

Cattle genome sequencing and its implications for selection and conservation

Mario Calus
Sipke-Joost Hiemstra
Han Mulder
Mari Smits
Jack Windig
Roel Veerkamp

Animal Breeding & Genomics Centre



Contents of this presentation

- Background of available (genomic) data
- Availability of thousands of genetic markers
- Implications for selection and conservation

Background

- 10 years ago:
 - Pedigree
 - Limited nr. of microsatellites (genetic markers)
- Genetic markers have been used both for selection and conservation purposes
- Recently, Single Nucleotide Polymorphisms (SNPs) have become the most commonly used genetic markers

Background

- Genome sequencing work has speeded up SNP discovery
- Technical developments made high-throughput genotyping possible
- ~50,000 SNP arrays are commercially available for cattle, pigs and poultry for <200 Euro per animal
- Size of SNP arrays will soon be 5 to 10-fold larger
- Wide availability of whole genome sequence info of individual animals is foreseen

Possible use of SNP data

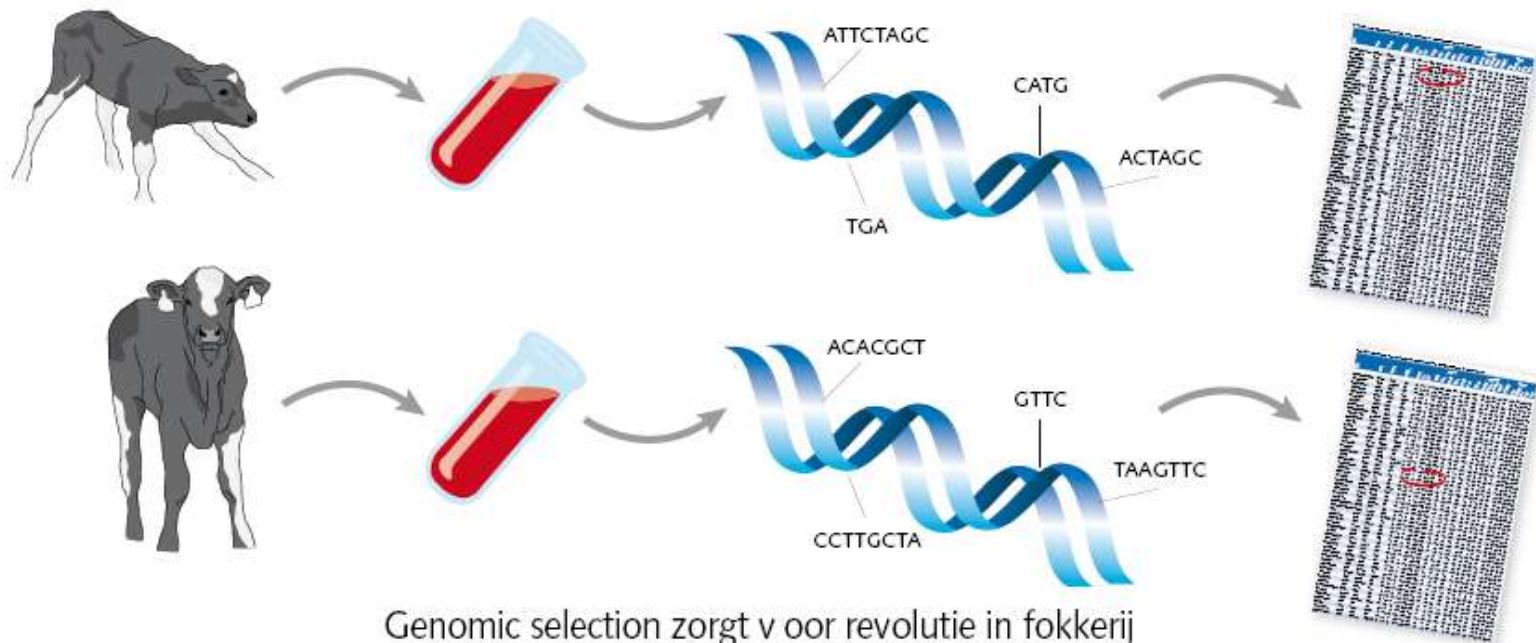
For selection purposes:

- Genome-wide association studies
- Genomic selection

For conservation purposes:

- Estimation of kinships between animals / populations
- Estimation of effective population size (N_e) in recent and more distant past
- Population admixture

Genomic selection: revolution of evolution?



Genomic selection zorgt voor revolutie in fokkerij

Selecteren op DNA-niveau

Tien jaar geleden waren het nog futuristische ideeën, nu is genomische selectie in de fokkerijwereld het gesprek van de dag. Maar wat houdt het precies in? ASG-onderzoekers schetsen de achtergrond van de nieuwe techniek.

Genomische selectie wordt wel bestempeld als een van de meest significante ontwikkelingen in de veefokkerij sinds de invoering van de fokwaardechatting. Het principe van genomische selectie is eind jaren negentig ontwikkeld bij ASG in Lelystad. Toen waren het nog futuristische ideeën, die niet mogelijk waren door te hoge kosten voor het bepalen van de markers. Het afgelopen jaar is

de doortraak gekomen: 50.000 markers die de meeste genetische variatie kunnen verklaren. Het bepalen van die markers kan nu voor hetzelfde geld als waarvoor eerder honderd markers konden worden bepaald, circa 200 euro. Nederland loopt voorop bij de introductie van genomische selectie. ASG is nauw betrokken bij deze ontwikkeling en de implementatie en werkt samen met de onderzoeksafdelingen van CEV en de pluimvee- en varkensfokker Hendrix Genetics. De betrokken ASG-onderzoekers geven hier de achtergrond, schetsen de potentiële impact voor de melkveefokkerij en kijken naar mogelijke toepassingen op het individuele melkveebedrijf.

Topstier pas laat herkend
Een stierkalf krijgt de helft van zijn vaders genen

en de helft van zijn moeders genen. Niet elke nakomeling krijgt dezelfde helft van zijn ouders. Een belangrijke vraag is welke specifieke genen van beide ouders een nakomeling verkrijgt. In de melkveefokkerij is daarvoor het preestier-wachtstier-fokkersysteem ingevoerd. Hierbij wordt de gemiddelde prestatie van ongeveer twaalf nakomelingen van een stier gebruikt om te bepalen of deze stier de gewenste of de ongewenste genen van zijn ouders heeft geërfd. De totale waarde van de verkregen genen wordt uitgedrukt in fokwaarden. Het grote voordeel van het gebruiken van gegevens van dochters is dat hiermee vrij nauwkeurig de fokwaarde van een stier te bepalen is. Het grootste nadeel is echter dat het ongeveer vijf jaar duurt voordat een stier zijn eerste fokwaarde krijgt (zie figuur 1). Het duurt dus ook lang voordat een topstier als

zodanig wordt herkend en veehouders zo'n stier in grote mate benutten.

Van 30 naar 50.000 markers
In de afgelopen tien jaar is veel onderzoek gedaan naar het gebruik van genetische markers. Het doel hiervan was om op jonge leeftijd al te kunnen bepalen of een kalf bepaalde gewenste genen heeft geërfd van zijn of haar ouders. Zodoende is een strengere selectie van kalveren mogelijk, waardoor uiteindelijk een groter deel van de preestieren tot fokstier promoveert en een groter deel van de geselecteerde koeien stiermoeder wordt. Omdat het bepalen van markers altijd duur was en er weinig markers beschikbaar waren, is er veel onderzoek gedaan naar het vinden van de beste genetische markers. Veel markers werden getoet, maar



Genomic selection

- Selection based on many (currently ~50,000) SNPs
- Potential to speed up genetic progress
- Allows selection when recorded number of phenotypes is limited:
 - Difficult to record traits
 - Small populations
- Reliable breeding value early in life
 - Distinguish between related animals at young age (Daetwyler et al., 2007) (ΔF ↓)
 - Potential to decrease generation interval (ΔF ↑)
- Larger SNP-chips (> 300k SNPs) may allow across breed genomic selection => enhance perspectives of smaller breeds

Conservation purposes

- Identification of bottlenecks due to domestication, selection, and breed differentiation
- Unravel genetic structure of cattle breeds
- Identify signatures of selection
- Characterize diversity on different genomic regions (PhD Krista Engelsma)

How many animals should be SNP-genotyped?

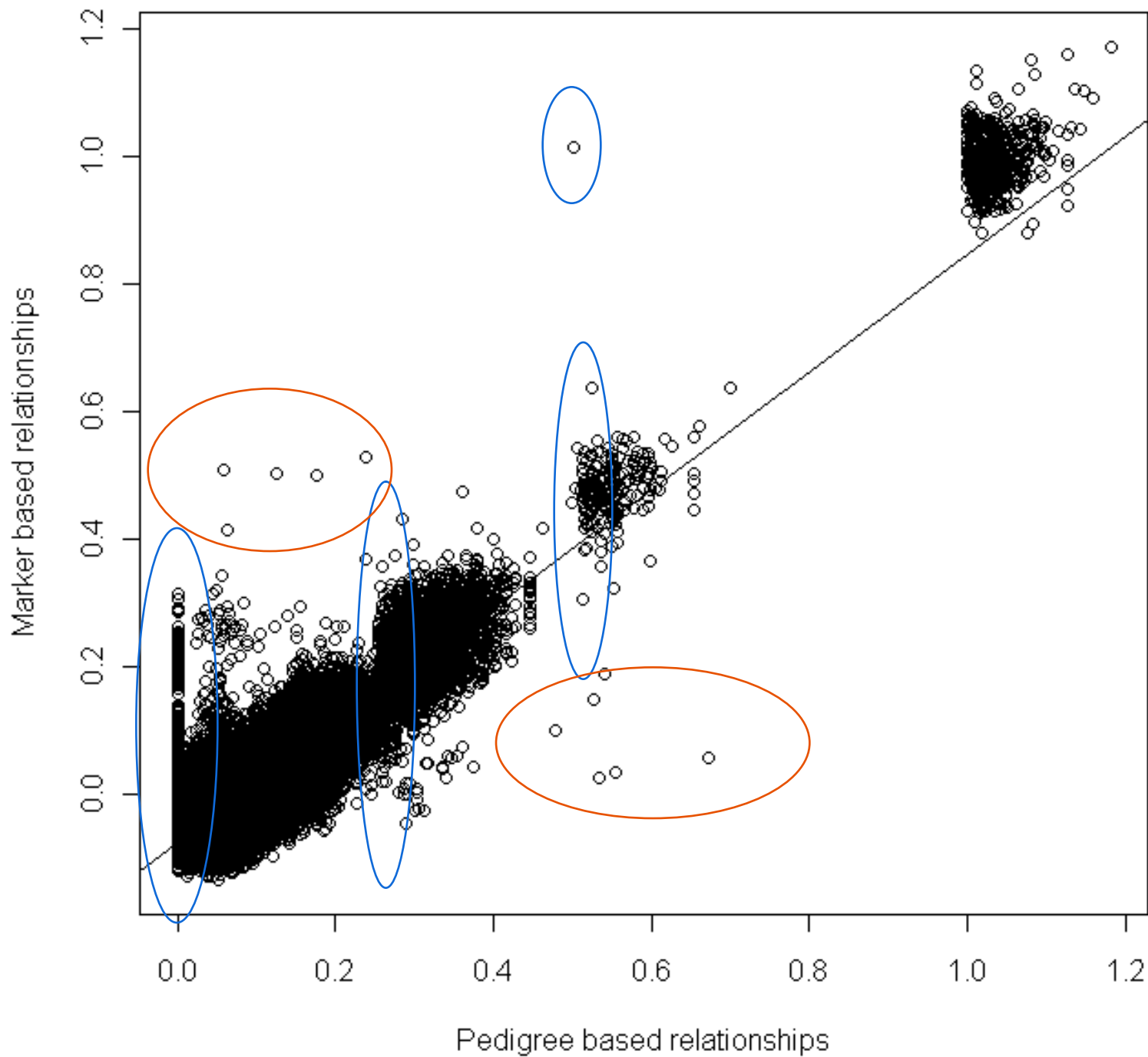
- For genomic selection: 1000+ (with reliable known phenotypic performance)
 - For conservation / analyzing population history, 25+ genotyped animals per breed may be sufficient
- => SNPs obtained for genomic selection can be used (when available) for conservation purposes

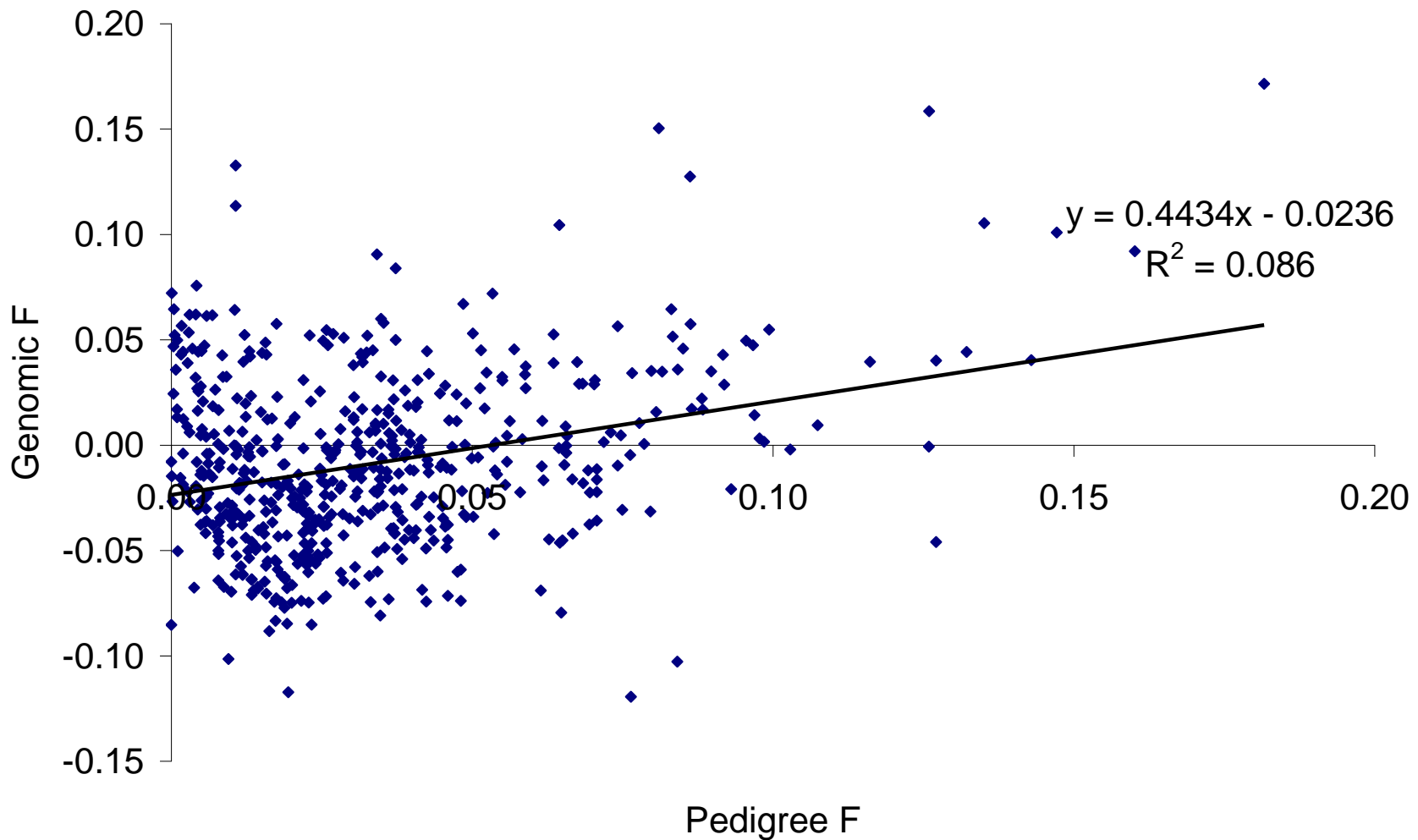
How many SNPs are required?

- For genomic selection: In cattle, 50k SNPs sufficient within a breed; ~300k required across breeds (De Roos et al., 2008)
- For conservation purposes (e.g. characterization of populations): ???

SNP versus pedigree information

- SNPs avoids the need to have pedigree links between two individuals (or populations)
- SNP information allows more precise estimation of kinships (important to maintain variation in selection programs)
- SNPs allow estimation of inbreeding without the need to specify an arbitrary base
- Pedigree and SNP information can be used to check against each other





=> How should genomic F be calculated?

=> Which one is going to be the method of choice?

Does availability of SNPs solve all problems?

- Genetic variation may result from different genomic factors (epigenetics, copy number variation, ...) that are poorly explained by SNPs
- Commercial SNP chips are optimized to be used across breeds:
 - Optimized for SNP variation that is present across breeds
 - Do not (easily) pick up rare (new) variants

Take home messages

- Abundant SNP information is (or will be) available for breeding animals
- Number of SNPs will soon increase
- Main application of SNPs is genomic selection
- SNP information opens up many opportunities for conservation purposes

Acknowledgements

- Involved companies:
Hendrix Genetics, CRV (HG)
- Netherlands Organisation for Scientific Research
(NWO – Casimir)
- RobustMilk