Genomic management of inbreeding in breeding schemes

Theo Meuwissen, Anna Sonesson, John Woolliams

Norwegian University of Life Sciences, Ås, Norway.
NOFIMA, Ås, Norway
Roslin Institute, Edinburgh, UK

Introduction

- Management ΔF by Optimum Contribution Selection (Meuwissen '97)
  - Maximise genetic gain / $\overline{EBV}$ of parents
  - Limit increase of coancestry / inbreeding
- In the era of genomics:
  - GEBV are by GBLUP / BayesA/B/C/R
    - Genomic relationshipmatrix **G**
- Question arises: move to Genomic Optimum Contributions (GOC)
  - Limit increase of genomic relationships **G**
  - Sonesson et al. (2012): YFS

However…

- GS-GOC resulted in correct ΔF_{genom} but low ΔG (Sonesson et al. 2012)
  - Compared to GS-AOC (yielded >>ΔF_{genom})
  - GS uses marker-set to achieve genetic progress /allele freq. changes
    - Allele freq. changes at QTL are desirable/markers are used as proxy for QTL
  - Can we use same marker-set to restrict allele freq. changes?

- F_{ped} = inbreeding at unlinked neutral loci
  - Do unlinked neutral loci exist in the genomics world?
  - F_{ped} underestimates real inbreeding in the genome

AIM:

1. What do we really want to achieve by managing ΔF?
   – Which measure of inbreeding should be used

2. Which OC method achieves this goal best
   – GOC / AOC / or a new G-matrix orthogonal to the direction of selection

3. Does ΔF management directly hinder ΔG and vice versa
   – ΔG => allele freq. changes (at QTL/ markers)
   – ΔF management => limit allele freq. changes

Goals of inbreeding management

Goals of inbreeding management

1. Want allele freq. changes at QTL in favorable direction
2. Avoid consequences from inbreeding:
    1. Inbreeding depression mainly at ‘fitness’ traits
    2. Loss of genetic variation at traits currently not of interest
        - Hypothetical trait may become of future interest
        - Selective sweeps should not erase most of the genetic variation
    3. Recessive disease alleles drifting to high frequency
        - Although disease mutations may be mapped and selected against
        - Diverts selection pressures away from breeding goal

The solution for the three problems

- Maintain genomic heterozygosity (Het)
  1. Inbreeding depression is proportional to loss of heterozygosity ($\Delta Het*d$)
  2. Genetic variance of hypothetical trait is Het_t*a^2
  3. Recessive diseases are not expressed in heterozygot form

New definition of rate of inbreeding

- $Het_t = Het_0 (1\text{-}F)$

  where F = inbreeding at the locus

    «the locus» = a neutral locus, but may be linked to QTL / markers

- F is thus inbreeding at neutral linked locus (F_{NL})

  – F_{ped} = inbreeding at neutral unlinked locus

- The rate of inbreeding at neutral linked loci is:

$$\Delta F_{NL} = \frac{Het_{t-1} - Het_t}{Het_{t-1}} = \frac{Hom_t - Hom_{t-1}}{1 - Hom_{t-1}}$$

Comparing alternative OC schemes

Comparing alternative OC methods

- AOC: pedigree relationships
- GOC: genomic relationships (VanRaden 2008; type II)
- GOC|b: genomic relationship orthogonal to the direction of selection **b**:

$$\mathbf{G|b}=\frac{\boldsymbol{XRX'}}{m}\ with\ \boldsymbol{R}=\boldsymbol{I}-\frac{\boldsymbol{bb'}}{\boldsymbol{b'b}}$$

b = solution vector of SNP effects (changes from generation to generation)

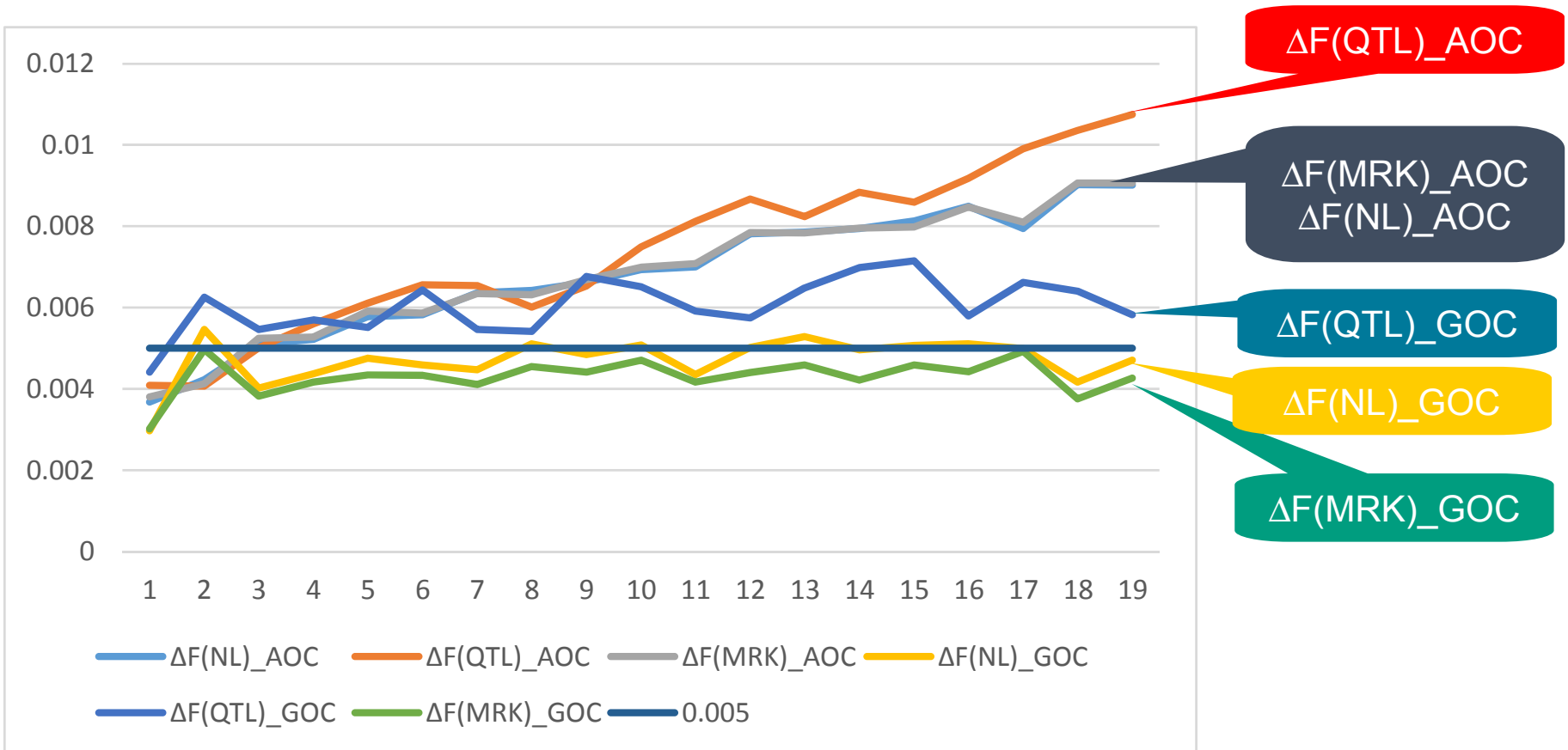
X = marker genotype matrix

- Desired rate of inbreeding: 0.5%.

Simulated fish breeding scheme

- Genome
  - 10 chromosomes of size 1 Morgan
  - SNP panel of 9000 SNPs
  - 1000 QTL loci (not in SNP panel)
  - 1000 neutral linked (not QTL/ not in SNP panel)
    - Monitor ΔF_{NL}.
- Genomic selection scheme:
  - 2000 fish/generation
    - 1000 selection candidates (only genotyped)
    - 1000 sibs (genotyped + phenotyped (h^2=0.4))

ΔF : AOC vs. GOC


0.012
0.01
0.008
0.006
0.004
0.002
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
ΔF(NL)_AOC
ΔF(QTL)_AOC
ΔF(MRK)_AOC
ΔF(NL)_GOC
ΔF(QTL)_GOC
ΔF(MRK)_GOC
0.005
ΔF(QTL)_AOC
ΔF(MRK)_AOC
ΔF(NL)_AOC
ΔF(QTL)_GOC
ΔF(NL)_GOC
ΔF(MRK)_GOC

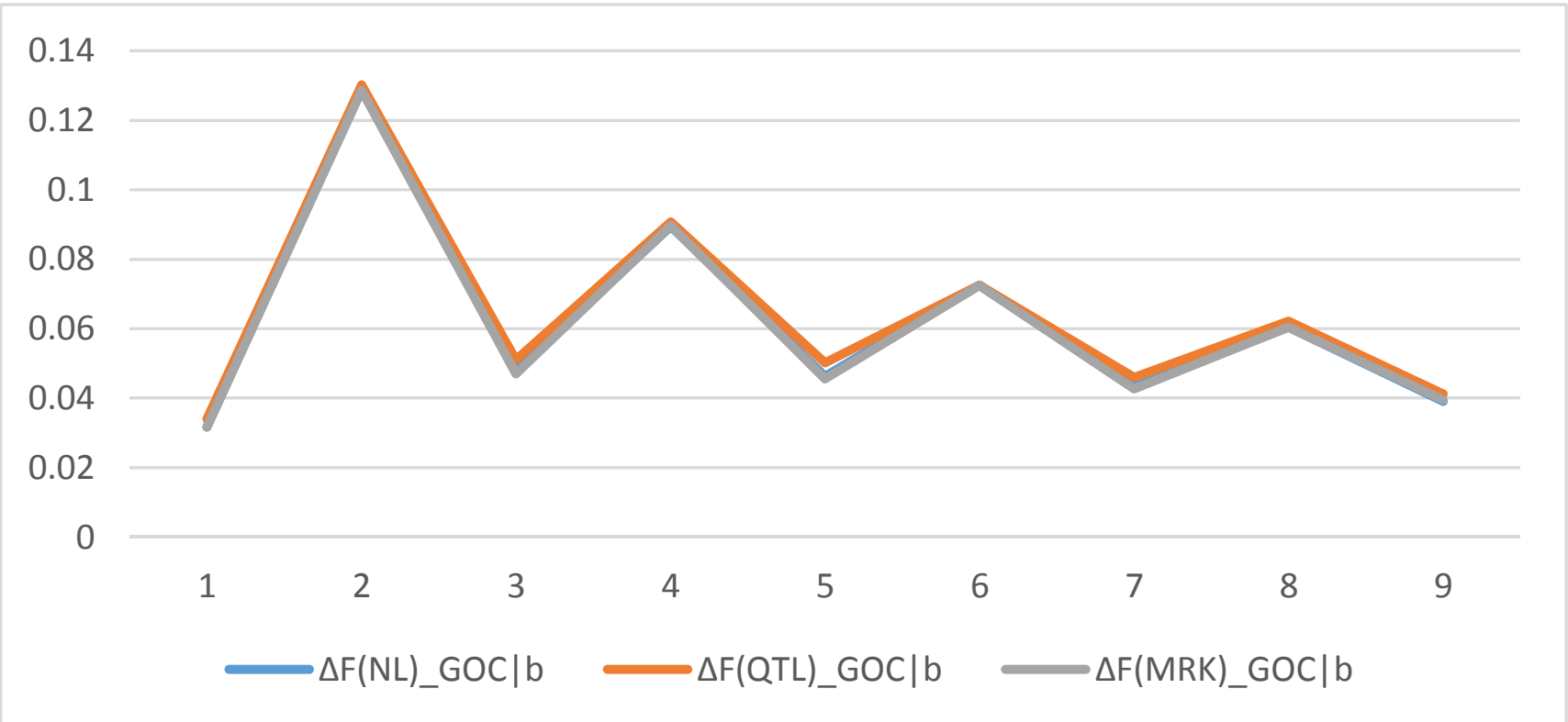
Results (so far):

- GOC maintains ΔF restriction at neutral linked loci
- ΔF(QTL) >> ΔF(MRK)
  - GS causes allele frq. changes at SNPs => associated changes at QTL
    - Freq. changes at QTL >> freq. changes at SNPs

Results (so far):

- GOC maintains ΔF restriction at neutral linked loci
- ΔF(QTL) >> ΔF(MRK)
  - ~~GS causes allele frq. changes at SNPs => associated changes at QTL~~
    - ~~Freq. changes at QTL < freq. changes at SNPs~~
- AOC does not control the rate of inbreeding
  - ΔF_{NL} keeps increasing over 20 generations of selection

ΔF when G|b method was used

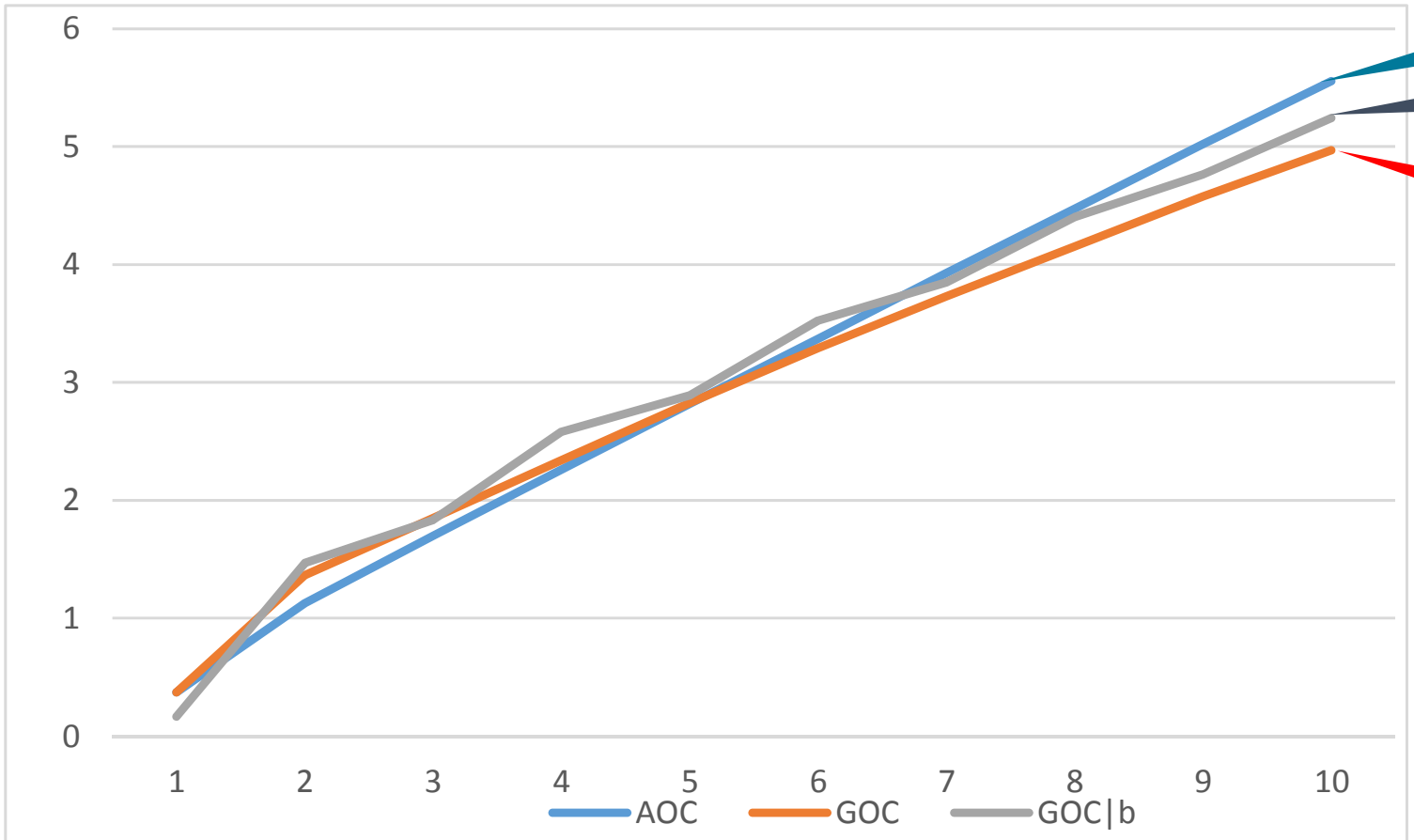


Results from G|b method

- $\Delta F\_GOC|b$ exceeded ΔF_{NL} restriction

- ΔF is very similar at QTL, SNPs and NL-loci
  - Seems drift at NL-loci and SNPs follows that at QTL
    - Majority of inbreeding is due to inbreeding at QTL, which is unrestrained

Genetic improvement



AOC

GOC|b

GOC

ΔG results

- Initially GOC and GOC|b yield highest genetic gain
- GOC|b gives generaly more gain than GOC
  - But also more ΔF_{NL}.
- AOC yields lower ΔG intially but highest ΔG in the long term
  - due to the ever increasing ΔF_{NL}.
  - ΔG does not decrease even though genetic variance reduces

Discussion

Define ΔF_{NL} as 'Genomic ΔF'

- Inbreeding is about the unknown risks of breeding
  - About variation outside the breeding objective / causal variants
  - Variation at neutral loci which may be linked to causal variants
  - Genomic ΔF measures inbreeding due to genome and pedigree structure
    - ΔF_{ped} is inbreeding rate due to pedigree structure alone

- Selection is about the causal variants
  - Correlated responses are due to pleiotropic effects of causal variants
    - Seperates risks due to correlated responses from those due to inbreeding

WGS data and ΔF management

- ΔF management was directed at:
  - inbreeding depression at ‘fitness’ traits
  - maintaining genetic variance at hypothetical trait
  - avoiding drift at recessive disease loci
  - all these refer to ‘anonymous’ loci that occur in the genome
- But WGS data contains all these loci:
  - I.e. WGS data over several generations => genomic ΔF (directly)
    - This differs from GEBV estimation: WGS data ≠> GEBV

Runs of homozygosity: F_{ROH}

- Distinguishes IBD from IBS
  - 'Long' ROH => IBD
    - Using too long ROH : misses out on true IBD
  - HOM is special case of F_{ROH} (runs of 1 SNP)
- Problem using long ROH:
  - Long ROHs recombine and break up
    - Thus F_{ROH} does not accumulate in the same way as inbreeding does

$$F_{t+1} \neq F_t + (1 - F_t)\Delta F$$

  - So if using F_{ROH} don't use too long runs
    - Probability of recombination over the considered time period should be negligible

Conclusions

On genomic ΔF

- Define the genomic rate of inbreeding as ΔF at neutral linked loci
- Adresses 3 main problems of inbreeding
    1. Inbreeding depression ‘fitness’ traits
    2. Genetic variation at hypothetical trait
    3. Drift at recessive disease loci
- Seperates risk from ΔF vs. from correlated selection responses
- WGS data can measure genomic ΔF
    - Contains all these ‘anonymous’ loci
    - Overwhelms the (relatively few) causal variants

Conclusions wrt GOC schemes

- GOC controls the rate of genomic inbreeding (ΔF_{NL})
- AOC did not control ΔF_{NL}
  - ΔF_{NL} kept on increasing over 20 generations
- G|b relationship matrix resulted also in too high genomic ΔF.
- Hypothesis: GS affects freqs of SNPs => freq changes at QTL
  - Disproved: inbreeding at QTL >> inbreeding at SNPs
  - Can combine GOC and GS using a SNP panel

Acknowledgements:

- Christian Sørensen, Mark Henryon, Gebreyohans Gebregiwergis, Peer Berg

- Norwegian Research Council (prj.nr. *226275/E40*)