GS cattle workshop

Genotyping strategy and reference population

- Effect of size of reference group (Esa Mäntysaari, MTT)
- Effect of adding females to the reference population (Minna Koivula, MTT)
- Value of using GS at herd level (Line Hjortø, VFL)

- Discussion

18:30 Dinner

Effect of size of reference group

Effect of adding females to the reference population

Esa Mäntysaari , Minna Koivula,
Timo Knürr, Ismo Strandén,

MTT, Biotechnology and Food Research,
Biometrical genetics

Contents

- Prediction of accuracy of genomic prediction R^2
  - Genomic Model
  - Effect of number of reference bulls on R^2
- Effect of adding genotyped females into reference population
  - Using single step evaluation and deregressed cow EBVs
- Conclusion:
  - How many cows are needed

Accuracy of Genomic evaluation

- Several equations exist for predicting the accuracy of DGV
  - Daetwyler et al, 2008; Goddard, 2009; Hayes et al. 2009; Goddard et al. 2011; Meuwissen et al. 2013)

  - Generally reliability of prediction for the animals that have no phenotypes themselves:

$$R_{DGV}{}^{2} = w\frac{Nref * h^{2}}{Nref * h^{2} + Me}$$

where

- *w* is the proportion of genetic variance that can be predicted by genomic model
- *Nref* is the number of animals with genotypes and phenotypes
- h^2 is the prediction accuracy of the phenotypes
- Me is the number of haplotypes segregating in the population

Accuracy of Genomic evaluation

- Several equations exist for predicting the accuracy of DGV
  - Daetwyler et al, 2008; Goddard, 2009; Hayes et al. 2009; Goddard et al. 2011; Meuwissen et al. 2013)

  - Generally reliability of prediction for the animals that have no phenotypes themselves:

$$R_{DGV}{}^{2} = w\frac{Nref * h^{2}}{Nref * h^{2} + Me}$$

where

- *w* is the proportion of omic model
- *Nref* is the number of a
- h^2 is the prediction acc
- Me is the number of ha

w

Relative to the genetic structure of the trait and the genotyping tool, for example SNP density

Accuracy of Genomic evaluation

- Several equations exist for predicting the accuracy of DGV
  - Daetwyler et al, 2008; Goddard, 2009; Hayes et al. 2009; Goddard et al. 2011; Meuwissen et al. 2013)

  - Generally reliability of prediction for the animals that have no phenotypes themselves:

$$R_{DGV}{}^{2} = w\frac{Nref * h^{2}}{Nref * h^{2} + Me}$$

where

- *w* is the proportion genomic model
- *Nref* is the number
- h^2 is the prediction
- Me is the number

Nref

In principal independent non-related animals (with the same amount of information

Accuracy of Genomic evaluation

- Several equations exist for predicting the accuracy of DGV
  - Daetwyler et al, 2008; Goddard, 2009; Hayes et al. 2009; Goddard et al. 2011; Meuwissen et al. 2013)

  - Generally reliability of prediction for the animals that have no phenotypes themselves:

$$R_{DGV}{}^{2} = w\frac{Nref * h^{2}}{Nref * h^{2} + Me}$$

where

- *w* is the proportion of omic model
- *Nref* is the number of a
- h^2 is the prediction acc
- Me is the number of ha

h^2

Accuracy of observation:

heritability or reliability (with same amount of information for any)

Accuracy of Genomic evaluation

- Several equations exist for predicting the accuracy of DGV
  - Daetwyler et al, 2008; Goddard, 2009; Hayes et al. 2009; Goddard et al. 2011; Meuwissen et al. 2013)

  - Generally reliability of prediction for the animals that have no phenotypes themselves:

$$R_{DGV}{}^{2} = w\frac{Nref * h^{2}}{Nref * h^{2} + Me}$$

where

- *w* is the proportion of … nomic model
- *Nref* is the number of a…
- h^2 is the prediction acc…
- Me is the number of ha…

Me

Depends on genetic structure of trait, and population. Mostly related to effective population size N_e

Me = 2 Ne Lm / log(NeLm)

Ne =effective pop size

Lm = genome size

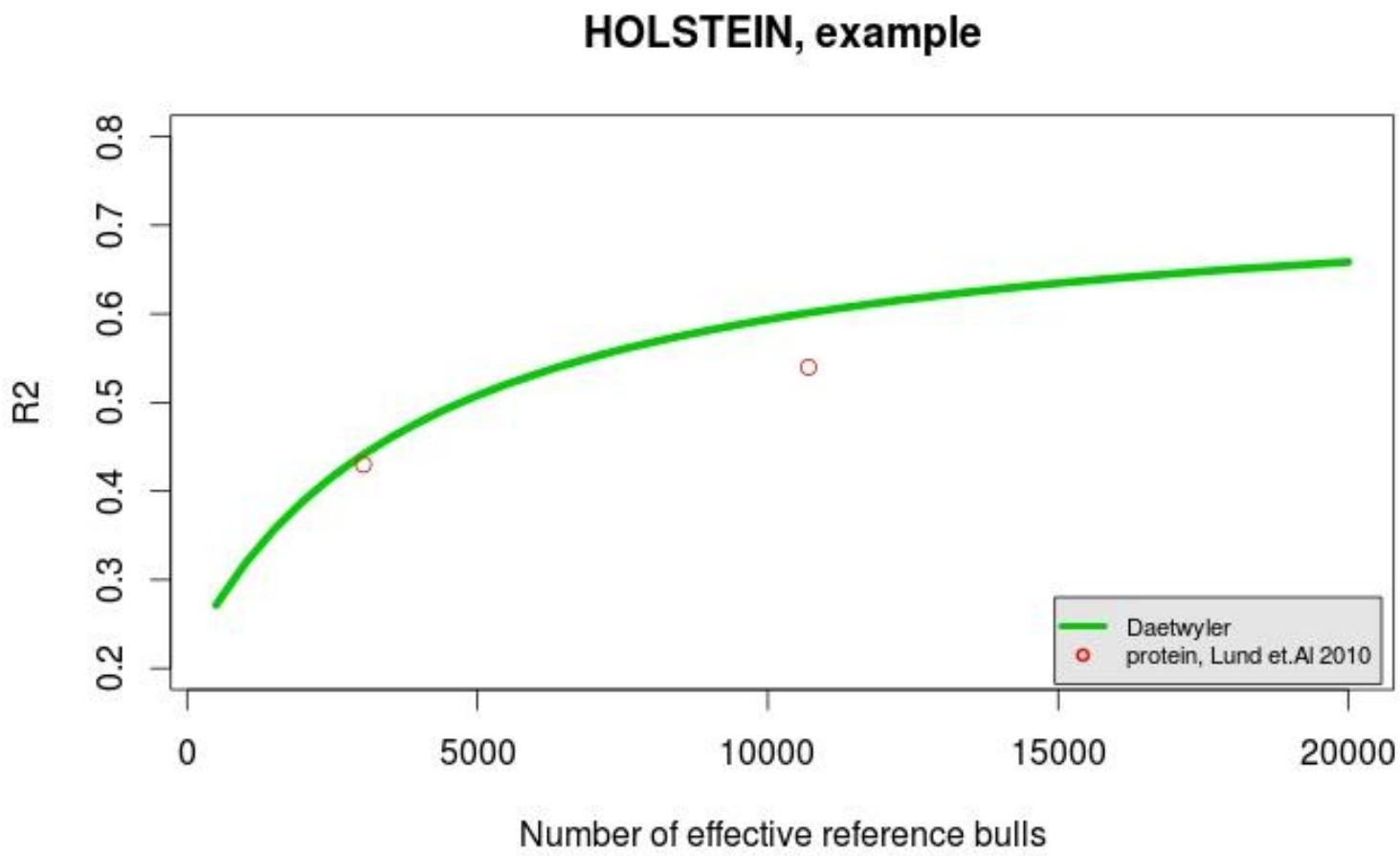
Accuracy of Genomic evaluation (2)

- The prediction generally fits poorly to our data
  - Mäntysaari et al (now) suggest a correction that takes into account the dependencies within data

$$R_{DGV}{}^{2} = w\frac{Nref_1 + \Delta(Nref - Nref_1)}{Nref_1 + \Delta(Nref - Nref_1) \; + \; Me/h^2}$$

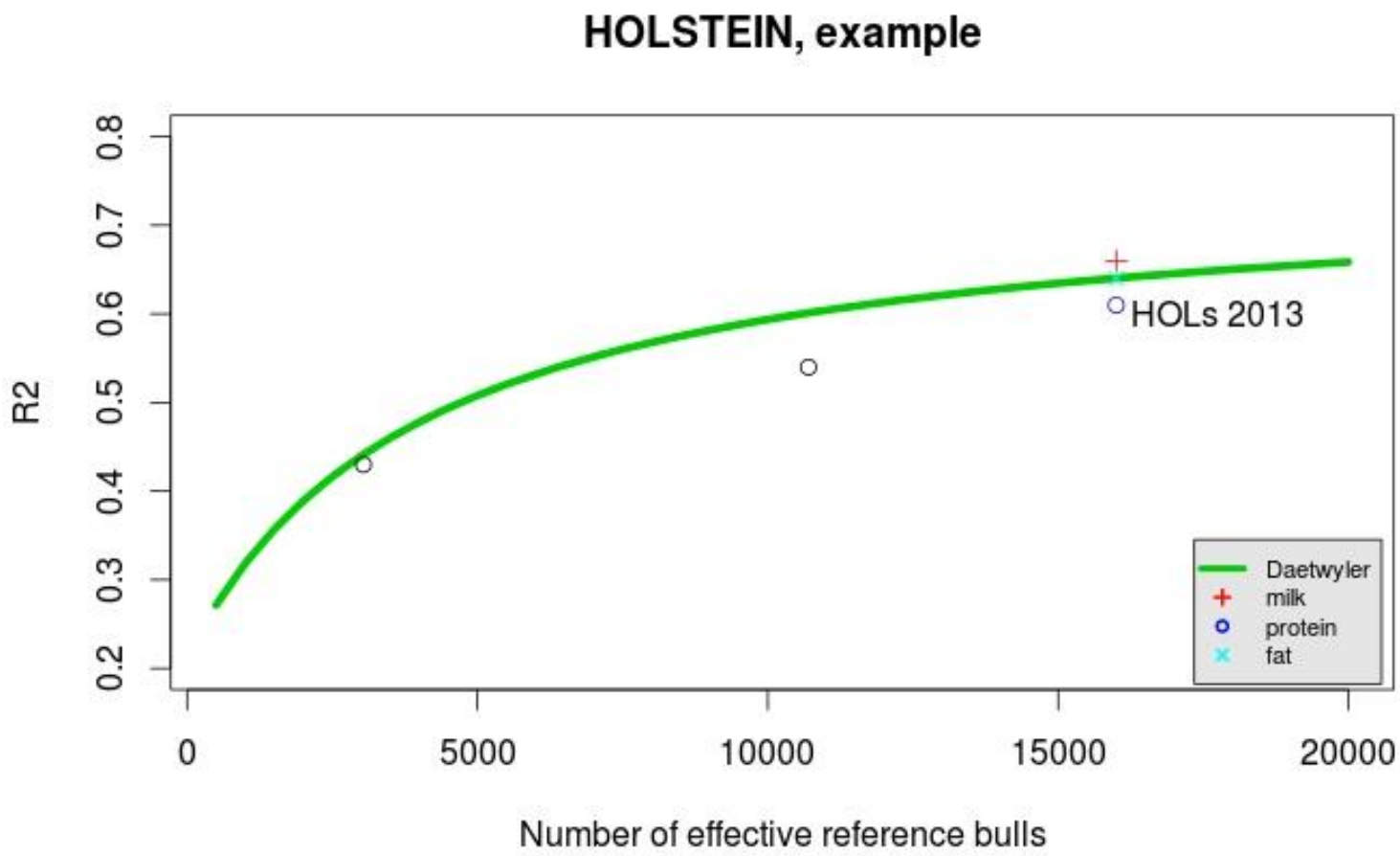
where

- Δ (Nref) is a proportion of data increase that is independent from the smallest reference population level $Nref_1$
- else, as before



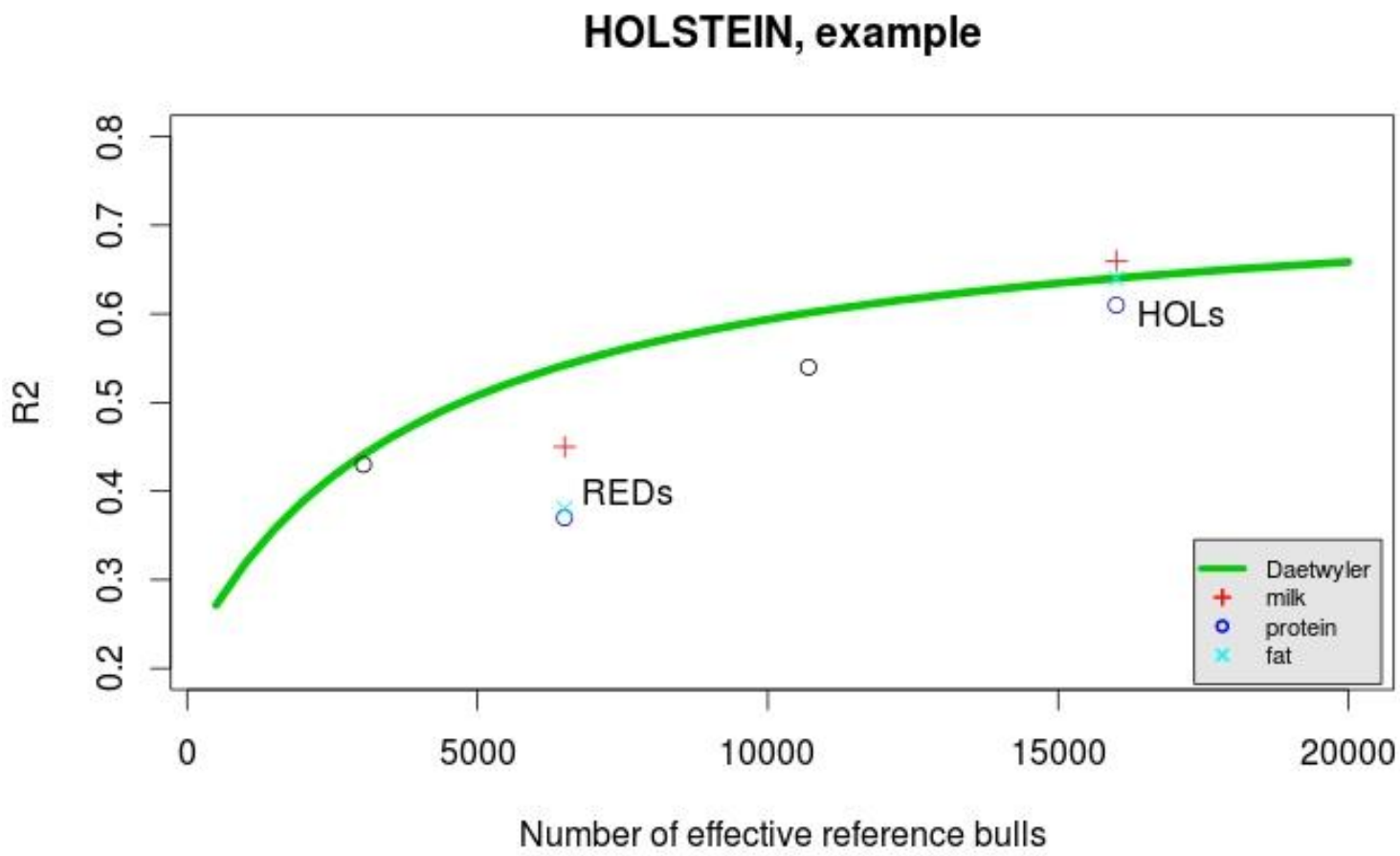
Assume:

- h^2 =0.85
- Ne=100
- w = 0.75
- Δ Nref = 30%



Assume:

- h^2 =0.85
- Ne=100
- w = 0.75
- Δ Nref = 30%



Assume:

- h^2 =0.85
- Ne=100
- w = 0.75
- Δ Nref = 30%

Prediction of sire GBLUP accuracy

Empirical data from GBLUP control model run by Knürr et al. EAAP 2013

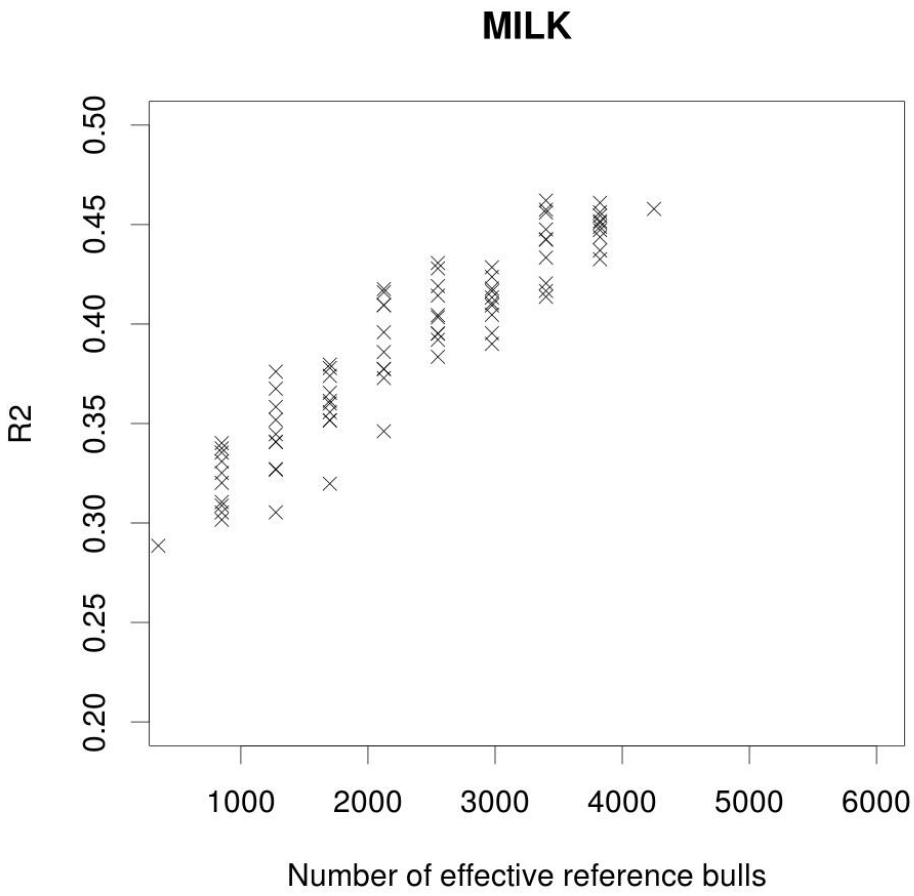
- Reduced data with 4250 training bulls
- 38194 SNPs
- DRPs received from NAV
- GBLUP model w. 10% polygenic
-

Prediction of sire GBLUP accuracy

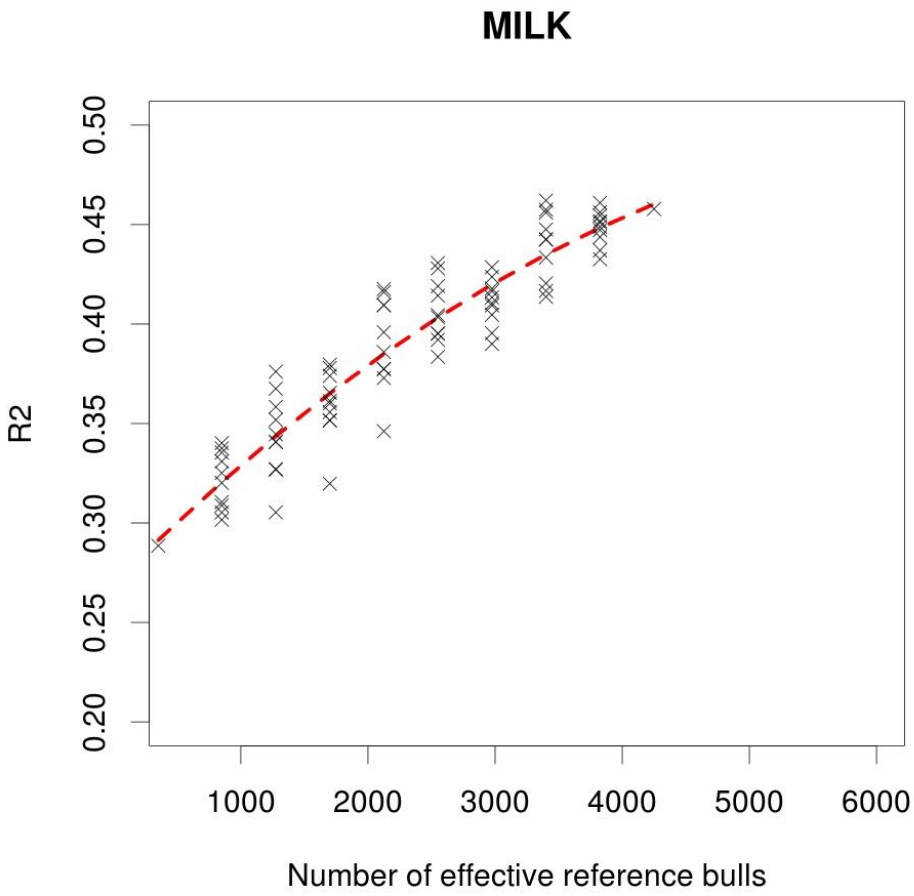
Reduced reference data sets:

1. Remove all the bulls w.out sons: Minimal reference set 351 bulls
2. From 3900 non-parents: use sampling 20,30,…,100% reference population size
   - 10 replicates for each size of reference population
3. Use each sample to estimate GBLUP for validation bulls

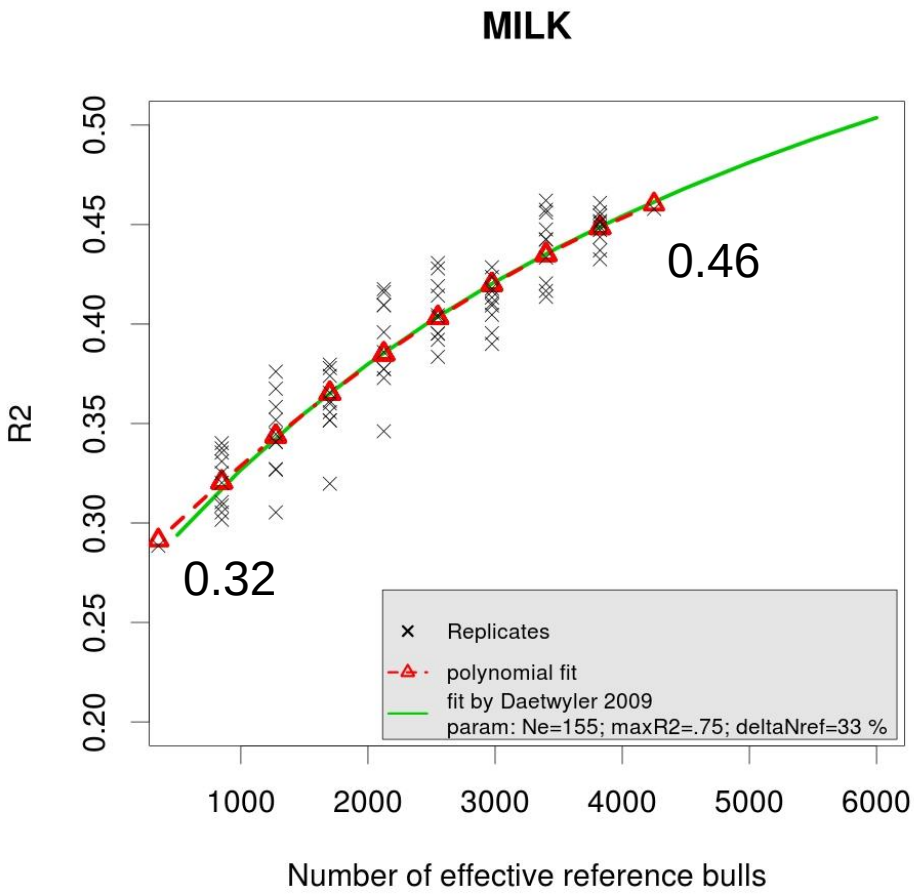
→ Validation $R^2_{ref\%}$



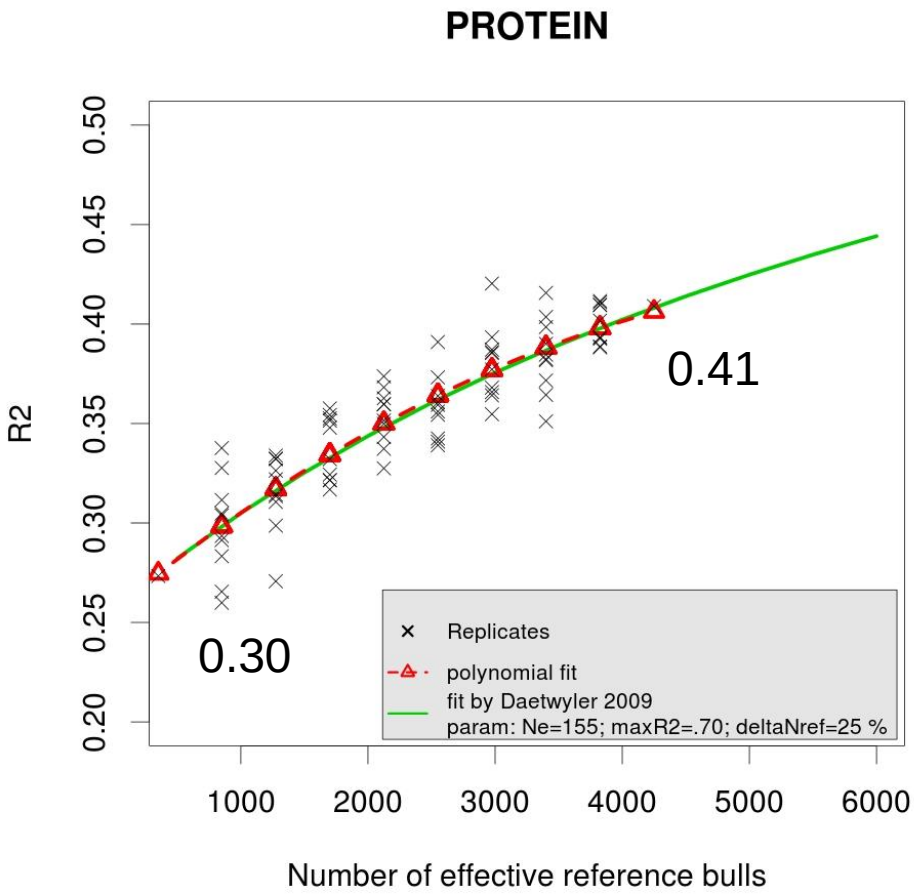
- 82 samples, the full model R^2 was 0.46



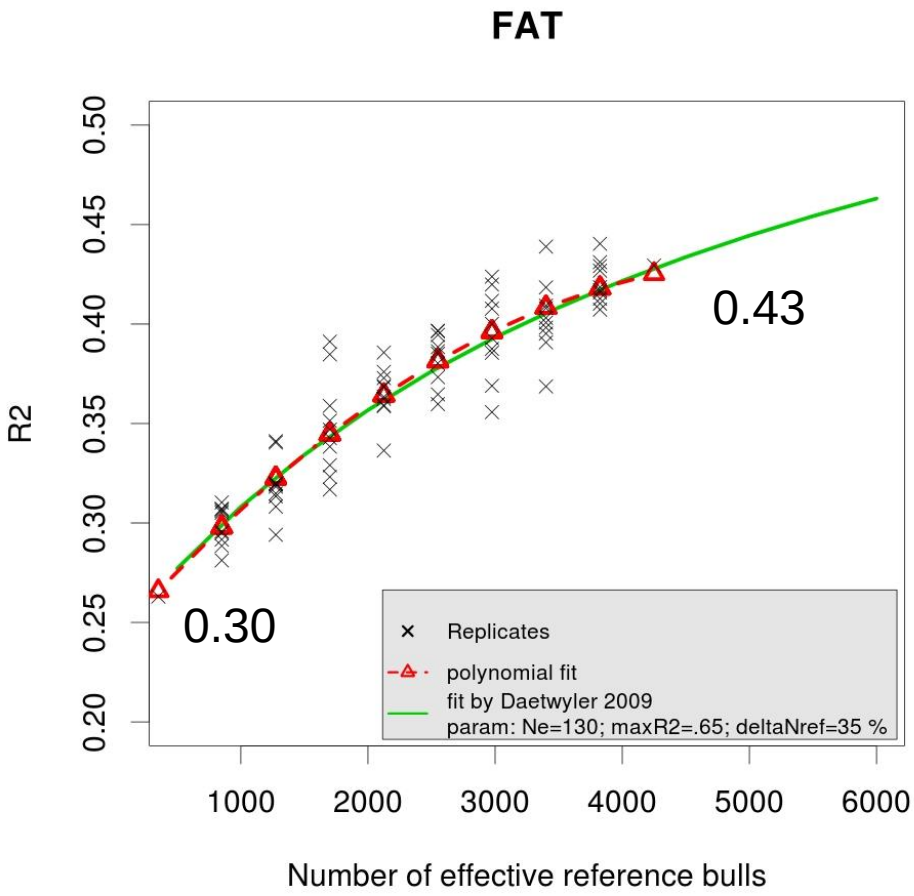
- 82 samples
- Second order polynomial fit shows clear curvature



- Modified Daetwyler prediction fits well:
  - Effective population size 155
  - If base curve is fitted to 850 bulls (R^2=0.32), each extra bull contributes only 33% more
  - Suggests that asymptotic accuracy is R^2=0.75



- Modified Daetwyler prediction fits well:
  - Effective population size 155
  - If base curve is fitted to 850 bulls (R^2=0.30), each extra bull contributes only 25% more
  - Suggests that asymptotic accuracy is R^2=0.70



- Modified Daetwyler prediction fits well:
  - Effective population size 135
  - If base curve is fitted to 850 bulls (R^2=0.30), each extra bull contributes only 35% more
  - Suggests that asymptotic accuracy is R^2=0.65

	1000 reference bulls contribute %-units	Reference population to get R^2=0.55
Milk	2.22	8800
Protein	1.94	16000
Fat	1.86	14600

- The prediction model does not work well
- Only the milk equation is usable:
  - Increasing the validation R^2 from 0.46 to 0.55 would require 4400 new bulls genotyped
  - Or theoretically genotyping 10200 cows

Single step evaluation with cow DRPs - animal model DRPs

Minna Koivula, Ismo Strandén,
Esa A. Mäntysaari

Data sets and evaluation models:

- Nordic production test-day data from July 2013
  - 3.7 million cows with records
  - 4.9 million animals in the Nordic Red pedigree

- DRPs for cows with edc >0
  - DRP = μ + EBV + ε
  - Heritabilities
    - Milk 0.48
    - Protein 0.48
    - Fat 0.49

  - DRPs for 3,072,815 RDC cows

For one step and validation:

Reduced data I

- DRPs of young genotyped cows included
  - 2,947,546 million cow DRPs
  - 3137 genotyped cows in reference population
  - Daughters of validation bulls removed

Reduced data II,

- DRPs of **genotyped bulldams** excluded
- DRPs genotyped **young cows** excluded
- Daughters of validation bulls removed
  - 2,944,409 million cow DRPs

- Genotype data (September 2013)
  - 46943 markers
  - 9107 genotyped animals in Nordic Red pedigree
    - 5315 bulls and 3792 cows

Single step model with cow DRPs:

- Pedigree extracted for 9107 animals with genotypes
- $\mathbf{G^*} = \mathbf{G_w}^{-1} - \mathbf{A_{22}}^{-1}$

1) $\mathbf{A}^{-1}$constructed using full pedigree file with all animals
2) **G** -matrix scaled with $\Sigma 2pq$ and $\Sigma G_{ii} / \Sigma A_{ii}$
   a) 0.20 weight for polygenic $\mathbf{A_{22}}$ in $\mathbf{G_w}$ (Chistensen and Lund)
   b) $\mathbf{G}^* = 1.6^* \mathbf{G_w}^{-1} - 0.5^* \mathbf{A}^{-1}$ (Mistzal), where w=0.10

Mean EDCs and DRPs ± SDs
for genotyped reference cows and full cow DRP data

- Genotyped reference cows (n=3137)
  - Milk
    - 0.738 ± 0.26
    - 2112.81 ± 1357.20
  - Protein
    - 0.669 ± 0.26
    - 80.268 ± 46.86
  - Fat
    - 0.671 ± 0.28
    - 80.82 ± 60.51
- Full data
  - Milk
    - 0.971 ± 0.23
    - 653.89 ± 1416.67
  - Protein
    - 0.917 ± 0.25
    - 22.657 ± 50.84
  - Fat
    - 0.926 ± 0.27
    - 25.23 ± 62.01

Mean EDCs, DRPs and DYDs ± SDs for validation bulls and reference bulls

- Validation bulls (n= 769)
  - Milk
    - 18.83 ± 8.19
    - 2259.82 ± 615.36
    - 990.21 ± 330.37
  - Protein
    - 17.53 ± 7.94
    - 79.31 ± 17.92
    - 37.93 ± 9.85
  - Fat
    - 17.19 ± 7.86
    - 87.77 ± 23.94
    - 39.39 ± 12.12
- Reference bulls
  - Milk
    - 17.70 ± 116.47
    - 1159.15 ± 718.19
    - 393.31 ± 808.93
  - Protein
    - 17.03 ± 113.01
    - 36.41 ± 24.51
    - 14.02 ± 28.24
  - Fat
    - 16.82 ± 112.22
    - 41.84 ± 28.82
    - 14.85 ± 35.18

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	b_1	R^2	b_1	R^2
PA	**0.878**	**0.361**	**0.750**	**0.269**	**0.674**	**0.286**
$GEBV_{w80}$	**0.858**	**0.506**	**0.762**	**0.414**	**0.758**	**0.473**
$GEBV_{BullG\_w80}$	**0.866**	**0.482**	**0.772**	**0.402**	**0.766**	**0.460**
$GEBV_{Misztal}$	**0.994**	**0.511**	**0.890**	**0.430**	**0.874**	**0.482**
$GEBV_{BullG\_M}$	**0.968**	**0.488**	**0.876**	**0.419**	**0.874**	**0.472**

Reduced Data I

- 769 candidate bulls, born 2005 – 2009
- DRPs of genotyped cows included
  - **3137 genotyped cows in reference population**

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	b_1	R^2	b_1	R^2
PA	**0.878**	**0.361**	**0.750**	**0.269**	**0.674**	**0.286**
$GEBV_{w80}$	**0.858**	**0.506**	**0.762**	**0.414**	**0.758**	**0.473**
$GEBV_{BullG\_w80}$	**0.866**	**0.482**	**0.772**	**0.402**	**0.766**	**0.460**
$GEBV_{Misztal}$	**0.994**	**0.511**	**0.890**	**0.430**	**0.874**	**0.482**
$GEBV_{BullG\_M}$	**0.968**	**0.488**	**0.876**	**0.419**	**0.874**	**0.472**

Reduced Data I

- 769 candidate bulls, born 2005 – 2009
- DRPs of genotyped cows included
  - **3137 genotyped cows in reference population**

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	b_1	R^2	b_1	R^2
PA	**0.878**	**0.361**	**0.750**	**0.269**	**0.674**	**0.286**
$GEBV_{w80}$	**0.858**	**0.506**	**0.762**	**0.414**	**0.758**	**0.473**
$GEBV_{BullG\_w80}$	**0.866**	**0.482**	**0.772**	**0.402**	**0.766**	**0.460**
$GEBV_{Misztal}$	**0.994**	**0.511**	**0.890**	**0.430**	**0.874**	**0.482**
$GEBV_{BullG\_M}$	**0.968**	**0.488**	**0.876**	**0.419**	**0.874**	**0.472**

Reduced Data I

- 769 candidate bulls, born 2005 – 2009
- DRPs of genotyped cows included
  - **3137 genotyped cows in reference population**

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	b_1	R^2	b_1	R^2
PA	**0.894**	**0.364**	**0.778**	**0.273**	**0.724**	**0.289**
$GEBV_{w80}$	**0.870**	**0.491**	**0.784**	**0.409**	**0.784**	**0.461**
$GEBV_{BullG\_w80}$	**0.882**	**0.491**	**0.794**	**0.409**	**0.802**	**0.466**
$GEBV_{Misztal}$	**1.012**	**0.501**	**0.926**	**0.428**	**0.894**	**0.473**
$GEBV_{BullG\_M}$	**0.980**	**0.495**	**0.894**	**0.425**	**0.898**	**0.475**

Reduced Data II
- 769 candidate bulls, born 2005 – 2009
- DRPs of genotyped cows and bulldams excluded

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	B_1	R^2	b_1	R^2
PA	**-0.016**	**-0.003**	**-0.028**	**-0.004**	**-0.050**	**-0.003**
$GEBV_{w80}$	**-0.012**	**0.015**	**-0.022**	**0.005**	**-0.026**	**0.012**
$GEBV_{BullG\_w80}$	**-0.016**	**-0.009**	**-0.022**	**-0.007**	**-0.036**	**-0.006**
$GEBV_{Misztal}$	**-0.018**	**0.010**	**-0.016**	**0.002**	**-0.002**	**0.009**
$GEBV_{BullG\_M}$	**0-.012**	**-0.007**	**-0.018**	**-0.006**	**-0.024**	**-0.003**

Difference = Reduced Data I - Reduced Data II

-Effect of DRPs of 3137 reference cows

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	B_1	R^2	b_1	R^2
PA	**-0.016**	**-0.003**	**-0.028**	**-0.004**	**-0.050**	**-0.003**
$GEBV_{w80}$	**-0.012**	**0.015**	**-0.022**	**0.005**	**-0.026**	**0.012**
$GEBV_{BullG\_w80}$	**-0.016**	**-0.009**	**-0.022**	**-0.007**	**-0.036**	**-0.006**
$GEBV_{Misztal}$	**-0.018**	**0.010**	**-0.016**	**0.002**	**-0.002**	**0.009**
$GEBV_{BullG\_M}$	**0-.012**	**-0.007**	**-0.018**	**-0.006**	**-0.024**	**-0.003**

Difference = Reduced Data I - Reduced Data II

-Effect of DRPs of 3137 reference cows

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	B_1	R^2	b_1	R^2
PA	**-0.016**	**-0.003**	**-0.028**	**-0.004**	**-0.050**	**-0.003**
$GEBV_{w80}$	**-0.012**	**0.015**	**-0.022**	**0.005**	**-0.026**	**0.012**
$GEBV_{BullG\_w80}$	**-0.016**	**-0.009**	**-0.022**	**-0.007**	**-0.036**	**-0.006**
$GEBV_{Misztal}$	**-0.018**	**0.010**	**-0.016**	**0.002**	**-0.002**	**0.009**
$GEBV_{BullG\_M}$	**0-.012**	**-0.007**	**-0.018**	**-0.006**	**-0.024**	**-0.003**

Difference = Reduced Data I - Reduced Data II

-Effect of DRPs of 3137 reference cows

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	B_1	R^2	b_1	R^2
PA	**-0.016**	**-0.003**	**-0.028**	**-0.004**	**-0.050**	**-0.003**
$GEBV_{w80}$	**-0.012**	**0.015**	**-0.022**	**0.005**	**-0.026**	**0.012**
$GEBV_{BullG\_w80}$	**-0.016**	**-0.009**	**-0.022**	**-0.007**	**-0.036**	**-0.006**
$GEBV_{Misztal}$	**-0.018**	**0.010**	**-0.016**	**0.002**	**-0.002**	**0.009**
$GEBV_{BullG\_M}$	**0-.012**	**-0.007**	**-0.018**	**-0.006**	**-0.024**	**-0.003**

Difference = Reduced Data I - Reduced Data II

-Effect of DRPs of 3137 reference cows

Conclusion 2: if genotyped cow DRP excluded, better to exclude also the genotypes

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

	Milk		Protein		Fat	
	b_1	R^2	b_1	R^2	b_1	R^2
PA	**0.878**	**0.361**	**0.750**	**0.269**	**0.674**	**0.286**
$GEBV_{w80}$						
$GEBV_{BullG\_w80}$	**-0.008**	**-0.024**	**-0.010**	**-0.012**	**-0.002**	**-0.013**
$GEBV_{Misztal}$						
$GEBV_{BullG\_M}$	**-0.026**	**-0.023**	**-0.014**	**-0.011**	**0.000**	**-0.010**

Reduced Data I all genotypes - Bull genotypes only

- DRPs of genotyped cows included (*Reduced Data I)*
  - In Data I: 3137 genotyped cows in reference population

Conclusion: if genotyped cow DRP included, better to include also the genotypes

Validation results for bulls

$$R^2_{validation} = R^2_{model} / \overline{r^2_{DRP}}$$

Regression of DYD to $GEBV_R$ or EBV_R(PA)

BIAS in Single step GBLUP can be reduced !

	Milk		Protein		Fat	
	b_1	R^2	b_1	R^2	b_1	R^2
PA						
$GEBV_{w80}$						
$GEBV_{BullG\_w80}$						
$GEBV_{Misztal}$	**0.136**	**0.005**	**0.128**	**0.016**	**0.116**	**0.009**
$GEBV_{BullG\_M}$	**0.102**	**0.006**	**0.104**	**0.017**	**0.108**	**0.012**

Reduced Data I
Misztal – w80

Results of including cows into reference

- For milk and fat the improvement in R^2 due to 3137 reference cows was substantial:
  0.9-1.5 %-units
  - Furthermore,
    these additional 3137 cows genotyped would correspond 684 bulls
  - Each extra 4.3-4.6 cows would equal to extra bull
  - TO REACH R^2=0.55 by increasing cows genotyped ?
    - 3137*(0.55-0.46)/0.015 = 18822

Conclusions

- Theory:
  - Using Daetwyler et al. (2008) increase in R^2 = 3 %-units / 1000 bulls
  - Using conservative Daetwyler increase in R^2 = 1.1 %-units /1000 bulls
  - Using conservative Daetwyler increase in R^2 = 1.4 %-units / 3000 cows

- Based on SS GBLUP (and milk)
  - Increase in validation R^2 = 1.5 %-units / 3137 cows

- Estimate of number of chromosome segments is different for different traits?

- Theoretical prediction of increase R^2 is relative to ”information count” in data:

  If the h^2 is lower, the value of cow data is less

  when h^2=0.10
  - Using Daetwyler increase in R^2 = 2.8 %-units / (4000♂ →5000 ♂)
  - Using Daetwyler increase in R^2 = 1.4 %-units / (4000 ♂ →4000♂+3000♀)

THANK YOU !