Methods for prediction of breeding values for young animals in single-step evaluations

Eduardo Pimentel, C. Edel, R. Emmerling, K.-U. Götz

Institute of Animal Breeding

Motivation

❑ Single-step model is computationally very demanding

LfL
Animal Breeding

Motivation

❑ Single-step model is computationally very demanding

❑ **ssGEBV of young animals should be rapidly available**

LfL
Animal Breeding

Motivation

❑ Single-step model is computationally very demanding

❑ **ssGEBV of young animals should be rapidly available**

➔ With increasing number of genotypes setting up the full model every time may become impracticable

LfL
Animal Breeding

Motivation

❑ Single-step model is computationally very demanding

❑ ssGEBV of young animals should be rapidly available

➔ With increasing number of genotypes setting up the full model every time may become impracticable

➔ Need for efficient and cost-effective alternative with negligible loss in accuracy

LfL
Animal Breeding

Objective

❑ Single-step model is computationally very demanding

❑ ssGEBV of young animals should be rapidly available

➔ With increasing number of genotypes setting up the full model every time may become impracticable

➔ Need for efficient and cost-effective alternative with negligible loss in accuracy

❑ The objective here was to develop such an alternative

LfL
Animal Breeding

Material and Methods

❑ Fleckvieh data from April 2017 (here the "previous run")

➔ Single-step univariate runs for 23 conformation traits

➔ MiX99 with input of $[\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}]$

➔ 3,172,636 animals and around 1.4 million records

➔ 62,559 genotyped animals in $\mathbf{G} = 0.8\mathbf{G}_0 + 0.2\mathbf{A}_{22}$

❑ 1,768 further young animals for prediction and validation

LfL
Animal Breeding

Material and Methods

- [ ] Validation through correlation with ssGEBV of <u>full model</u>:

  ➜ 62,559 + 1,768 = <u>64,327</u> genotyped animals in $\mathbf{G}$

$$\mathbf{G} = 0.8\mathbf{G}_0 + 0.2\mathbf{A}_{22}$$

- [ ] <u>Alternative models</u>:

  ➜ Data from previous run (62,559 genotyped animals)

  ➜ Prediction of ssGEBV for the 1,768 young animals

LfL *Animal Breeding*

Material and Methods

□ <u>G model:</u>

$$\hat{\mathbf{a}}_y = \mathbf{C}\mathbf{G}^{-1}\hat{\mathbf{a}}_2$$

$(62{,}559 \times 1)$

$(1{,}768 \times 1)$

ssGEBV of the young animals

ssGEBV of the genotyped animals from previous run

LfL
Animal Breeding

❑ <u>G model:</u>

$$\hat{\mathbf{a}}_y = \mathbf{C}\mathbf{G}^{-1}\hat{\mathbf{a}}_2$$

$(62,559 \times 1)$

$(1,768 \times 1)$

ssGEBV of the young animals

ssGEBV of the genotyped animals from previous run

$(62,559 \times 62,559)$

$\mathbf{G}^{-1}$ between genotyped animals from previous run

LfL
Animal Breeding

Material and Methods

❑ <u>G model:</u>

$$\hat{\mathbf{a}}_y = \mathbf{CG}^{-1}\hat{\mathbf{a}}_2$$

$(62{,}559 \times 1)$

ssGEBV of the genotyped animals from previous run

$(1{,}768 \times 1)$

ssGEBV of the young animals

$(62{,}559 \times 62{,}559)$

$\mathbf{G}^{-1}$ between genotyped animals from previous run

$(1{,}768 \times 62{,}559)$

Genomic relationships between young animals and genotyped animals from previous run

LfL
Animal Breeding

Material and Methods

☐ <u>G model:</u>

$$\hat{\mathbf{a}}_y = \mathbf{C}\mathbf{G}^{-1}\hat{\mathbf{a}}_2$$

$(62{,}559 \times 1)$

ssGEBV of the genotyped animals from previous run

$(1{,}768 \times 1)$

ssGEBV of the young animals

$(62{,}559 \times 62{,}559)$

$\mathbf{G}^{-1}$ between genotyped animals from previous run

$(1{,}768 \times 62{,}559)$

Genomic relationships between young animals and genotyped animals from previous run

Both $\mathbf{C}$ and $\mathbf{G}$ with $0.2\mathbf{A}$

LfL
Animal Breeding

Material and Methods

□ SNP model:

$$\begin{bmatrix} \mathbf{1'1} & \mathbf{1'M} \\ \mathbf{M'1} & \mathbf{M'M} + \mathbf{I}k \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1'\hat{a}}_2 \\ \mathbf{M'\hat{a}}_2 \end{bmatrix} \qquad k = \frac{\sigma_e^2}{\sigma_g^2}$$

$(62{,}559 \times m)$

centered genotypes of the animals from previous run

$(m \times 1)$

estimated SNP effects

LfL
Animal Breeding

Material and Methods

□ <u>SNP model:</u>

$$\begin{bmatrix} \mathbf{1'1} & \mathbf{1'M} \\ \mathbf{M'1} & \mathbf{M'M} + \mathbf{I}k \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1'\hat{a}}_2 \\ \mathbf{M'\hat{a}}_2 \end{bmatrix} \qquad k = \frac{\sigma_e^2}{\sigma_g^2}$$

$$k = \frac{\sigma_e^2}{\sigma_g^2} = \frac{0.2\sigma_a^2}{0.8\sigma_a^2\gamma} \qquad \gamma = \frac{1}{\sum 2p_i(1-p_i)}$$

$$k = 0.25 \sum 2p_i(1-p_i)$$

LfL
Animal Breeding

❑ <u>SNP model:</u>

$$\begin{bmatrix} \mathbf{1'1} & \mathbf{1'M} \\ \mathbf{M'1} & \mathbf{M'M} + \mathbf{I}k \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1'\hat{a}}_2 \\ \mathbf{M'\hat{a}}_2 \end{bmatrix} \qquad k = \frac{\sigma_e^2}{\sigma_g^2}$$

$$\hat{\mathbf{e}} = \hat{\mathbf{a}}_2 - (\hat{\mu} + \mathbf{M}\hat{\mathbf{g}})$$

$$\hat{\mathbf{e}}_y = \mathbf{A}_{y2}\mathbf{A}_{22}^{-1}\hat{\mathbf{e}}$$

$$\hat{\mathbf{a}}_y = (\hat{\mu} + \mathbf{M}_y\hat{\mathbf{g}}) + \hat{\mathbf{e}}_y$$

LfL
Animal Breeding

Results

	Min	1Q.	Mean	3Q.	Max
G model	0.9851	0.9916	**0.9927**	0.9941	0.9949
SNP model	0.9744	0.9888	**0.9896**	0.9920	0.9929

LfL
Animal Breeding

Results

All young animals

N = 1768
r = 0.9842

Full model (y-axis)
SNP model (x-axis)

❑ Illustration with rump angle
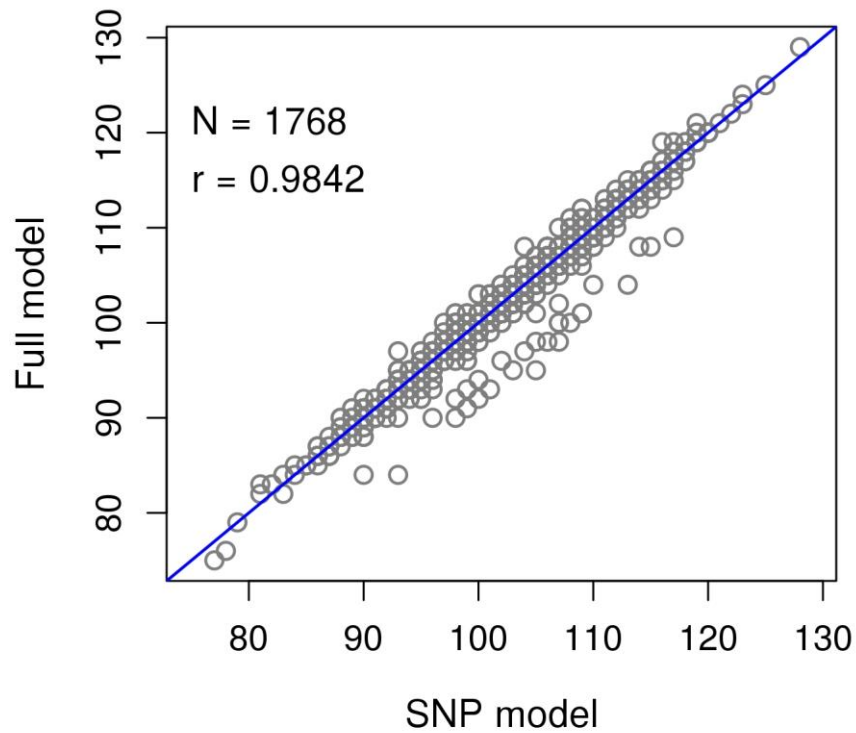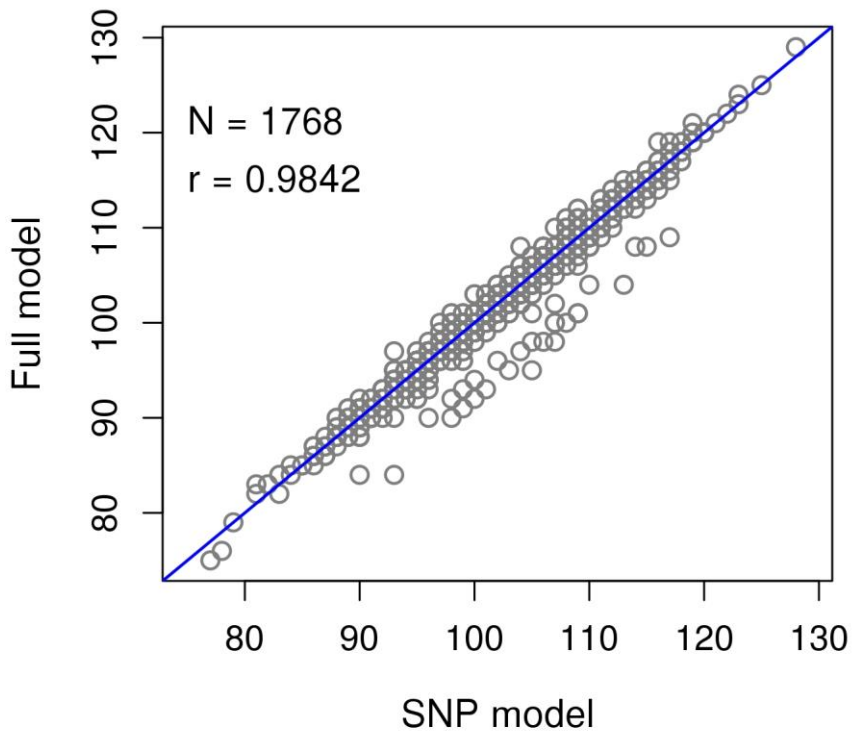
LfL
Animal Breeding

Results

All young animals

N = 1768
r = 0.9842

- ❑ Closer look at different groups of young animals

- ❑ Parents in the previous run? If yes, with phenotype / genotype?

Results

All young animals
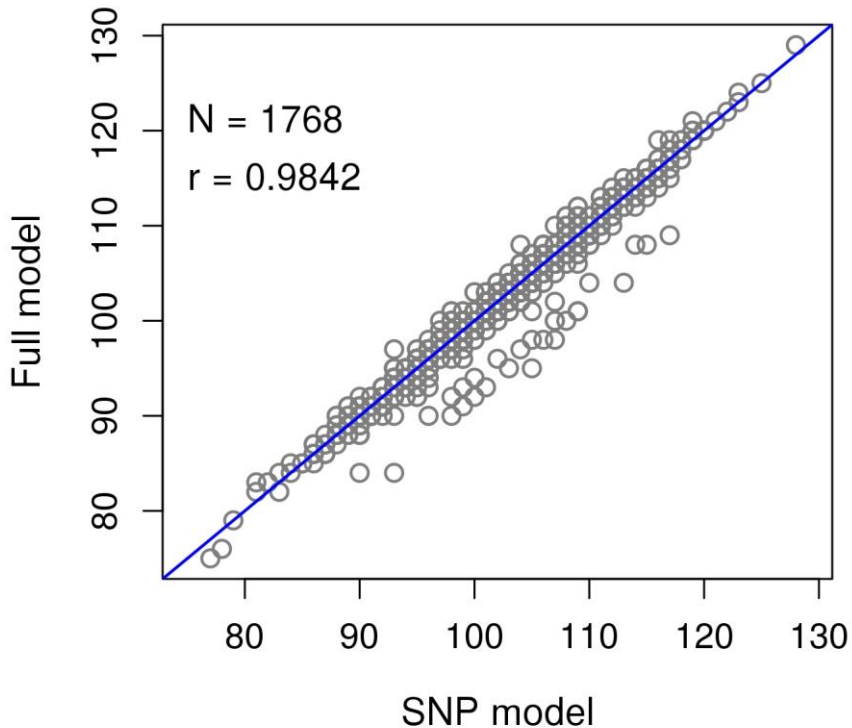
N = 1768
r = 0.9842

Full model (y-axis)
SNP model (x-axis)

- Closer look at different groups of young animals

- Parents in the previous run? If yes, with phenotype / genotype?

- Sires of young animals are usually represented in $\hat{a}_2$

- Information from the male side is usually well transmitted through SNP effects

LfL Animal Breeding
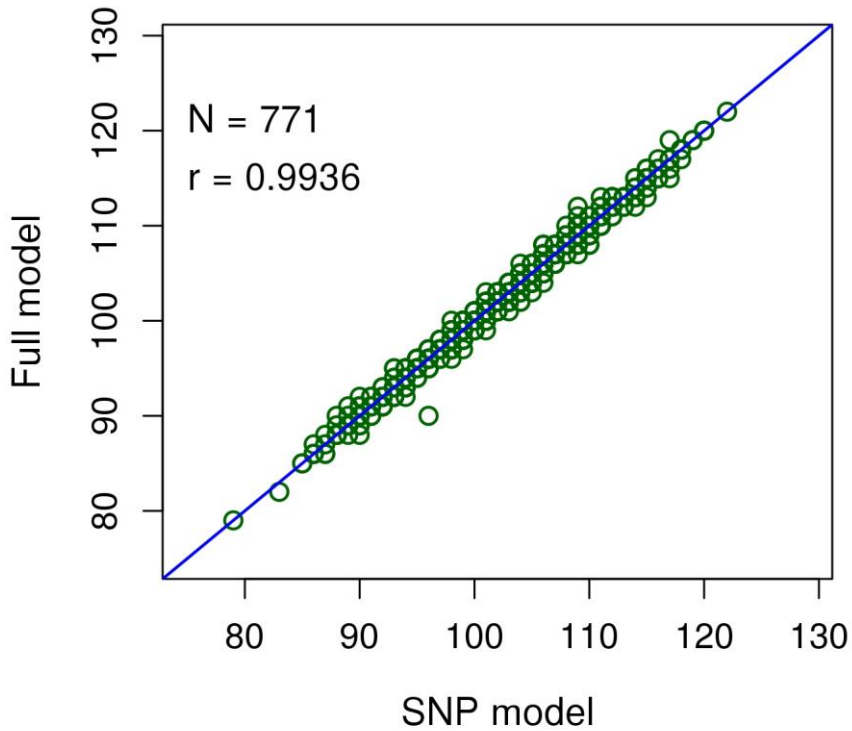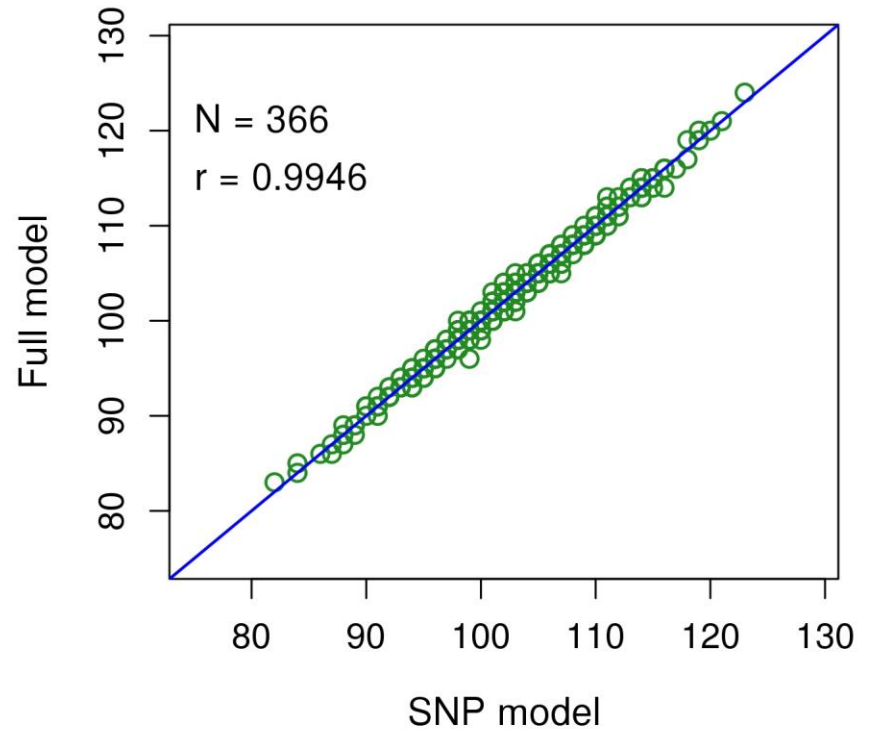
Results

Results

Dam has genotype but no phenotype
N = 193
r = 0.996
(x-axis: SNP model, y-axis: Full model)
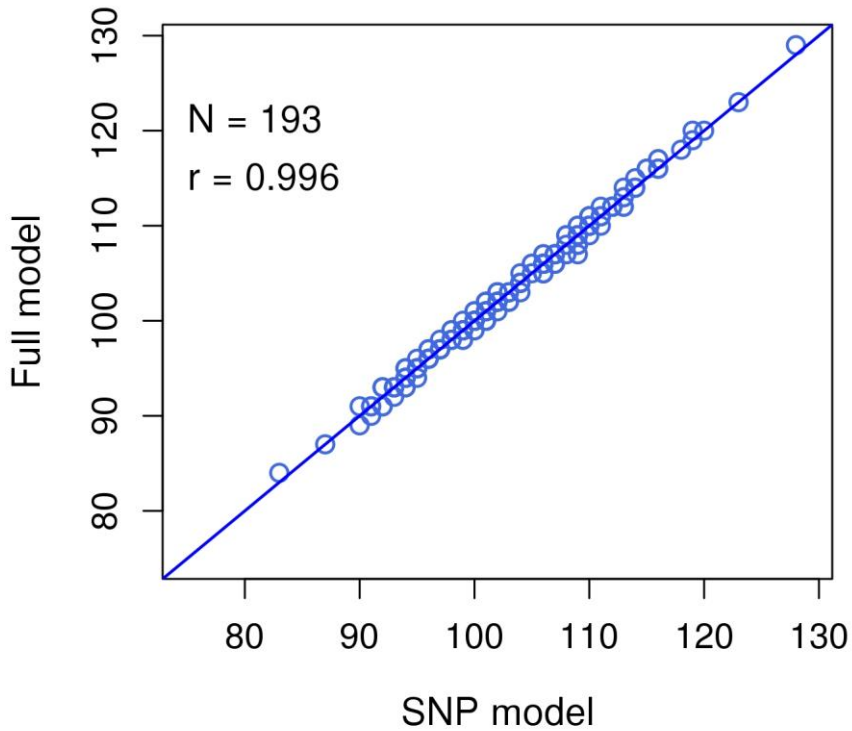
Dam has genotype and phenotype
N = 55
r = 0.997
(x-axis: SNP model, y-axis: Full model)
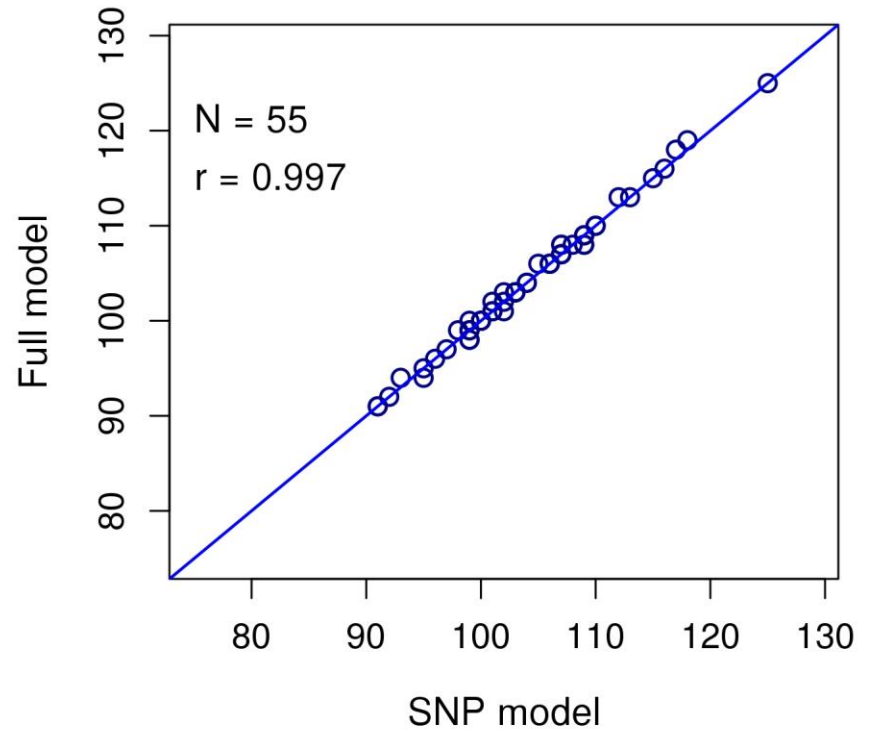
LfL
Animal Breeding

Results

Dam has phenotype but no genotype

N = 342
r = 0.9817

Full model

SNP model

Missing dam

N = 41
r = 0.8438

Full model

SNP model

LfL
Animal Breeding

Results

Dam has phenotype but no genotype

N = 342
r = 0.9817

Full model (y-axis): 80, 90, 100, 110, 120, 130
SNP model (x-axis): 80, 90, 100, 110, 120, 130
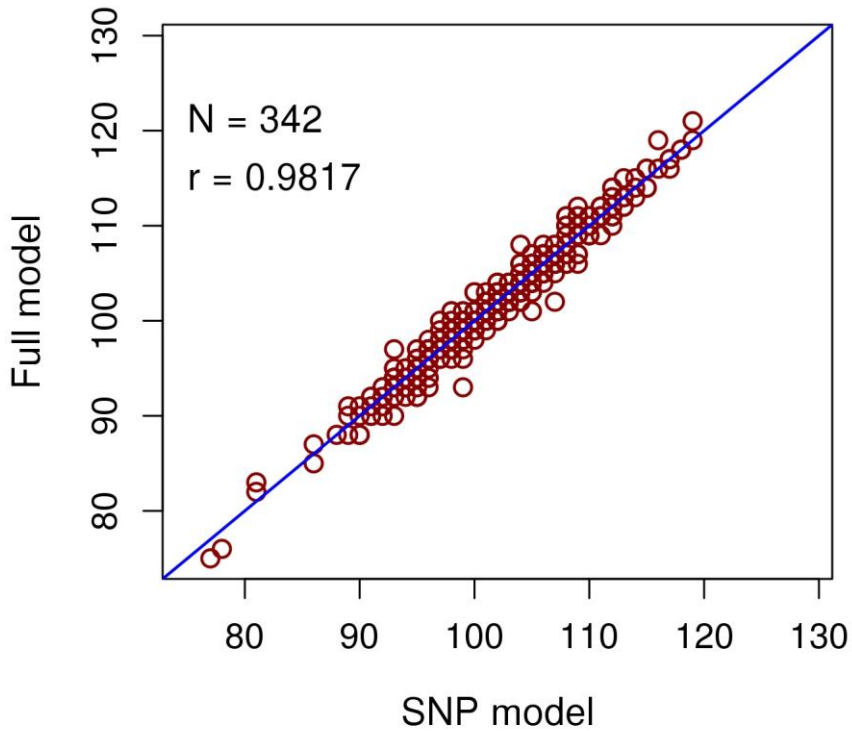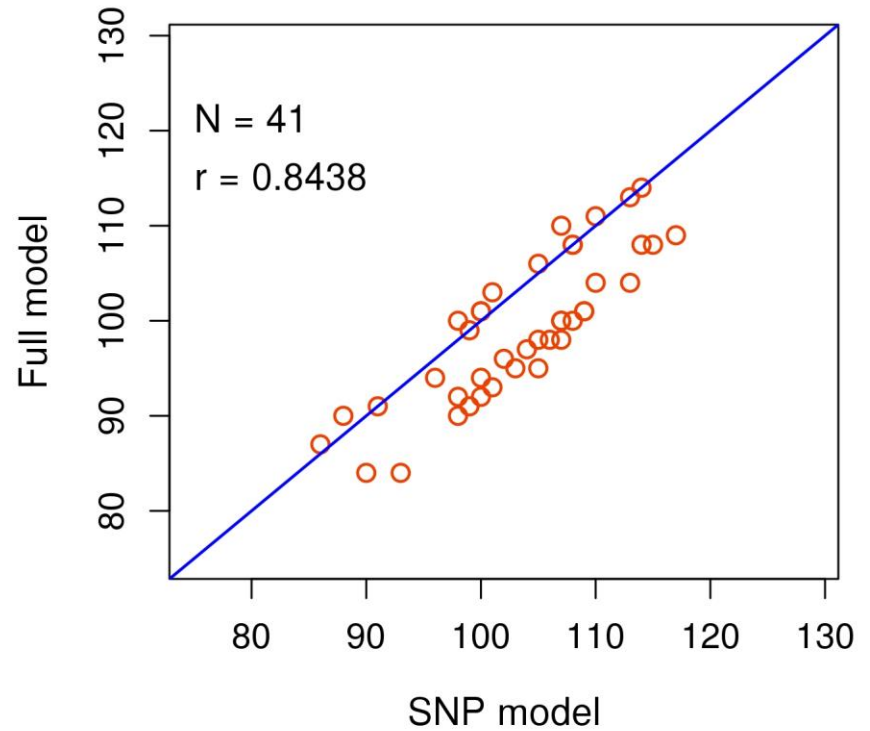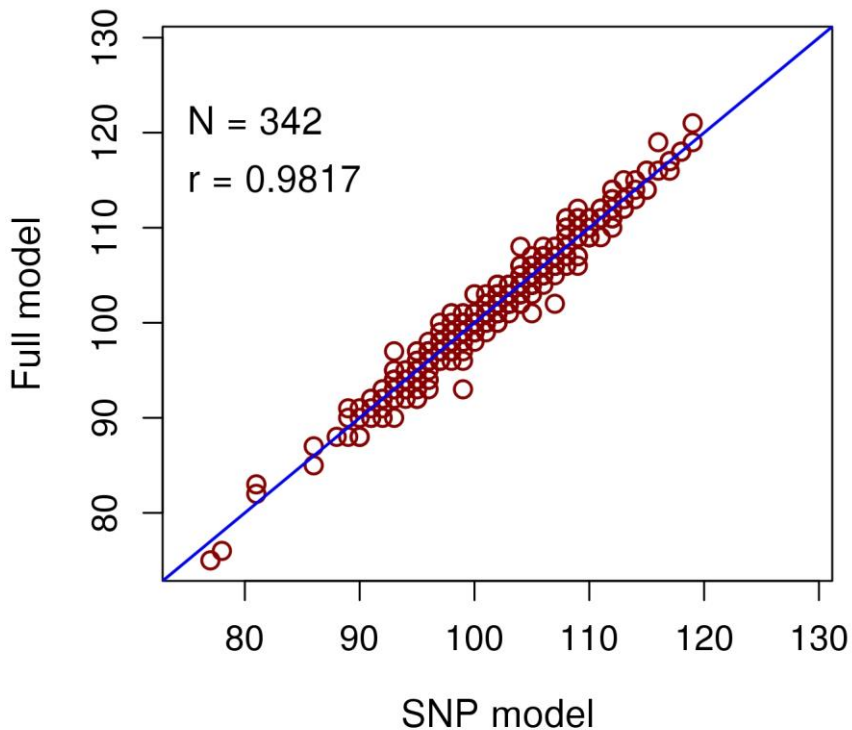
❑ Transmission of dam information through MGS only

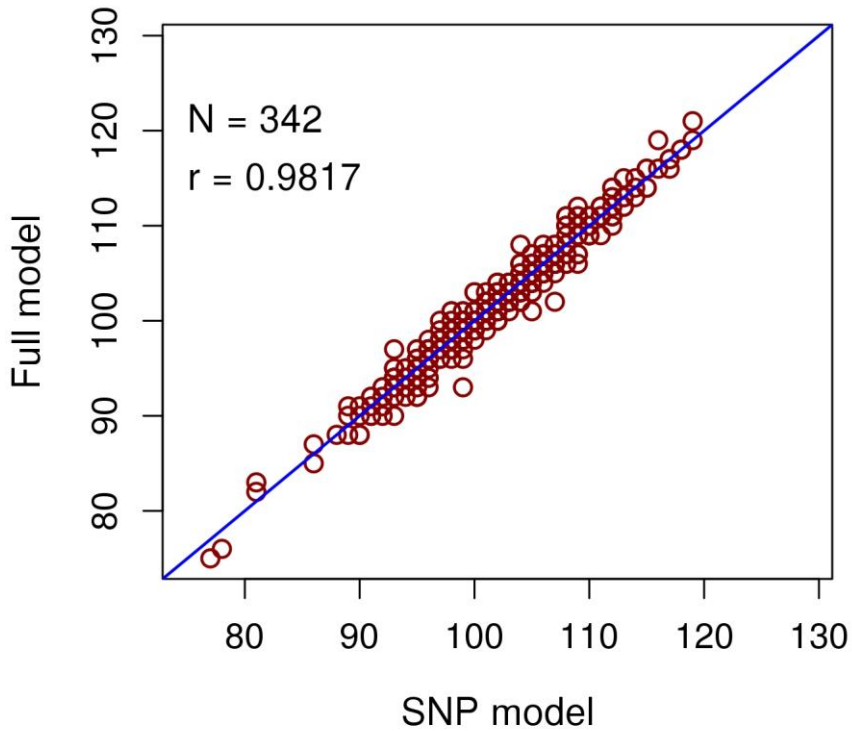❑ Mendelian sampling information of dam not contained in $\hat{\mathbf{a}}_2$
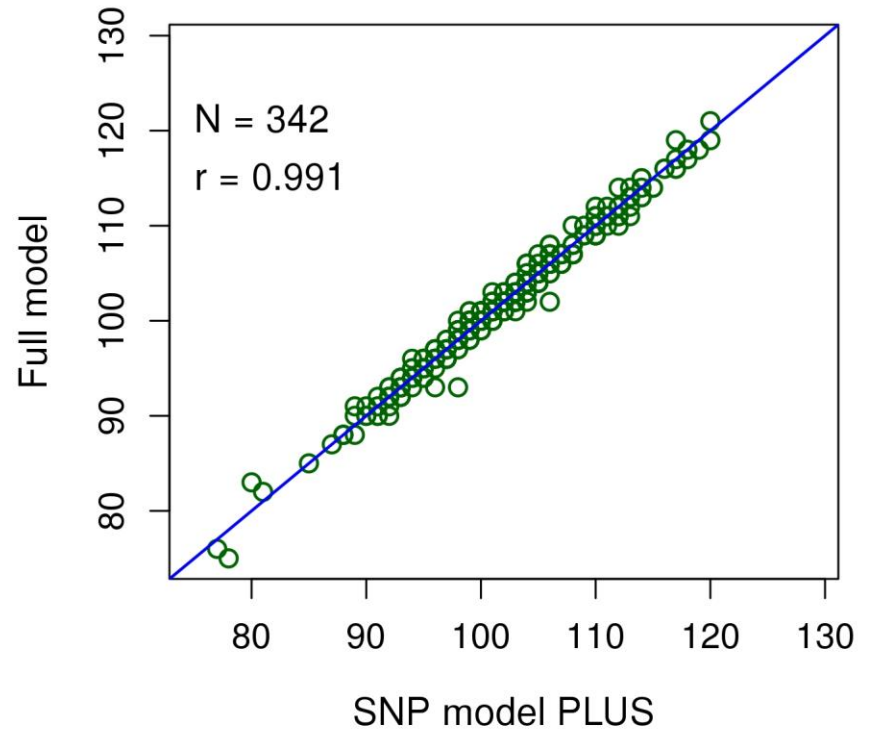
❑ Add to ssGEBV of young animal:

$$0.5\hat{\mathbf{m}}(1 - \hat{r}^2_{imp})$$

LfL
Animal Breeding

Results

Dam has phenotype but no genotype

N = 342
r = 0.9817

Dam has phenotype but no genotype

N = 342
r = 0.991

Results

Missing dam

N = 41
r = 0.8438

Full model (y-axis: 80, 90, 100, 110, 120, 130)
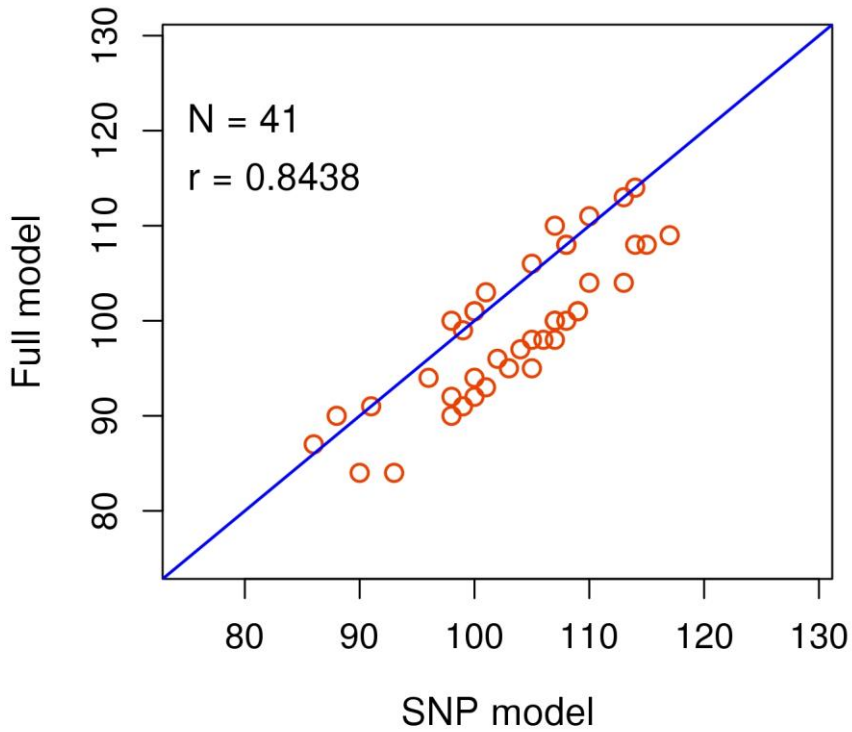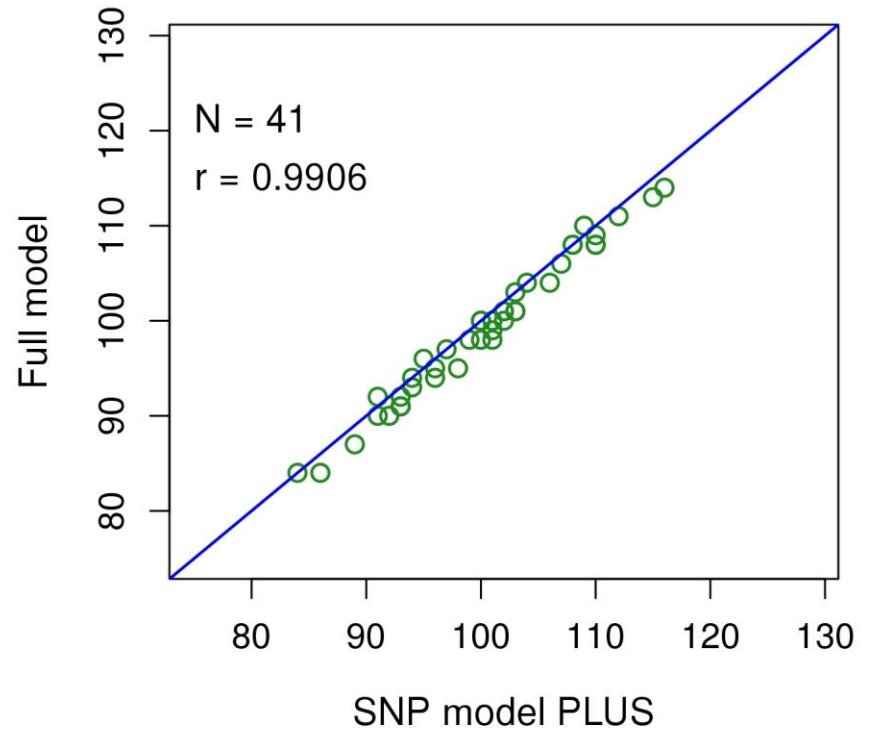
SNP model (x-axis: 80, 90, 100, 110, 120, 130)
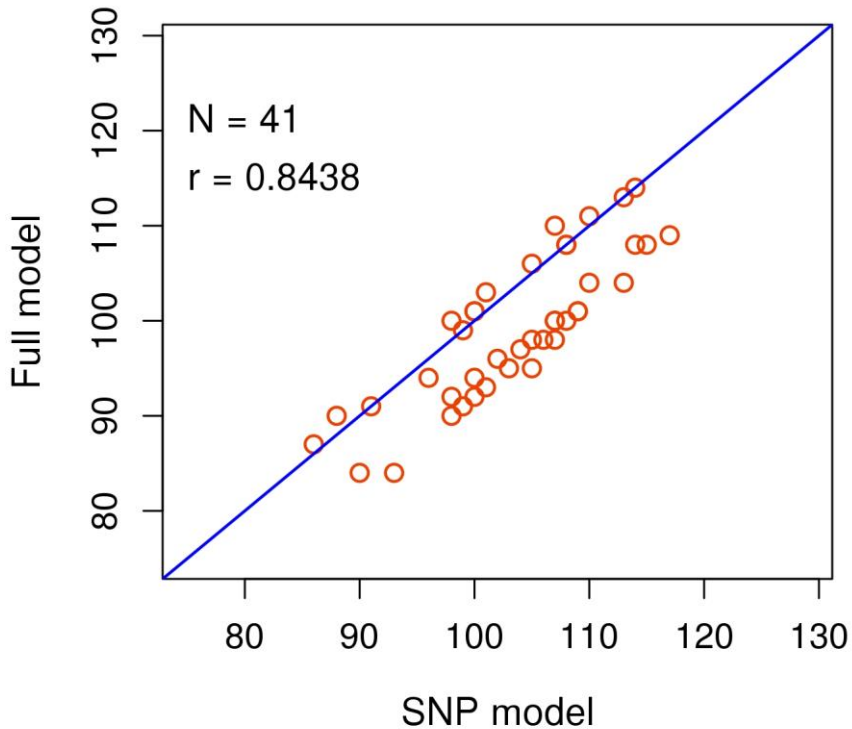
- ❑ Missing dam replaced by genetic group in the full model

- ❑ Genetic group information not contained in $\hat{\mathbf{a}}_2$

- ❑ Add half of the group solution from previous run to ssGEBV of young animal

LfL
Animal Breeding
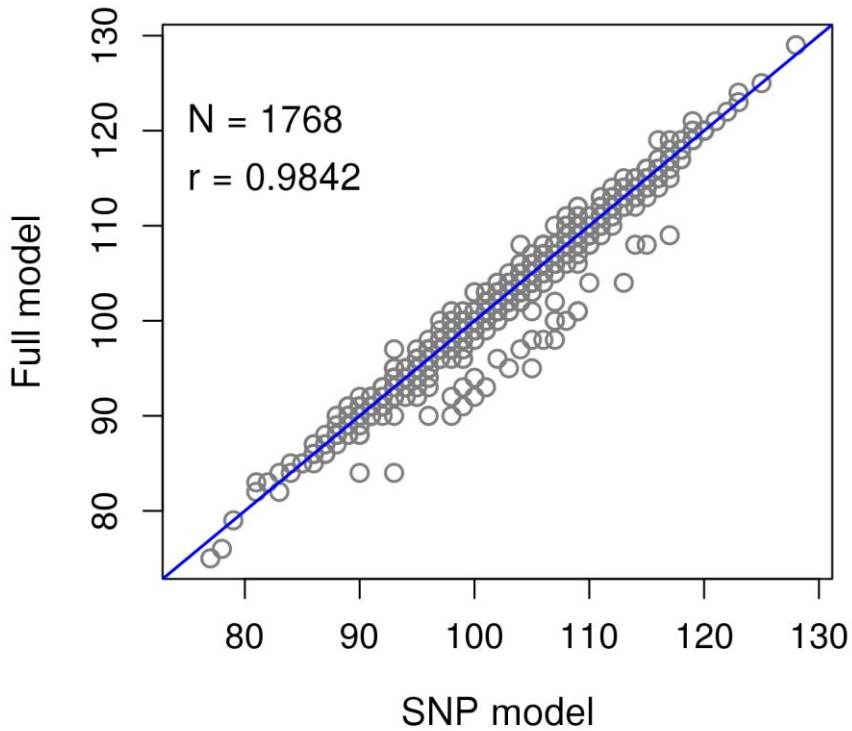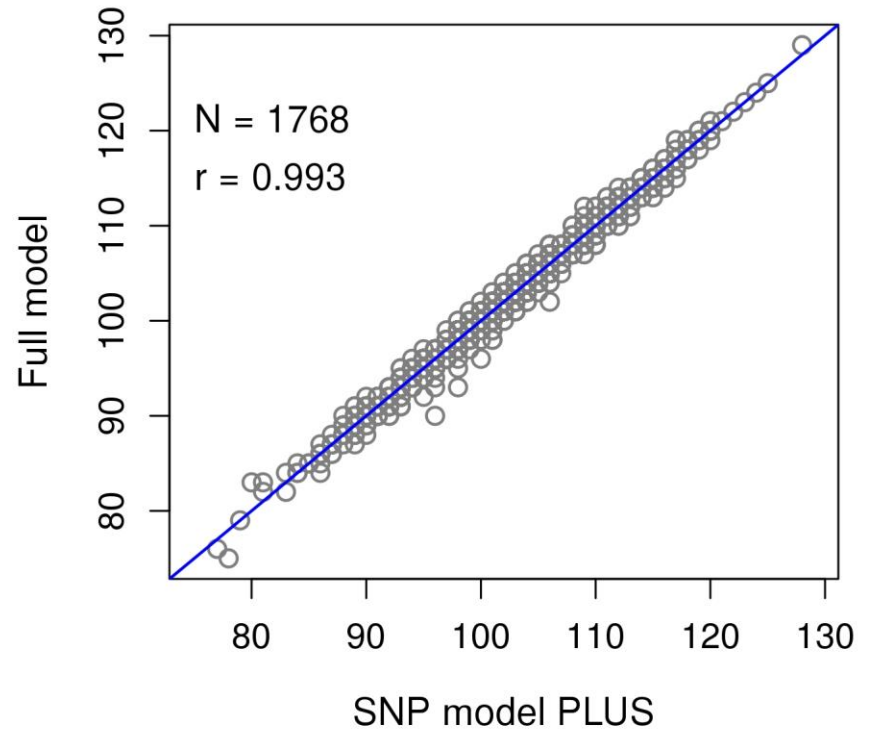
Results

Results

Results

	Min	1Q.	Mean	3Q.	Max
G model	0.9851	0.9916	**0.9927**	0.9941	0.9949
SNP model	0.9744	0.9888	**0.9896**	0.9920	0.9929
SNP model PLUS	0.9897	0.9927	**0.9932**	0.9941	0.9949

Results

$\mathbf{G}^{-1}$

	Min	1Q.	Mean	3Q.	Max
G model	0.9851	0.9916	**0.9927**	0.9941	0.9949
SNP model	0.9744	0.9888	**0.9896**	0.9920	0.9929
SNP model PLUS	0.9897	0.9927	**0.9932**	0.9941	0.9949

$\mathbf{M'M}$

$(m \times m)$

Results

$\mathbf{G}^{-1}$

	Min	1Q.	Mean	3Q.	Max
G model	0.9851	0.9916	**0.9927**	0.9941	0.9949
SNP model	0.9744	0.9888	**0.9896**	0.9920	0.9929
SNP model PLUS	0.9897	0.9927	**0.9932**	0.9941	0.9949

$$\hat{\mathbf{e}}_y = \mathbf{A}_{y2}\mathbf{A}_{22}^{-1}\hat{\mathbf{e}}$$

$\mathbf{M'M}$

$(m \times m)$

Results

$\mathbf{G}^{-1}$

	Min	1Q.	Mean	3Q.	Max
G model	0.9851	0.9916	**0.9927**	0.9941	0.9949
SNP model	0.9744	0.9888	**0.9896**	0.9920	0.9929
SNP model PLUS	0.9897	0.9927	**0.9932**	0.9941	0.9949

$$\hat{\mathbf{e}}_y = \mathbf{A}_{y2}\mathbf{A}_{22}^{-1}\hat{\mathbf{e}}$$

$\mathbf{M'M}$

$(m \times m)$

or simply PAs from $\hat{\mathbf{e}}$

LfL
Animal Breeding

Results

$\mathbf{G^{-1}}$

	Min	1Q.	Mean	3Q.	Max
G model	0.9851	0.9916	**0.9927**	0.9941	0.9949
SNP model	0.9744	0.9888	**0.9896**	0.9920	0.9929
SNP model PLUS	0.9897	0.9927	**0.9932**	0.9941	0.9949

$\mathbf{M'M}$

$(m \times m)$

- ❑ Save only $\hat{\mathbf{g}}$ and $\hat{\mathbf{e}}$

- ❑ Prediction even on a daily basis

- ❑ Suitable for a database system

LfL
Animal Breeding

Conclusions

❑ Prediction of single-step breeding values for young animals can be well approximated using SNP effects estimated from systems of size equal to the number of markers (i.e., without setting up $\mathbf{G}$)

❑ Especially meaningful for single-step implementations without the explicit computation of the inverse of the genomic relationship matrix (e.g., ssGTBLUP)

Thanks!

100 years Animal Breeding in Grub
1918 - 2018

LfL
Animal Breeding

Trait	Abr.	G-model	SNP-model	SNP-model PLUS
Rump angle	BN	0.990	0.984	0.993
Rump length	BL	0.994	0.992	0.994
Muscularity	BM	0.985	0.974	0.991
Udder depth	EB	0.995	0.993	0.994
Udder	EU	0.991	0.987	0.992
Fetlock	FE	0.992	0.989	0.993
Feet & Legs	FU	0.995	0.993	0.994
Width	HB	0.994	0.992	0.994
Height at cross	KH	0.991	0.989	0.993
Front teat placement	SP	0.995	0.993	0.994
Frame	RA	0.991	0.989	0.992
Udder cleanliness	ER	0.992	0.989	0.991
Rump depth	RT	0.994	0.991	0.993
Rear udder length	SEL	0.992	0.989	0.990
Hocks	SPA	0.991	0.987	0.993
Hock angularity	SPW	0.995	0.993	0.995
Teat thickness	SD	0.994	0.992	0.995
Teat length	SL	0.993	0.991	0.994
Teat placement	SS	0.994	0.992	0.994
Foot angle	TR	0.992	0.988	0.994
Fore udder length	VEL	0.994	0.990	0.993
Fore udder attachment	VA	0.994	0.991	0.994
Suspensory ligament	ZB	0.995	0.992	0.994
Mean		**0.9927**	**0.9896**	**0.9932**