f_{xy} + P[AIS, not\ IBD]_l + e$
  - Can be solved in linear regression model
  - WEDS estimator

Oliehoek et al. Genetics 173: 483–496 (May 2006)

ANIMAL SCIENCES GROUP
WAGENINGEN UR

**Animal Breeding &
Genomics Centre**

Use for gene banks

- **Verification**
  - Breed
  - Parents

- **Characterization and quantification**
  - Differences within and between breeds
  - Amount of diversity
  - Specific parts of the genome

- **Selection**
  - Specific genes
  - Genome wide

Verification of breed

- **Each breed has unique combination of markers**
  - Breed can be determined for samples of unknown origin
    - Provided unique marker combinations have been determined
  - Hybrids can be detected
  - Identical breeds can be merged
- **Most popular program: Structure**
  - Example: three heath sheep breeds in NL
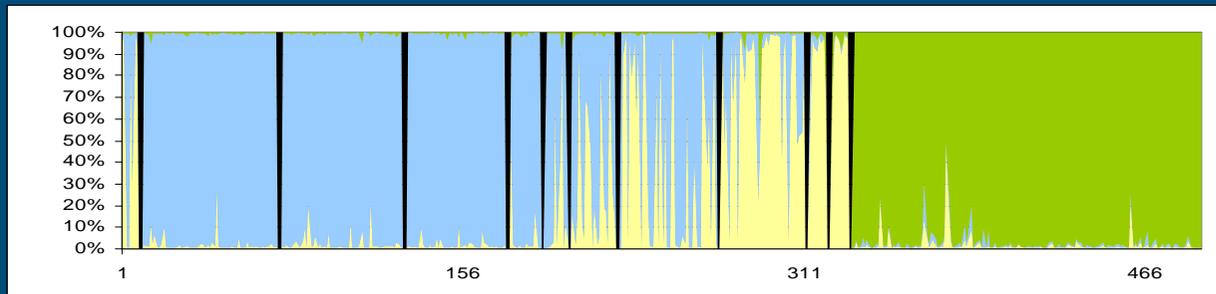
Verification of parentage

- Each offspring has one allel from its mother and the other from its father
- Suspected parent – offspring can be checked
  - Easier if both parent sampled
- Parent can be picked from number of potential parents
  - Easier if other parent sampled
- At least 15-20 microsatellites needed
- Gaps in pedigree may be filled in
  - Zeros in A-matrix can be updated

Characterization of Goat breeds

- ■ Econogene project
  - ● 30 Microsatellite markers
  - ● 45 traditional breeds
  - ● S. and C. Europe
  - ● 1426 goats

ANIMAL SCIENCES GROUP
WAGENINGEN UR

**Animal Breeding &
Genomics Centre**

Clustering
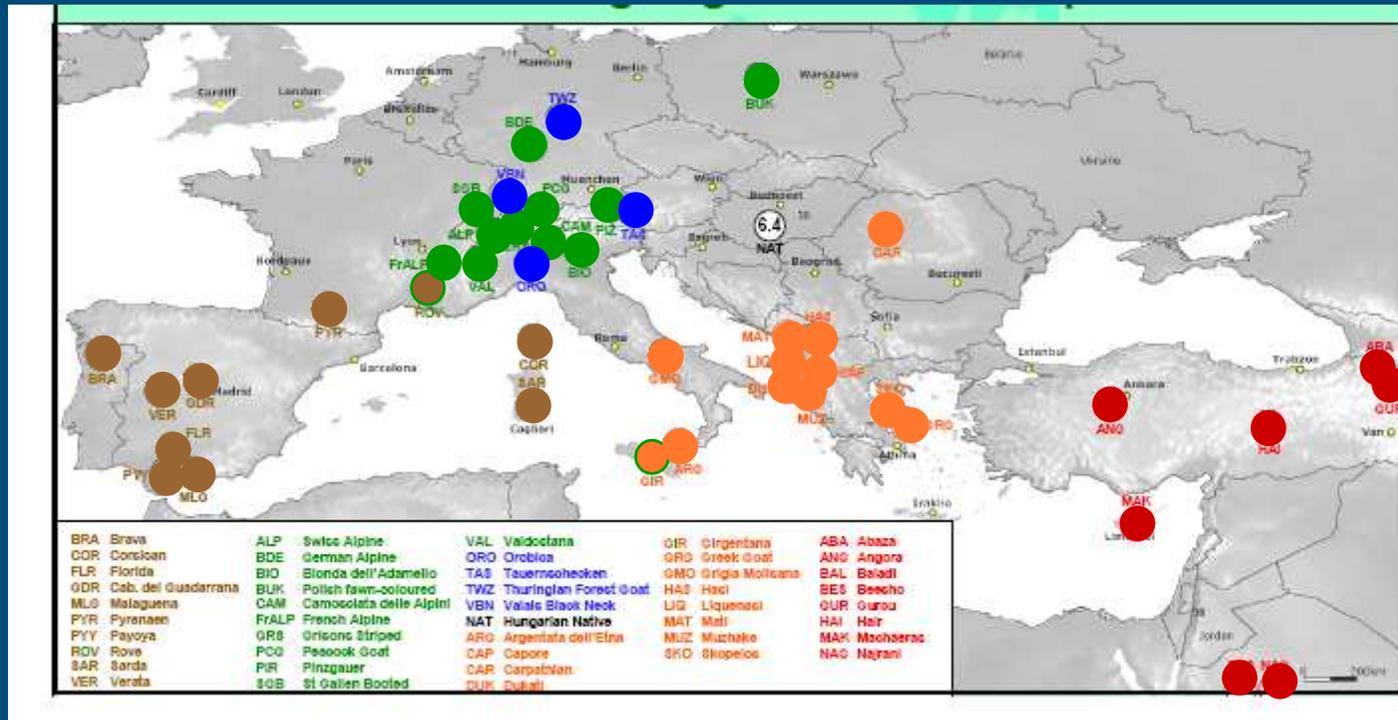
- **Breeds cluster geographically**
- **E – C – W Mediterranean, Alpine**
  - Cattle: same plus British, Scandinavian, NW lowlands, C France

ANIMAL SCIENCES GROUP
WAGENINGEN **UR**

**Animal Breeding &
Genomics Centre**

Number of alleles

Legend:
- 🟢 < 5.5
- 🟢 5.5 – 6
- 🟡 6 – 6.5
- 🟠 6.5 – 7
- 🟠 7 – 7.5
- 🔴 7.5 – 8
- 🔴 > 8
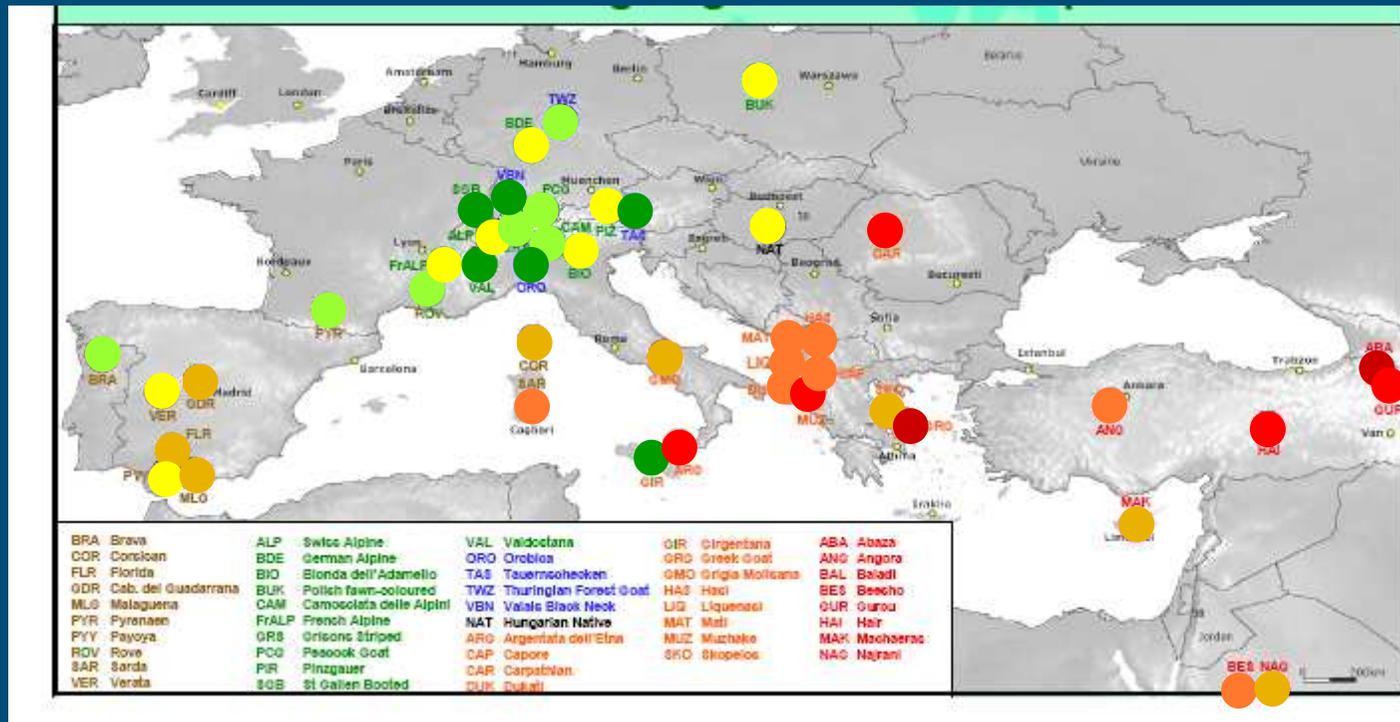
■ **Decreasing from SE to NW**
- Alleles lost with migration or
- Different breeding structure/selection

ANIMAL SCIENCES GROUP
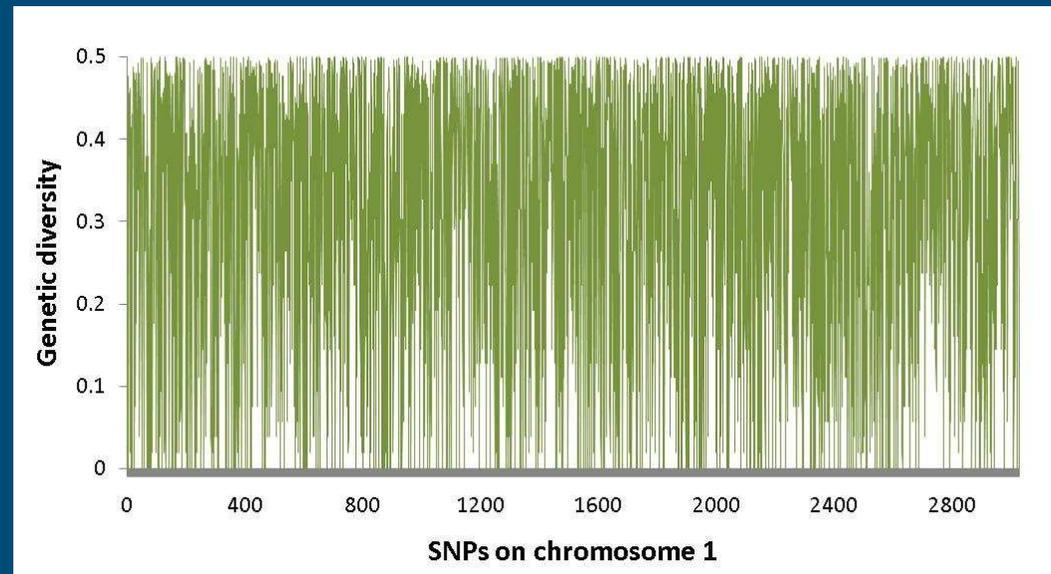**WAGENINGEN UR**

Animal Breeding & Genomics Centre

Marker estimated A-matrix of cattle in the Netherlands
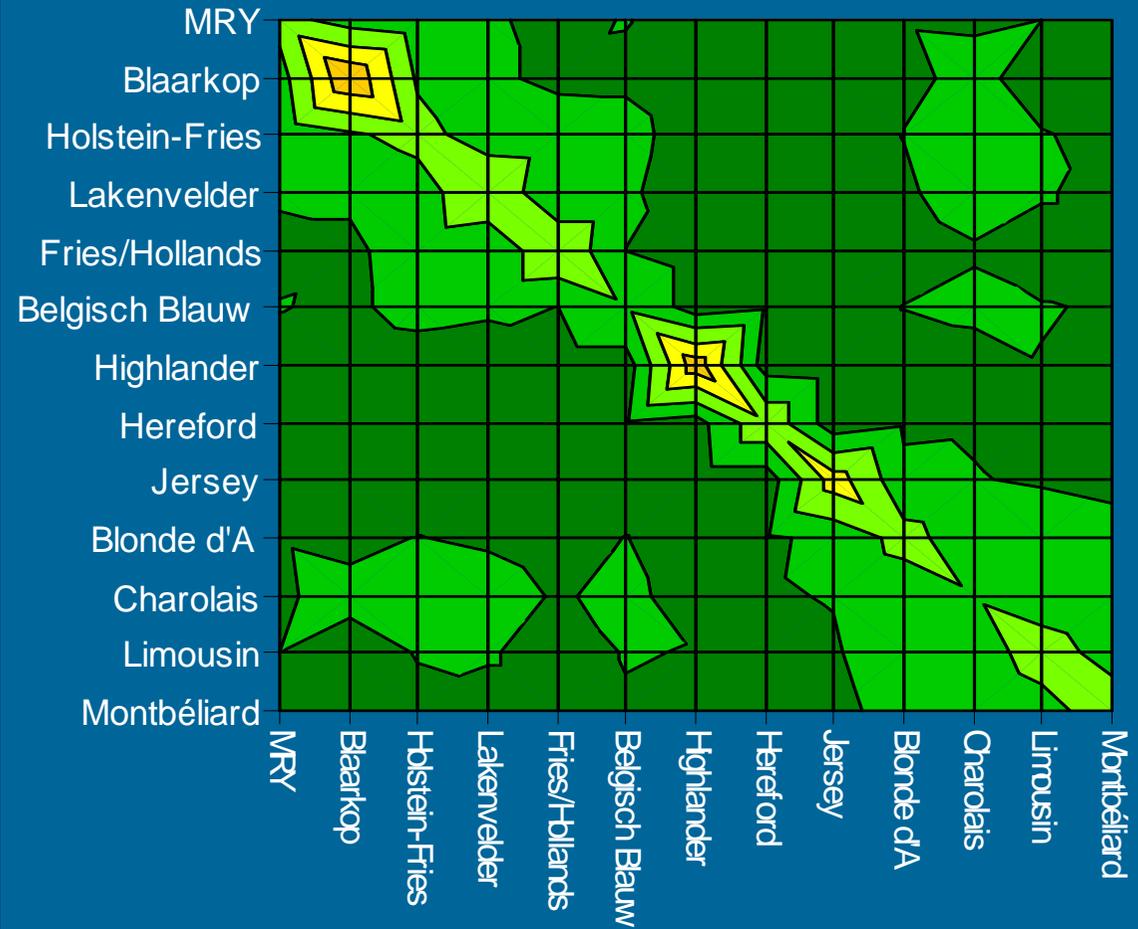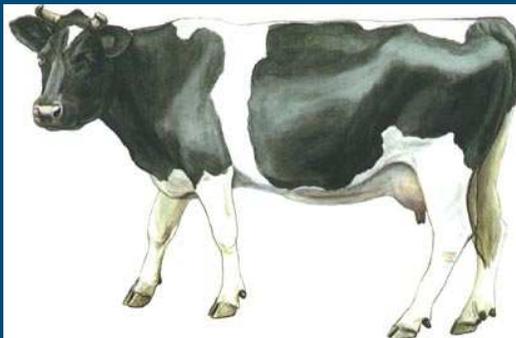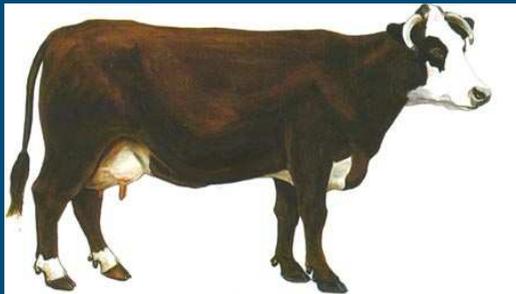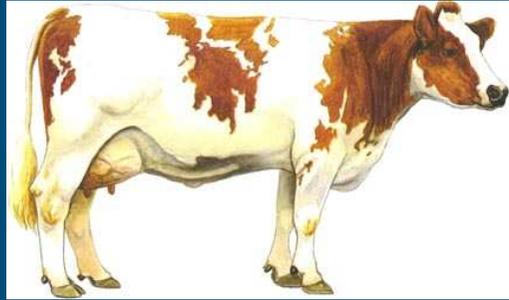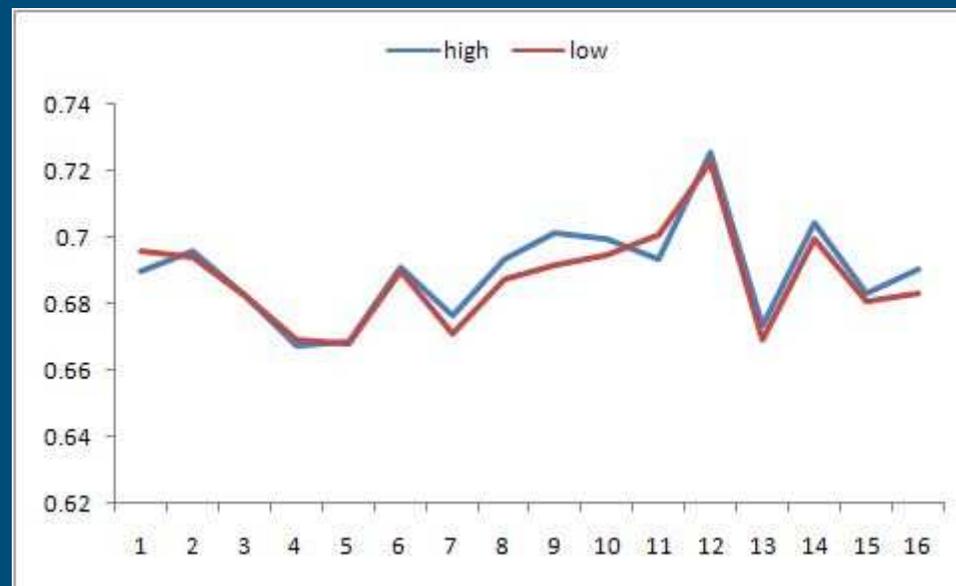
SNPs and genetic diversity

- **SNP variation in cattle**
- **Genetic diversity estimates based on SNP data**
  - Diversity estimated as average similarity

Extremely variable over SNPs!

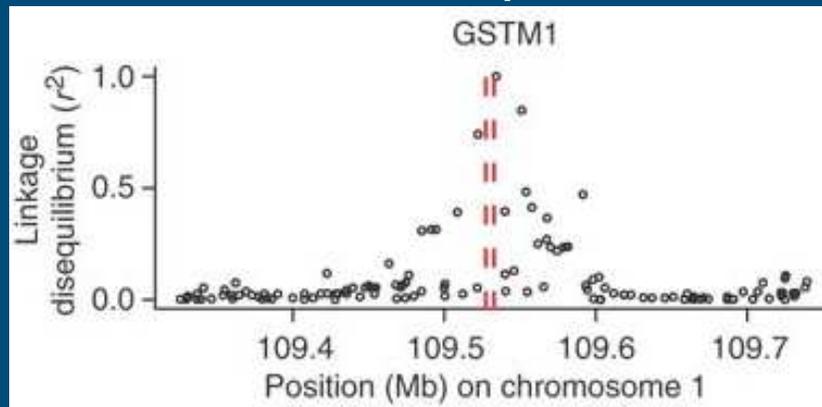# SNP diversity detailed on chromosome segments

- **Two groups of Holstein cattle**
  - Low producers and High producers
- **Genetic diversity per 5Mbp Stretches**
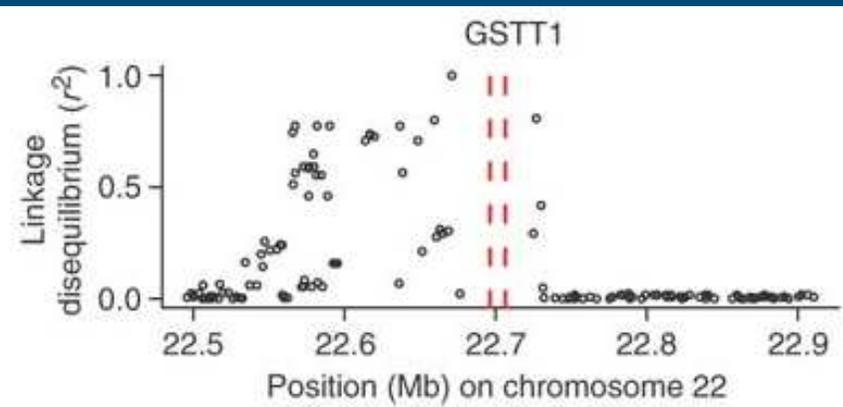 - Average similarity on chromosome 15

Effect of selection

- Signatures of selection → LD patterns over the genome:

Selection in past

Recent selection

Gene deletion in the human genome (McCarroll et al., 2006)

Single genes

- **DNA tests can be developed for traits based on a major gene / single gene**
  - Genetic defects
    - CVM, BLAD in Holstein cattle
  - Disease resistance
    - ARR allele in Prion gene in sheep
  - Morphological traits
    - Polled / Horned in cattle and sheep
    - Colour genes
- **Gene bank may specifically target such genes**
  - Is it worthwhile to conserve "bad" genes?
- **Priority should be with total diversity**

Genomic selection

- **DNA test for traits polygenic traits**
  - Most production traits
- **Breeding values estimated with SNPs**
  - Animals typed for large number of SNPs
  - Reference population with reliable breeding values
  - Association SNP profile with breeding value determined
  - Breeding value can be estimated at birth based on SNPs
  - Selection can be performed much quicker
  - Selection may shift from between family to within families
- **Gene banks**
  - May target genetic variation underlying specific traits
  - Can quantify loss/conservation for specific parts of the genome

**Animal Breeding &
Genomics Centre**