

# Accuracy and bias of genomic prediction for second generation candidates

Zengting Liu

H. Alkhoder, F. Reinhardt and R. Reents

vit, Germany, <http://www.vit.de/>

# Introduction: Candidates with 2 or 3 generations apart from the bull reference population



German Holstein genomic evaluation: November 2015

-- including 177,178 genotyped animals, 33,436 bulls in EuroGenomics RP

**Table 1. Percentage of candidates without sire in reference population**

Year of birth	Number of candidates	Regular traits	Late-measured (longevity)	Early measured (direct CE)
2013	29,438	18%	77%	0%
2014	32,048	80%	95%	4%
2015	19,288	96%	96%	44%

10-15% are third-generation candidates, e.g. embryos, for milk yield



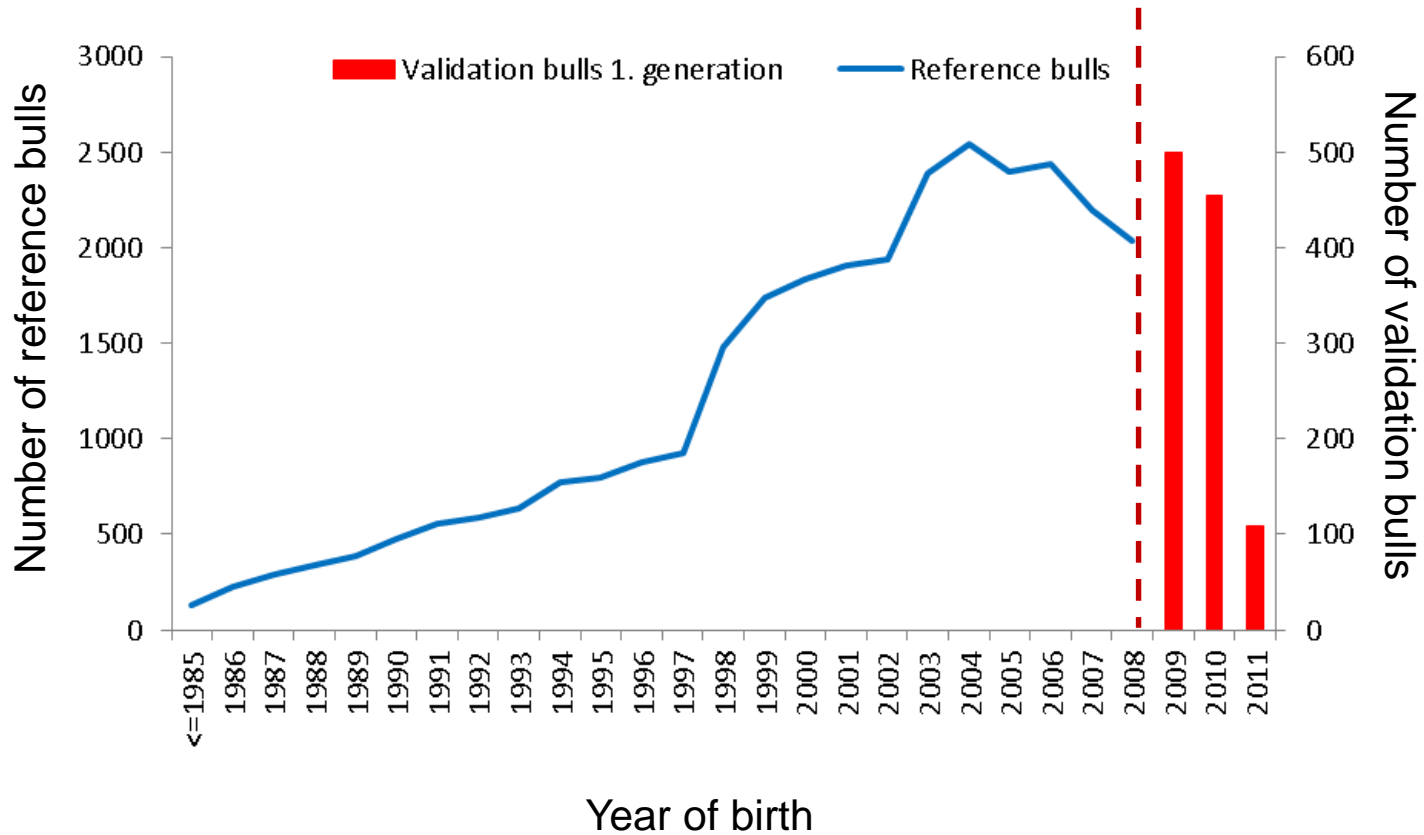
# Impacts of shorter generation intervals by genomic selection

- First generation candidates are disappearing
  - Distance btw. reference population (RP) and candidates increased
  - Top genomic bulls almost all from 2. or 3. generations
- Linkage disequilibrium (LD) between SNP markers and genes
  - Genomic evaluation usually does not account for the LD breakup from 1. to 2. generation
- Genomic validation methods work only for 1. generation candidates
  - LD breakup btw. RP and (1. generation) validation bulls considered
  - LD breakup btw. 1. and 2. (2. and 3.) generations NOT considered



# A regular genomic validation study for first generation candidates (I)

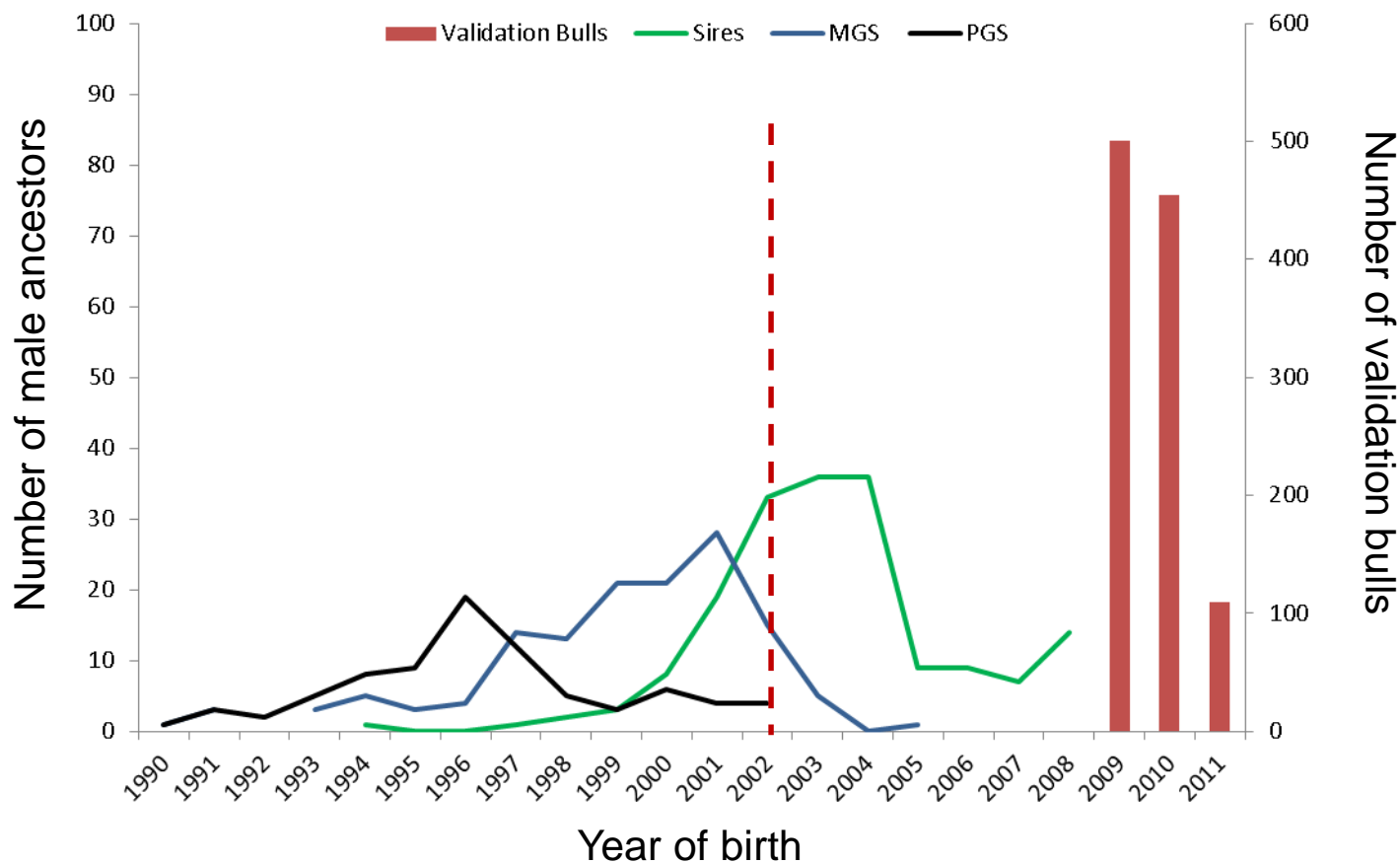
In total: 33,436 reference bulls  
 29,917 reference bulls  
 1063 national validation bulls



Milk yield, December 2015



# A regular genomic validation study for first generation candidates (II)



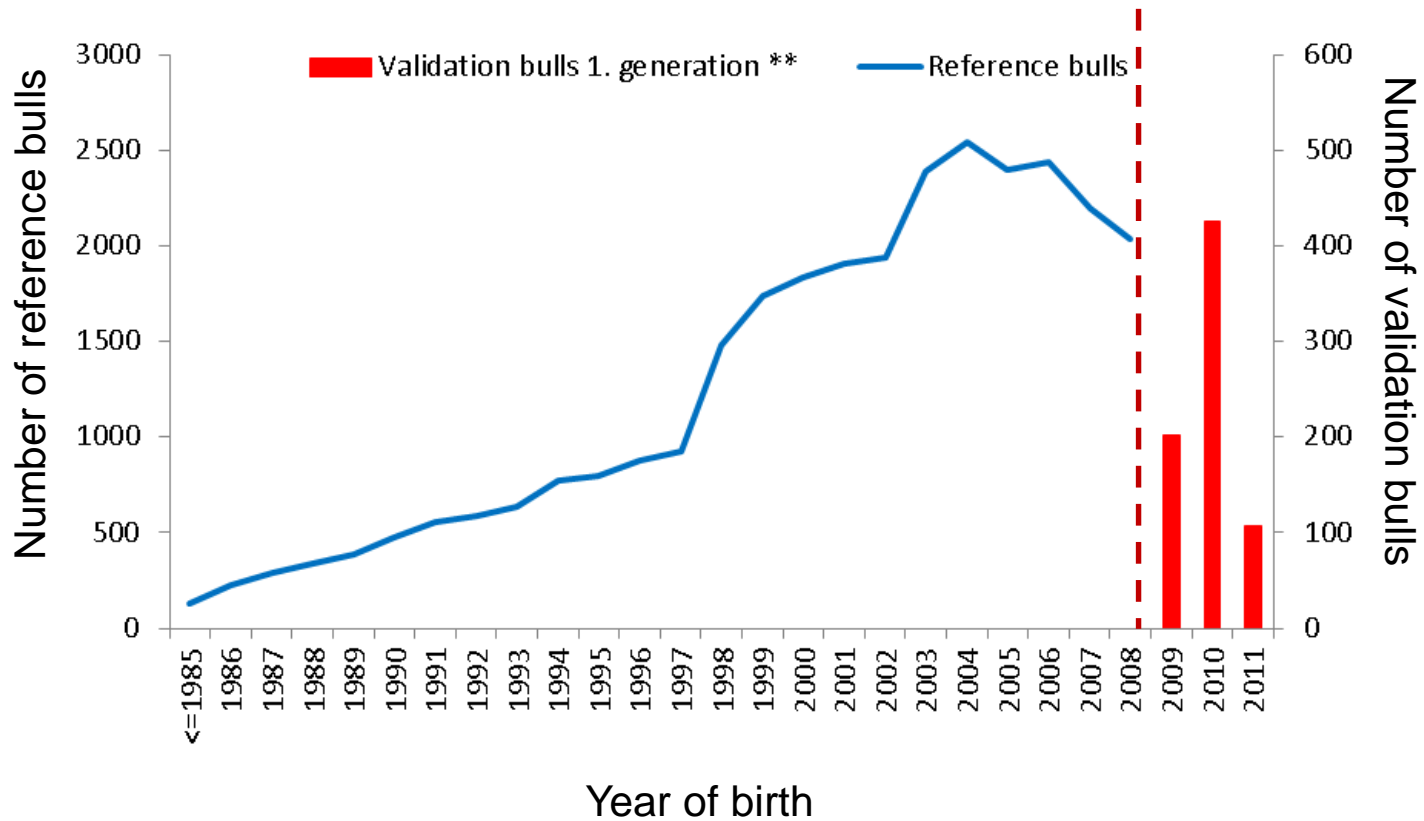
1063 validation bulls stemmed from 178 sires, 137 MGS and 81 PGS.

734 validation bulls with sire born after 2001, 329 validation bulls having sire older than 2003



# A regular genomic validation study for first generation candidates (III)

In total: 33,436 reference bulls  
 29,917 reference bulls  
 734 of 1063 national validation bulls

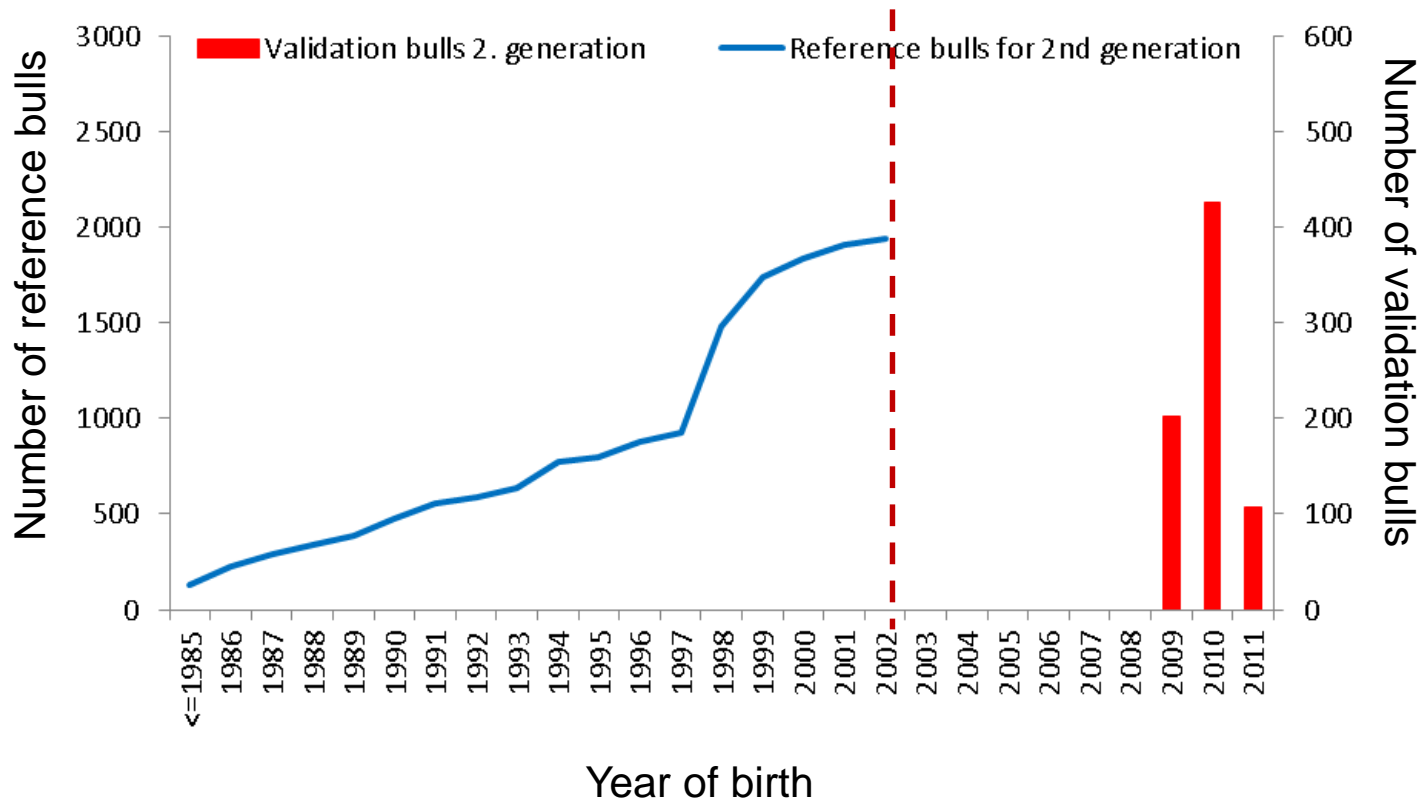


Milk yield, December 2015



# A genomic validation study for second generation candidates

In total: 33,436 reference bulls  
**15,912** reference bulls  
**734** of 1063 national validation bulls



Milk yield, December 2015

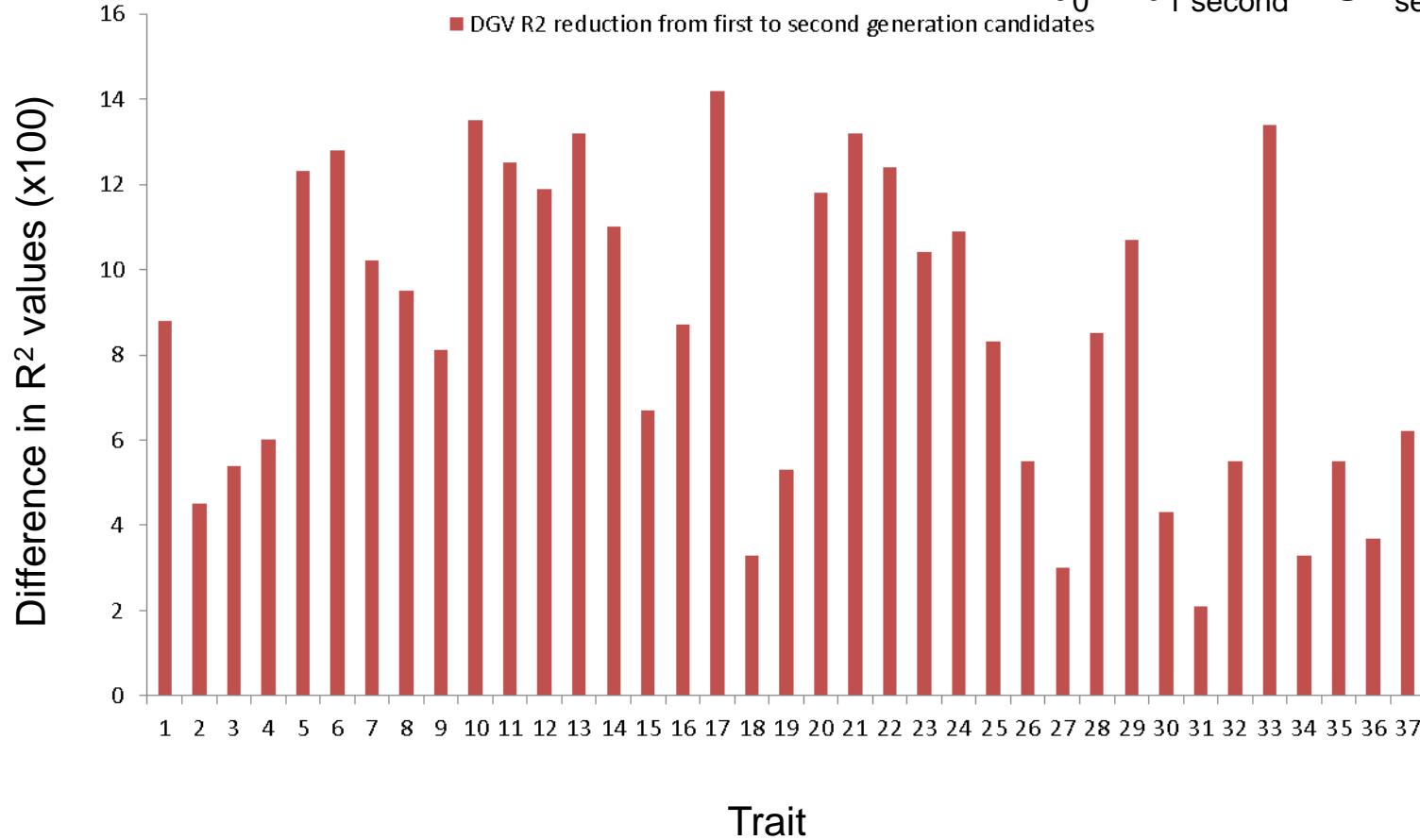


# Results: Validation R<sup>2</sup> reduction from first to second generations (DGV)

$$DRP = b_0 + b_1 \text{ first DGV}_{\text{first}} \quad R^2_{\text{first}}$$

$$DRP = b_0 + b_1 \text{ second DGV}_{\text{second}} \quad R^2_{\text{second}}$$

$R^2_{\text{first}} - R^2_{\text{second}}$



Average R<sup>2</sup> reduction: 0.086

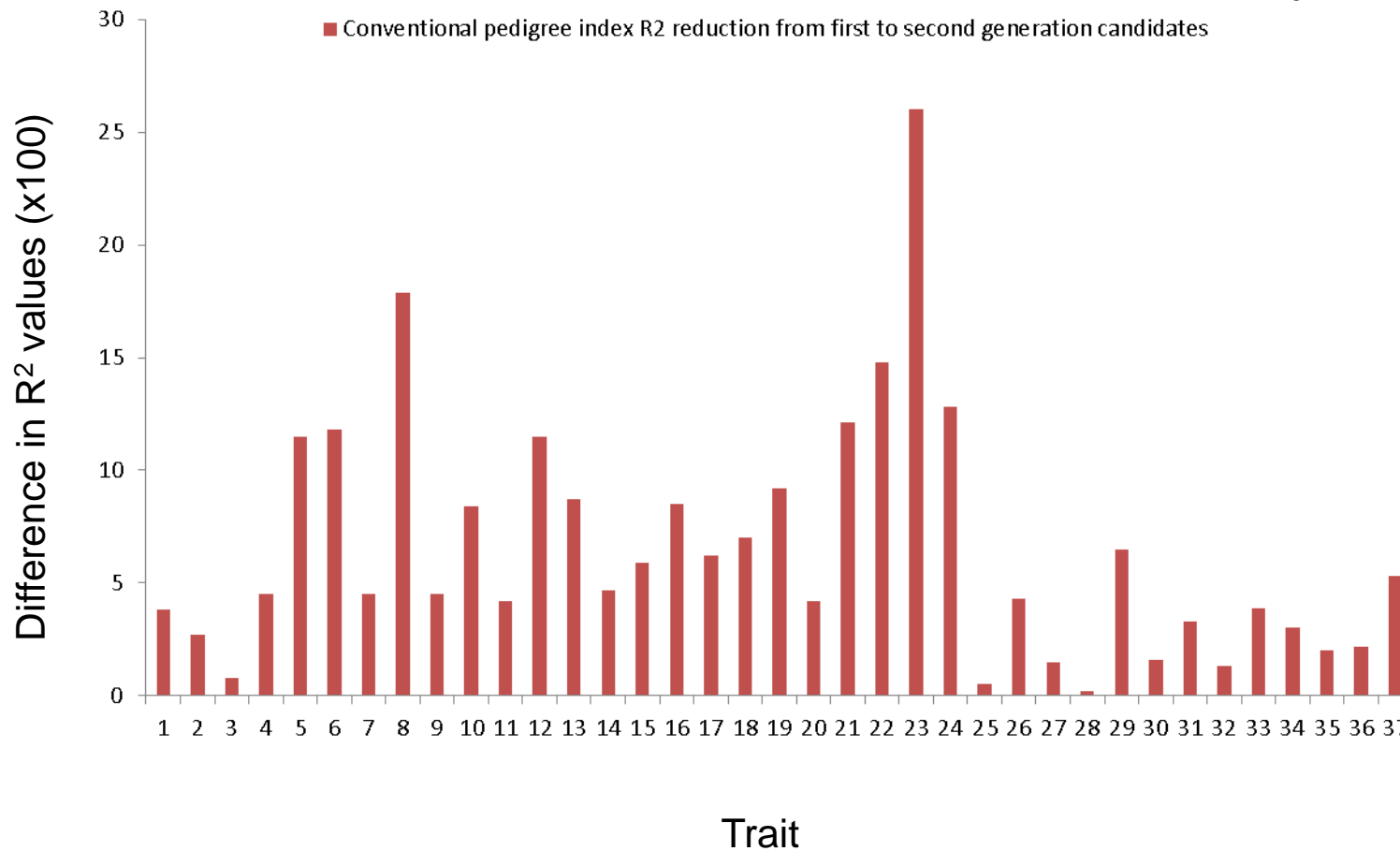




# Results: R<sup>2</sup> reduction from first to second generations (conventional male pedigree index)



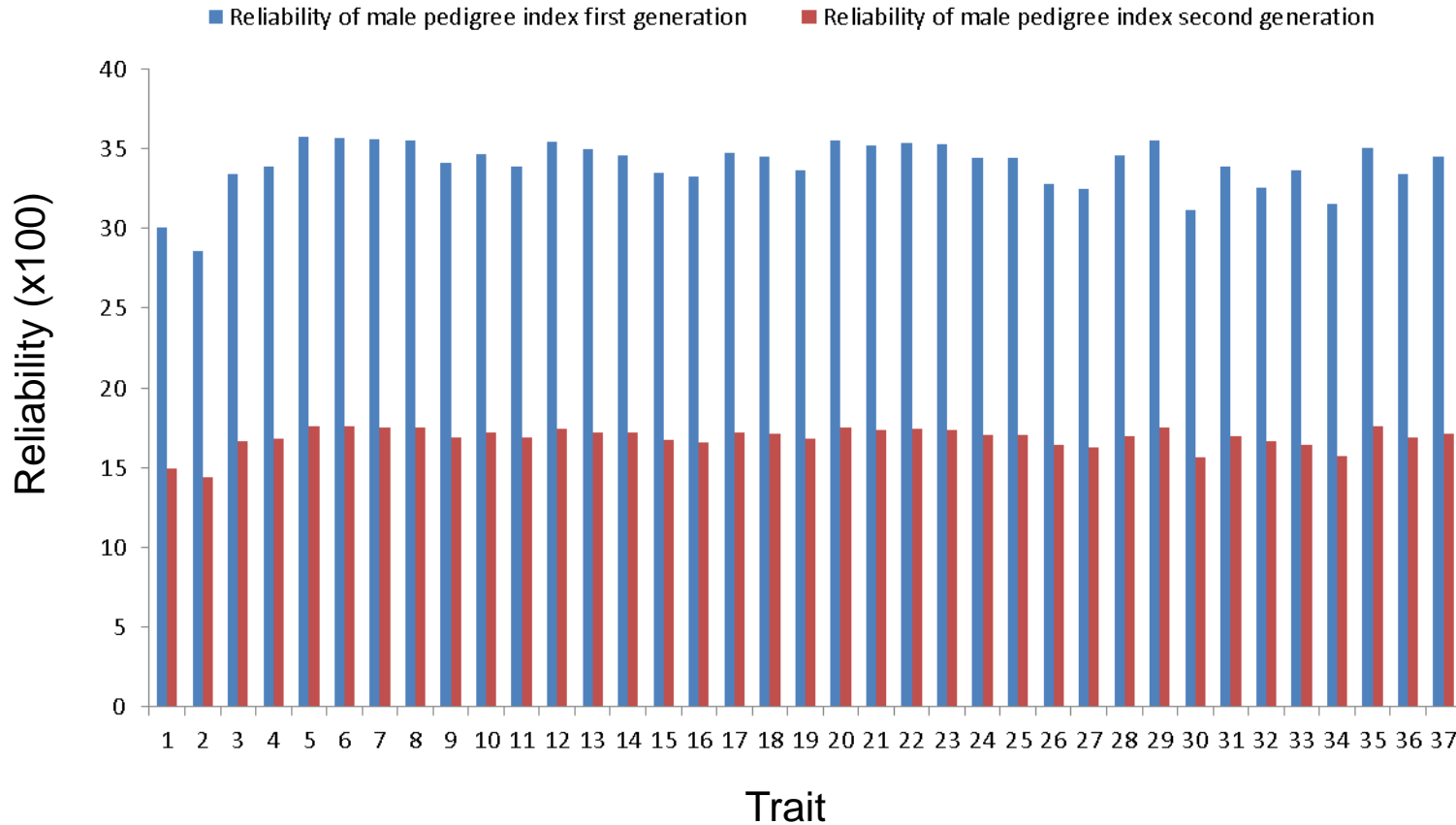
$$\text{DRP} = b_0 + b_1 \text{ mPI}$$



Average R<sup>2</sup> reduction: 0.065



# Results: Model reliabilities of first and second generations (conventional male pedigree index)



Average decrease in reliability of male PI: from 0.34 to 0.17

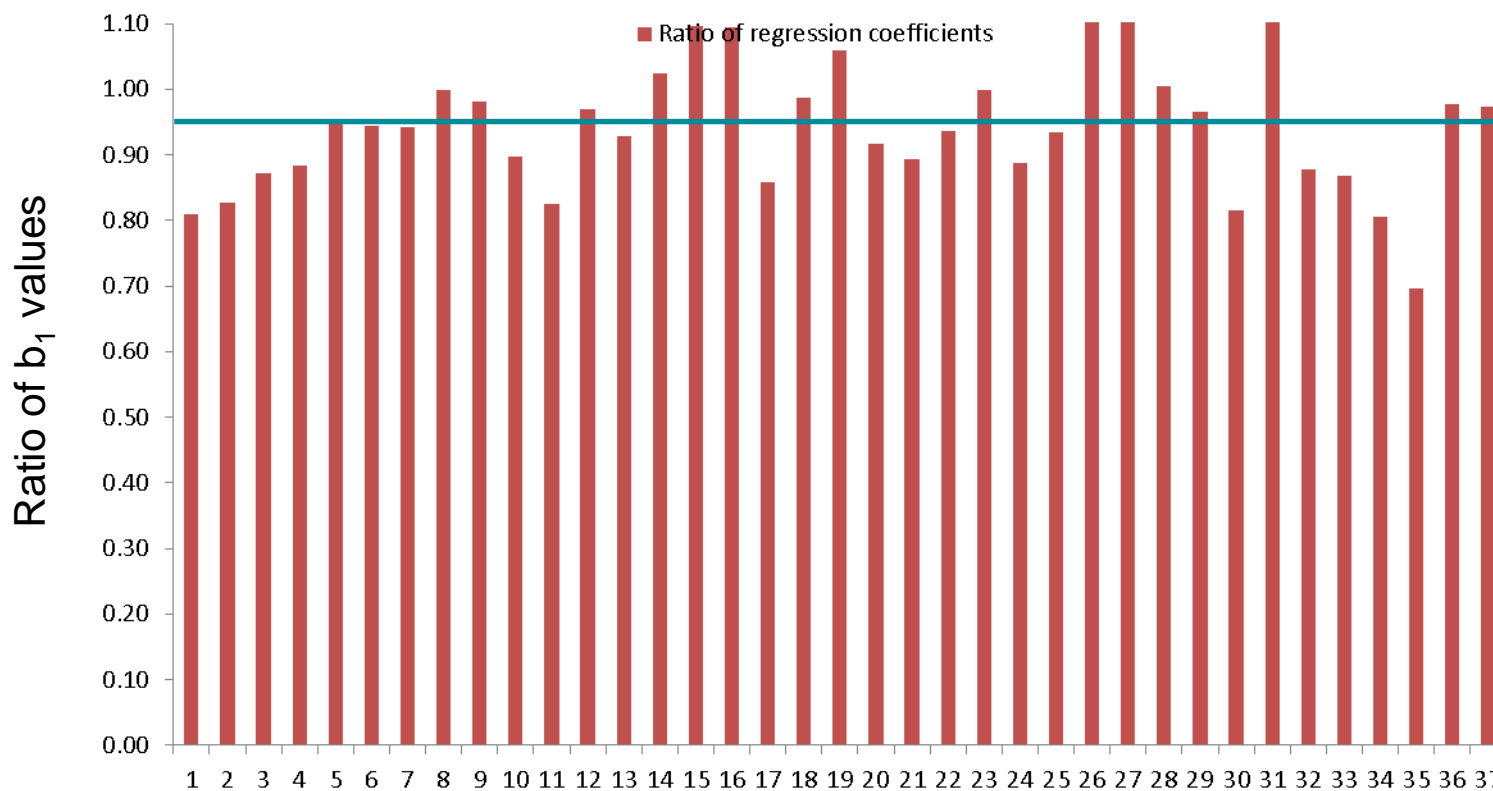


# Results: Bias of genomic prediction: ratio of $b_1$ of second over first generations (DGV)

$$\text{DRP} = b_0 + b_{1 \text{ first}} \text{DGV}_{\text{first}}$$

$$\text{DRP} = b_0 + b_{1 \text{ second}} \text{DGV}_{\text{second}}$$

$$f = \frac{b_{1 \text{ second}}}{b_{1 \text{ first}}}$$



Average ratio of regression coefficients: 0.95

Trait



# Implementation for routine evaluation

- Determining generation # for every genotyped animal, trait by trait
  - Third treated as second generation candidates
- Calculating male pedigree index as before
  - Reduced variance and reliabilities for 2. or 3. generations
- A shrinkage factor for reducing DGV variance of 2. or 3. generation candidates
  - $DGV = \mu + f \sum z_i * a_i$
  - $f=1$  for 1. generation candidates
  - $f = b_{1\text{second}} / b_{1\text{first}}$  for 2. or 3. generation candidates
  - Average  $f=0.95$  corresponding 2.5% less DGV variance caused by LD breakup from 1. to 2. generation
- By treating validation bulls as 1. or 2. generation candidates, all traits (GEBV) passed the GEBVTest
- Introduced for German Holsteins in April 2016



# Summary and Conclusions

- Genomic selection led to almost disappearing of 1. gen. candidates
  - Shortened generation intervals, e.g. sire – son pathway
  - Similar problem also for cow reference population
- Breakup of LD btw SNP markers & genes from 1. to 2. generations
  - Accuracy of genomic prediction drop by  $R^2 = 0.086$  (DGV)
  - Over-prediction bias increased:  $b_1$  ratio = 0.95 (DGV)
- Genomic prediction for 1. and 2. generation candidates
  - Fitting a RPG effect to reduce over-prediction for 1. generation
  - Multiplying a factor  $f$  to DGV to reduce over-prediction for 2. generation
- Extension of the current genomic validation procedure
  - For validating 1. generation candidates, AND
  - For validating 2. (3.) generation candidates



Thanks for your attention!

**vit**



IT-Solutions for Animal Production

# Revisiting the simulation study by Interbull WG Genomic Reliability

- Interbull working group on genomic reliability
  - M. Calus, P. M. VanRaden, M. Lidauer, V. Ducrocq, Z. Liu, B. Harris
  - Results presented by Harris at ITB Workshop in Verden, Feb 2015
- A very realistic simulation study, using QMSim, by Mario Calus
  - 45,000 females and 5000 males per generation
  - 1000 males chosen for mating based on EBV/GEBV (Scenario: SELECTION) or random (Scenario: NO)
  - Heritability = 0.3, bull reliability approximately 0.8 with 45 daughters
  - 9143 QTLs from a Normal distribution and 41,979 SNP markers
- Reference population and generations of candidates
  - 2 generations x 1000 training bulls (selection ratio of 1/5)
    - DYD, TBV, genotypes and pedigree available
  - 3 generations x 2000 male candidates (selection ratio of 2/5)
    - TBV, genotypes and pedigree available



# Interbull WG Genomic Reliability: Results of the scenario SELECTION

- GBLUP with 20% residual polygenic variance (P. M. VanRaden)
- Genomic validation:  $TBV = a + b \cdot GEBV$

Generation # of candidates	Validation $R^2$	Regression	$\frac{\sigma_{GEBV_j}}{\sigma_{GEBV_I}}$
1	0.30	0.96	1
2	0.26	0.94	0.95
3	0.22	0.92	0.89

- Assuming all generations have equal TBV variances
- If TBV/DYD variance decreases, the ratios become even smaller
- Higher selection intensity can lead to even smaller ratios





# Acknowledgements

- Interbull Working Group Genomic Reliability
  - Bevin Harris, Mario Calus, Paul VanRaden, Martin Lidauer, Vincent Ducrocq

