# Estimation and control of inbreeding using Genomics

Introductory talk

by Theo Meuwissen

5th Annual Gensap meeting

# Introduction

- Traditional pedigree based selection:
  - Rate of genetic gain
    - Increase profit
  - Concerned about rate of inbreeding ($\Delta$F)
    - Maintain genetic variation
    - More than 1% $\Delta$F / generation is generally considered too much
  - Optimal contribution selection:
    - Maximise genetic gain at a restricted rate of inbreeding
    - Restrict $\Delta$F to 1% / generation
    - Restrict $\overline{A_{parents}}$ to $\overline{A_{population}} + 2\%$

# Pedigree based definition of F:

- Probability that two alleles at '*unlinked neutral locus*' are 'Identical By Descent' (IBD)

- Requires definition of base population where:
  - All alleles are unrelated and non-inbred, i.e. non-IBD
  - If alleles are the same in base: called 'Alike in State' (AIS)
    - i.e. they are by chance the same

- Question: do *unlinked neutral loci* exists in the era of genomics?
  - my answer is: NO

# Enter the genomics era

- Genomic prediction (GBLUP):
  - genomic relationships (G) more accurate than A
  - But in ssGBLUP there are issues putting G and A on the same scale
  - i.e. their are difficulties in comparing A and G
- G matrix relationships:

Frequency in base popul.

$$G_{ij} = \frac{(X_i - p)(X_j - p)}{p(1-p)} = \frac{(X_i X_j - pX_i - pX_j + p^2)}{p(1-p)}$$

- $E(X_iX_j) = F*p+(1-F)p^2 = F*p(1-p) + p^2$

$$E(G_{ij}) = F$$

F at neutral linked locus (SNP)

(Powell et al. 2010)

# Alternative genomic relationships:

- G matrix using alternative genotype standardisations (and $p_0$)
- G based on Runs of Homozygosity (ROH)
- Molecular coancestry
  - Actual homozygosity of alleles
- Others…
  - differ wrt their correction for AIS

# But what do we really want?

1. At QTL: increase frequencies of good alleles
   - Loss of genetic variance is inevitable
2. Low inbreeding depression => high heterozygosity
3. Maintain genetic variation for fitness and current neutral traits
   - Genetic variance is $V_g = \sum 2p_j(1-p_j)a_j^2$
     - High heterozygosity
4. Low frequency of lethal recessive diseased animals
   - High heterozygosity

# But what do we really want?

1. At QTL: increase frequencies of good alleles
   - Loss of genetic variance is inevitable
2. Low inbreeding depression => **high heterozygosity**
3. Maintain genetic variation for fitness and current neutral traits
   - Genetic variance is $V_g = \sum 2p_j(1-p_j)a_j^2$
     - **High heterozygosity**
4. Low frequency of lethal recessive diseased animals
   - **High heterozygosity**

# So: we want heterozygosity

- Everywhere in the genome
  - at fitness, disease and at loci that may become of interest in the future
- I.e. at '*linked neutral loci*' instead of at unlinked neutral loci

- In case of Whole Genome Sequence data:
  - We have direct measures of heterozygosity/genetic variance of all loci
  - So fitness, disease and loci of future interest are included
  - No need to bother about founder populations, IBD and AIS

- SNPchip data are in principle the same
  - Except they are not a random sample from all the loci in the genome

# Conclusions

1. In the era of genomics we can and should define inbreeding as:
   - IBD at ***neutral linked loci***
2. Reducing inbreeding at neutral linked loci will:
   - Reduce loss of genetic variance at loci that may become of future interest
   - Reduce inbreeding depression for fitness and other traits
   - Reduce genetic defects drifting to high frequencies
3. WGS data measures heterozygosity/variance directly for:
   - fitness, disease and loci of future intrest
4. Point 2 describes exactly our goals for inbreeding management