How to blend DGVs and EBVs

Esa Mäntysaari, MTT

Gert Pedersen Aamand, NAV

Acknowledge

NAV Implementation group

Guosheng Su

Ole Fredslund Christensen

Rasmus Froberg Brøndum

Ulrik Sander Nielsen

Ismo Strandén

Esa Mäntysaari

Gert Pedersen Aamand

Nordisk Avlsværdi Vurdering •

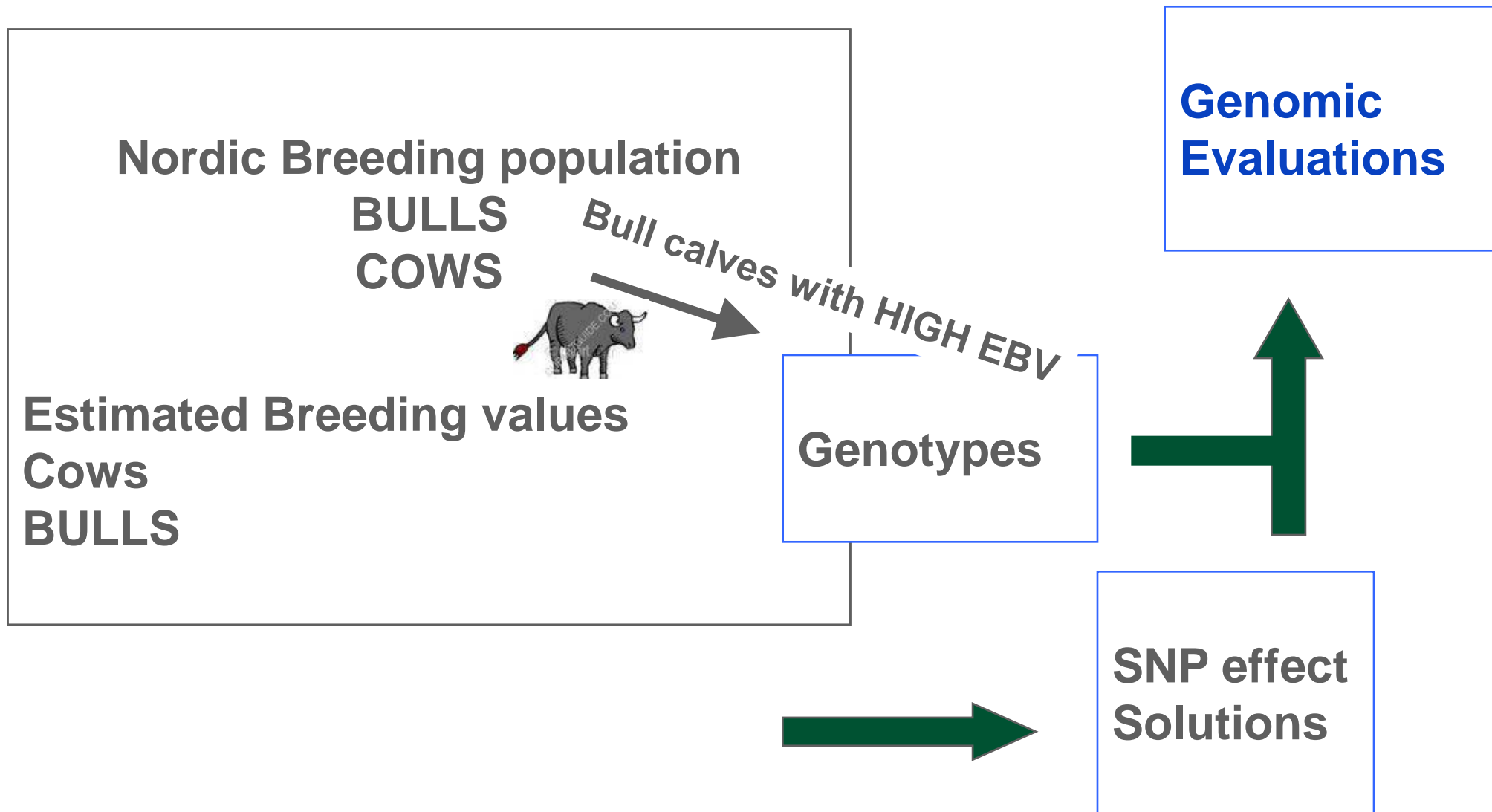# Review: What this far?

Nordic Breeding population
BULLS
COWS

Estimated Breeding values
Cows
BULLS

"Reference population"

SNP effect
Solutions

Review: What this far available

Nordic Breeding population
   BULLS
   COWS

Bull calves with HIGH EBV

Genomic Evaluations

Estimated Breeding values
Cows
BULLS

Genotypes

SNP effect Solutions

Review: Genomic evaluations

- **Genomic evaluations for top animals**
  - **Bull dams and bull calves**
  - **Genotypings selected based on EBV (NTM)**

- **Bull dam's own production has no effect on it's genomic index**
  - **Neither it's sons genomic index**
  - *Concept:*

    Genotype x SNP solutions = breeding value

Abbreviations

- **BV = Breeding value**
  - **True breeding value**

- **EBV = (Nordic) breeding value estimate**
  - **Estimated breeding value**

- **DGV (Genomic value based on SNP effects)**
  - **Direct Genomic Value**

Reliability of Genomic evaluations

Trait	Reliability, Holstein		
	Pedigree index EBV	DGV	Progeny test EBV
Longevity	0.17 – 0.21		0.50
Fertility	0.20 – 0.22		0.60
Calving difficulty	0.18 - 0.26	Reliability 40-50%	0.60
Udder health	0.17 - 0.25		0.70
Protein-index	0.31 - 0.38		0.90 - 0.98!
Yield-index	0.31 - 0.38		0.90

Genomic value ≠ Breeding Value

- **Not all genes detected by our SNP panel**

- **Not all SNP –effects are correct**
  - **Estimation model might be too simple: noise by epistatic and dominance effects**
  - **Not enough bulls and recombination in reference bull EBVs**
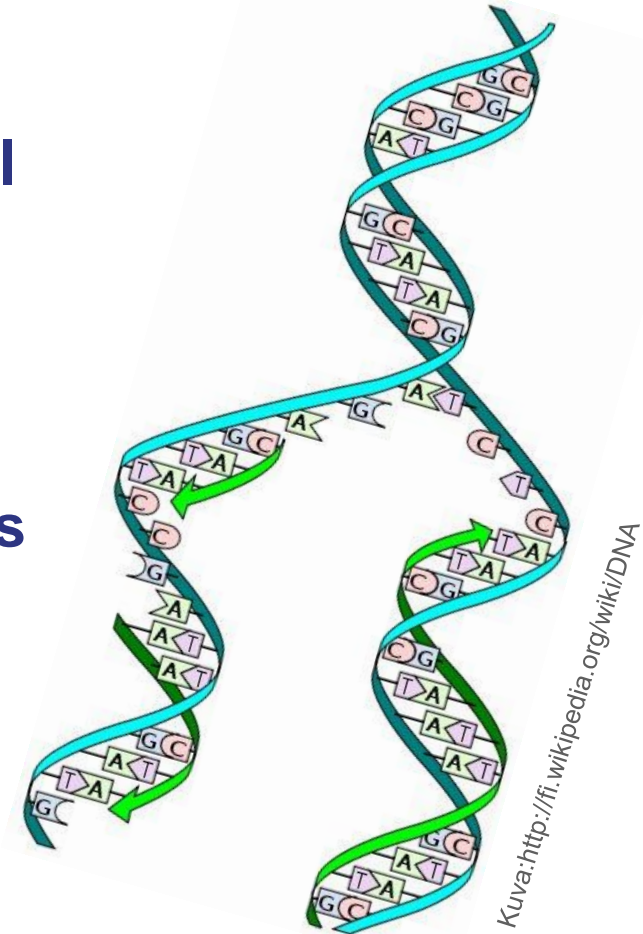
Kuva:http://fi.wikipedia.org/wiki/DNA

Combining:
Genomic and Performance information

More information is obtained if also the classical relationships are utilized

Reasons:

1. **More animals in (Nordic) evaluations than in estimation of DGVs**

   • **Some candidates have no sires etc.**

2. **More records in full pedigree than in sire evaluation**

Abbreviations

- ## DGV
  - ### Direct Genomic Value

- ## EBV
  - ### Estimated breeding value (for example NTM)

- ## GEBV (SNP effects + information from EBVs)
  - ### Genomic Enhanched Breeding value

GEBV = DGV & EBV
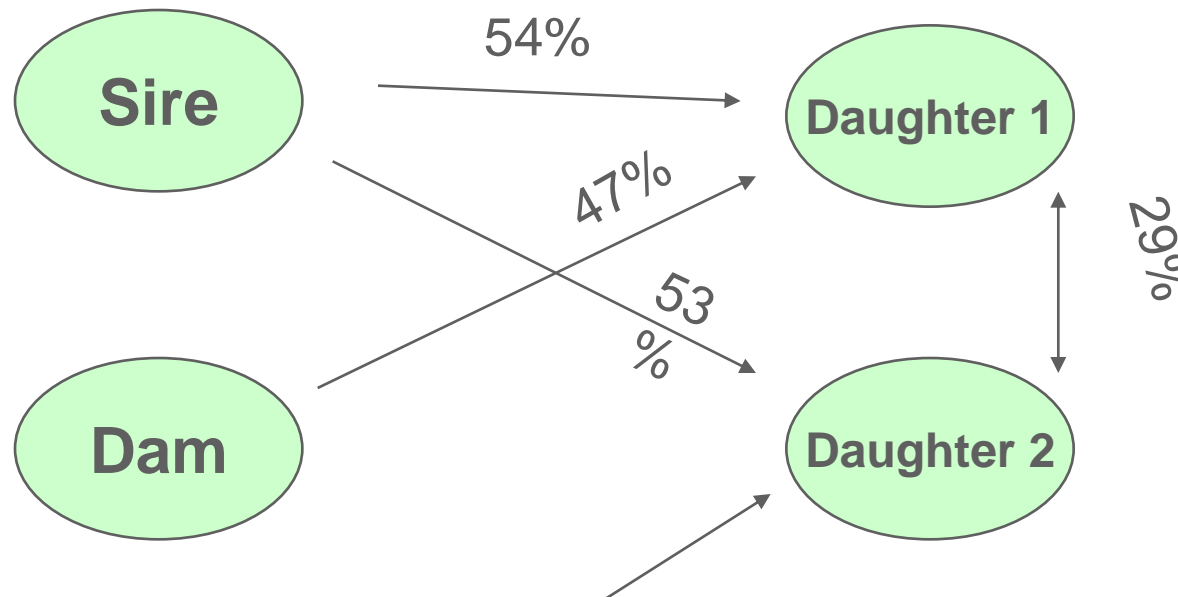
Genomic relationship matrix G

- **EBVs methodology is based on additive relationship matrix among animals**
    - **Resemblance between animals related**
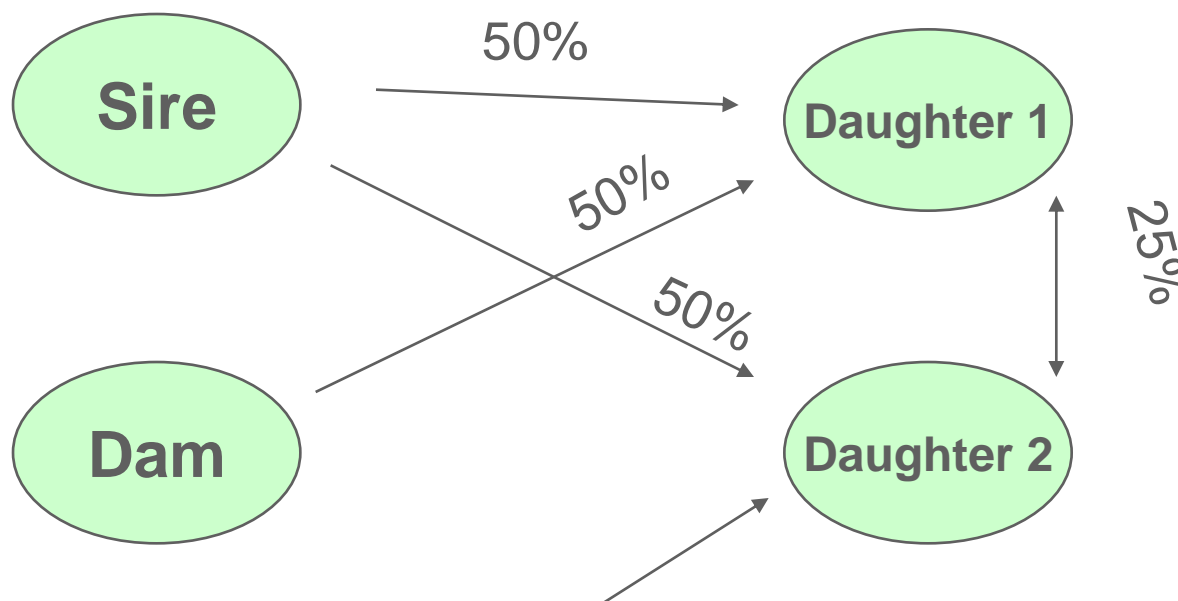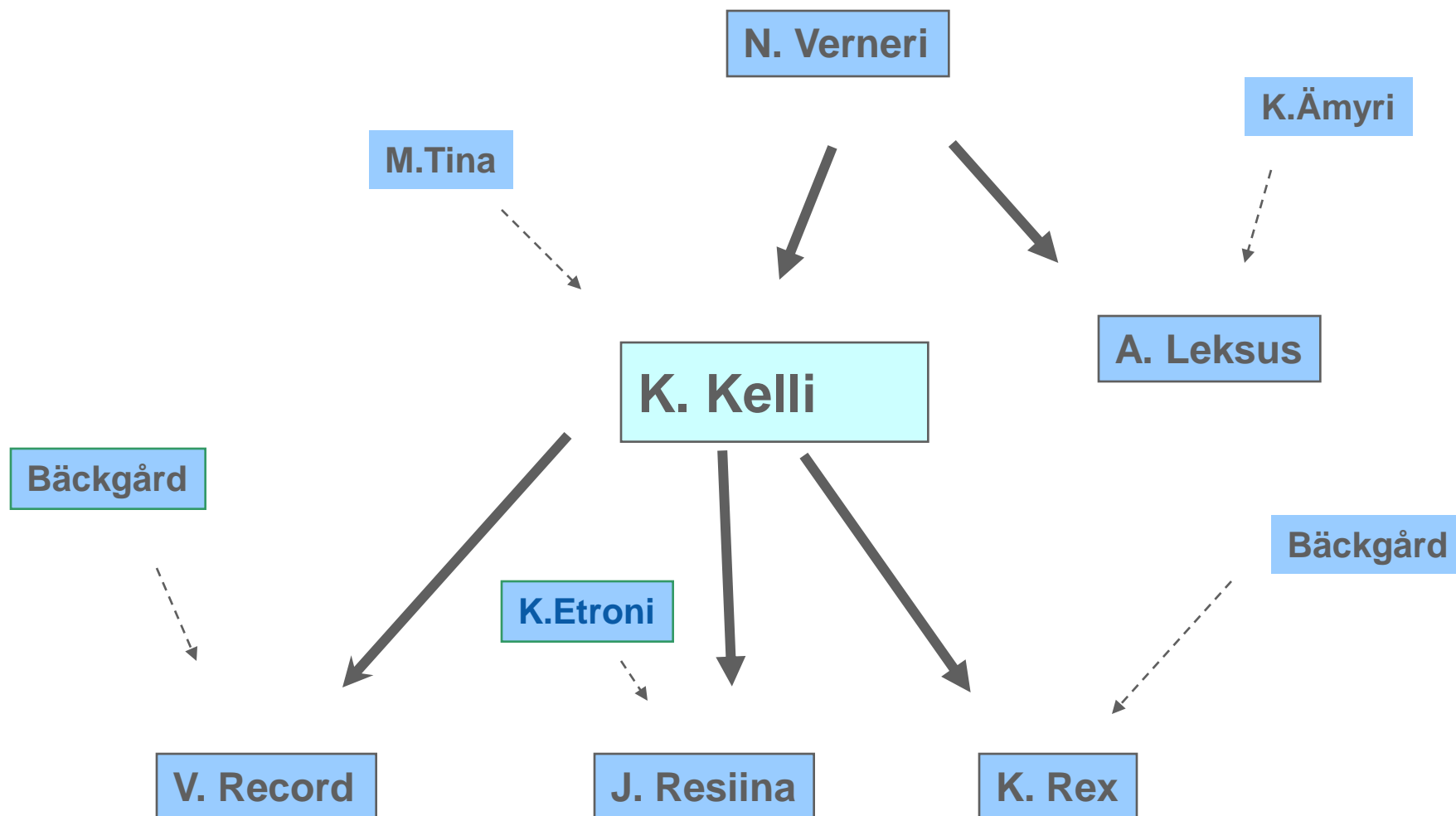

Relationship matrix **A**

Genomic relationship matrix G

- **Additive relationship** matrix could be replaced by **Genomic relationship** matrix
  - **Resemblance between all animals**
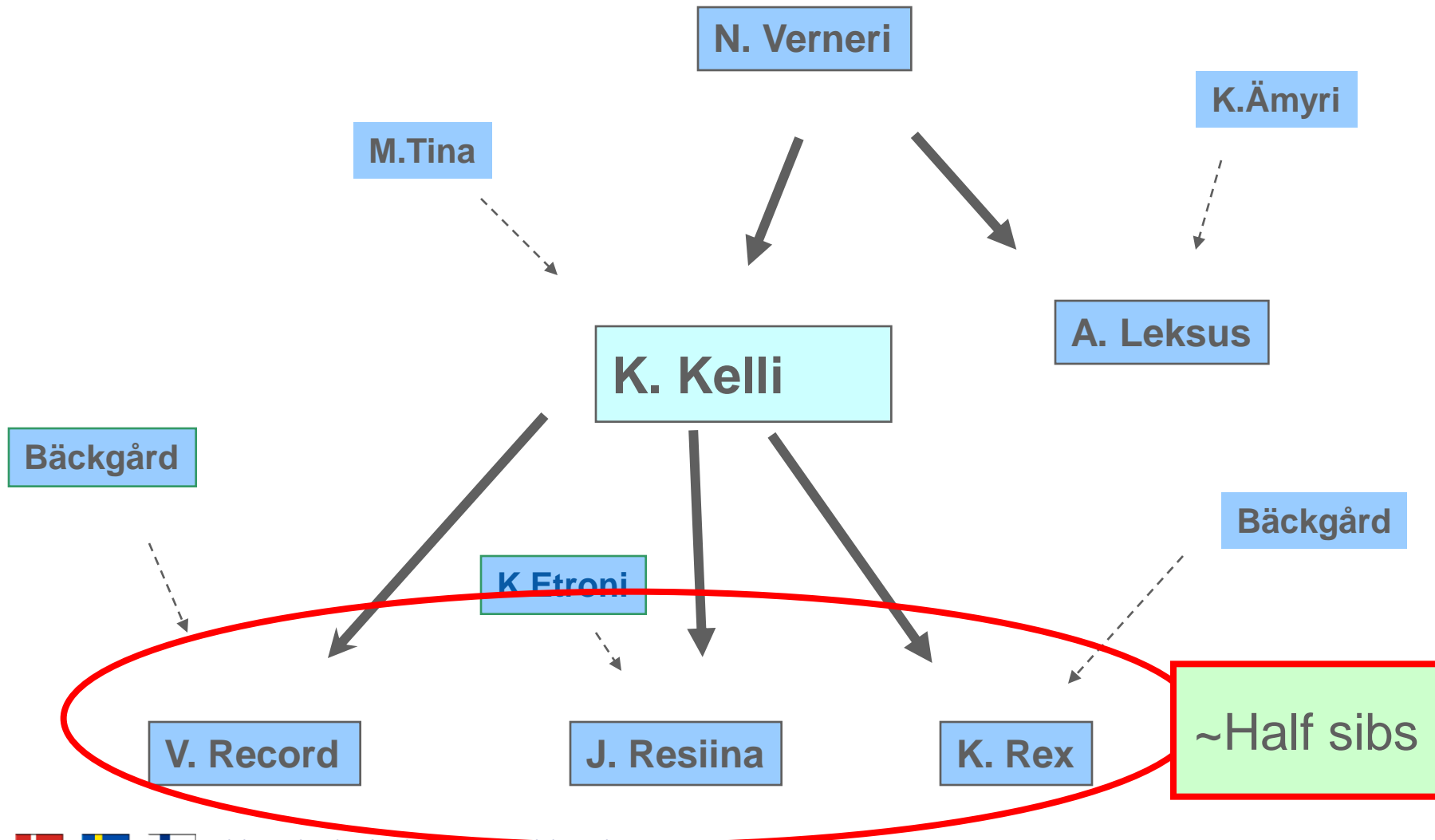

Sire → 54% → Daughter 1

Sire → 53% → Daughter 2

Dam → 47% → Daughter 1

Daughter 1 ↕ 29% Daughter 2

Dam → Daughter 2

Genomic Relationship matrix **G**

Example **bull** pedigree from Finnish AY

Example **bull** pedigree from Finnish AY

Genomic and Additive Relationships

	N. Verneri	K.Kelli	A.Leksus	K. Rex	V.Record	J. Resiina
N. Verneri	1 (1.03)	0.56	0.53	0.33	0.30	0.30
K.Kelli	.48	1.01 (1.04)	0.30	0.56	0.54	0.54
A.Leksus	.56	0.37	1.1 (1.02)	0.18	0.18	0.16
K.Rex	.20	0.55	0.20	1.09 (1.04)	0.37	0.31
V.Record	.31	0.59	0.22	0.45	1.17 (1.04)	0.31
J.Resiina	.21	0.54	0.18	0.29	0.33	1.07 (1.02)

Genomic and Additive Relationships

	N. Verneri	K.Kelli	A.Leksus	K. Rex	V.Record	J. Resiina
N. Verneri	1 (1.03)	0.56	0.53	0.33	0.30	0.30
K.Kelli	.48	1.01 (1.04)	0.30	0.56	0.54	0.54
A.Leksus	.56	0.37	1.1 (1.02)	0.18	0.18	0.16
K.Rex	.20	0.55	0.20	1.09 (1.04)	0.37	0.31
V.Record	.31	0.59	0.22	0.45	1.17 (1.04)	0.31
J.Resiina	.21	0.54	0.18	0.29	0.33	1.07 (1.02)

Combining all animals into G

- **DGV uses "relationships" among all animals in reference set and relates them to candidates**
  - **Some error exists in estimation of g_{ij}**

- **If relative not genotyped, then an average (additive) relationship should be used: COMBINING DGV & EBV**

- **G and A can be combined directly**
  *O. Christensen et al. 2009, A. Legarra et al 2009,
  I. Misztal et al 2009*

Derivation of GEBV

First approach:

GEBV = b_1*DGV + b_2*EBV$_{subset}$ + b_3*EBV$_{national}$

- EBV$_{subset}$ is a EBV that is calculated using animals in reference data set only
- Used in USA and the Netherlands (Van Raden method)
  - They try to minimize the covariance between DGV and EBV$_{national}$ when the SNP solutions for DGV are calculated

Second approach: GEBV=b_1*DGV + b_2*EBV$_{national}$

- DGV and EBV are weighted based on accuracy and covariance
- We use EBVs of reference bulls in calculation of DGVs, this leads high correlation between EBV$_{subset}$ and EBV$_{national}$

Weights in combining EBV and DGV

$$GEBV = b_1 \cdot DGV + b_2 \cdot EBV_{national}$$

- To derive b_1 and b_2 we need to know the accuracies of DGV and $EBV_{national}$, and their correlation

Consider EBV being PA (Parental Average)

- Accuracy depends whether sire (in reference group) has already daughter information
- Correlation between DGV and EBV also depends whether sire of candidate has daughters

DGV reliability for young sires

Table 1. $r^2_{GEBV,EBV}$ for all bulls in the test data, and expected reliability obtained from prediction error variance
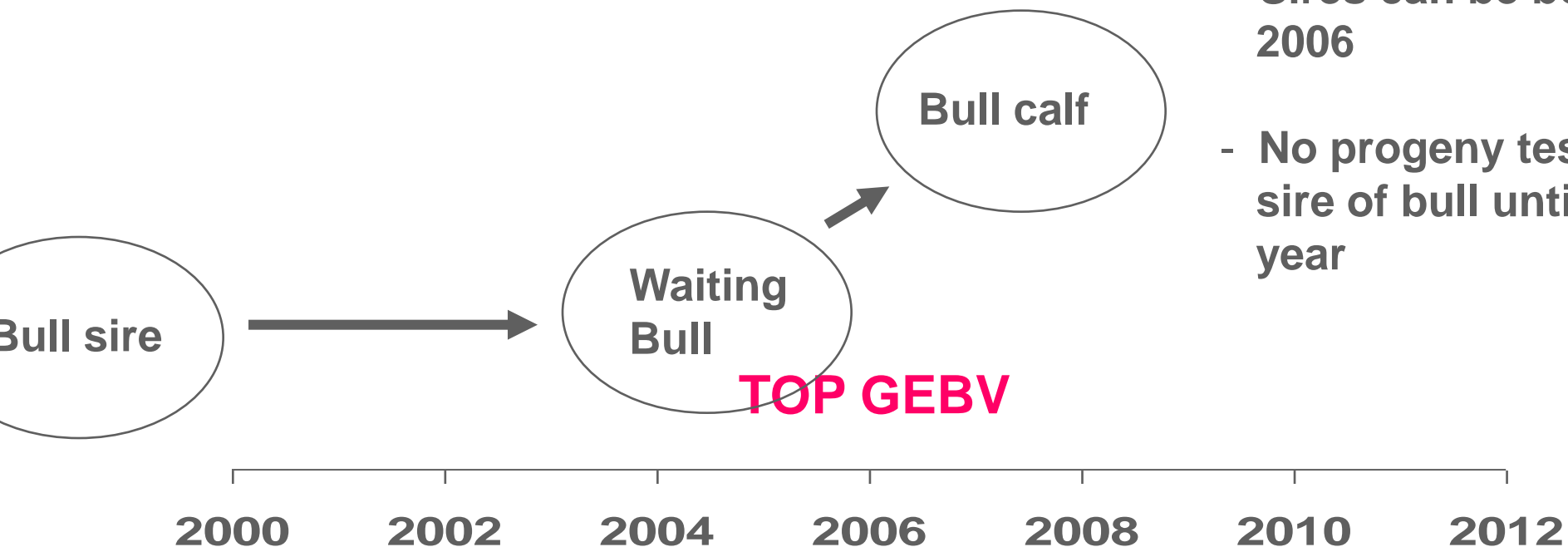
Trait	Sires included in reference data		Sires excluded from reference data	
	Expected reli	$r^2_{GEBV,EBV}$	Expected reli	$r^2_{GEBV,EBV}$
Fertility	0.566	0.412	0.487	0.326
Protein	0.528	0.412	0.493	0.367
Udder-h	0.557	0.435	0.486	0.335

Su, G 11.8.2009

Correlation between PA and EBV

Holstein	Sire in reference group	Sire not in reference
	PA-EBV	PA-EBV
Fertility	0.71	0,54
Protein	0,58	0,47
Udder health	0,70	0,50

We need to predict over generations!

Bulls born 2009
- Sires can be born 2006
- No progeny test for sire of bull until next year

Bull calf

Waiting Bull

Bull sire

TOP GEBV

2000   2002   2004   2006   2008   2010   2012

Reliability of GEBV

Correlation	Reliability		
PA-DGV	PA	DGV	GEBV
0,40	**0,30**	**0,50**	**0,58**
0,50	0,30	0,50	0,55
0,60	**0,30**	**0,50**	**0,52**
0,70	0,30	0,50	0,50

Critical for combining DGV and EBV

- **Reliabilities:**

  - **Reliability of EBV needs to be estimated from full Nordic data!**

  - **Reliability of DGV !**
    ? From validation tests R^2
    ? From selection index $PEV = (R^{-1} + \sigma^{-2}_a G^{-1})^{-1}$
    ? From posterior distribution of DGVs

- **Unbiasedness:**

  - **PA and DGV means have to be "in line" with the EBVs**

  - **Should we decide not to use bull dam EBVs in PA!**

Selection and genetic progress

Dairy Cattle breeding – overlapping generations

- **Need to be unbiasedly evaluated**
  old sires, young sires
  young heifers, old cows to be culled

- **Unbiasedness requires methods that estimate genetic trend correctly**

  - **BLUP accounts selection if visible in data**

  - **Selection on DGV ?**

Phenotypic trend = Genetic + Environmental trend

- **In correct model:**

Herd Mean 9700 kg

$\Rightarrow$

Environmental difference
Genetic difference: E[BV]
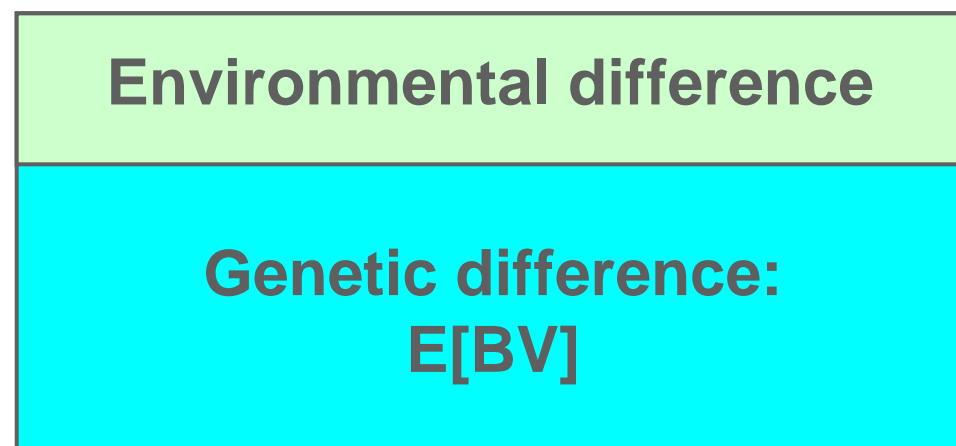
- **Selection on DGV → EBV ≠ E[BV]**

Back-blending of DGV into EBV

- **Incorporates DGVs into EBV calculations**

- **Assures right genetic trend in EBVs**

- **Improves accuracy of EBVs also for non-genotyped relatives**
  - **Consider bull dam with 3 genotyped sons**
  - **Heifer with genotyped (bull) dam**

- **Makes EBV+DGV combining unnecessary**

Blending methods

1. **Back-transform DGV into equivalent daughter performance (EDP)**

    •   *Ducrocq and Liu 2009, Liu et al. 2009*

2. **Consider DGV as a correlated trait**

    •   *Mäntysaari et al. (not implemented yet)*

3. **Combine G matrix and A matrix in BLUP breeding value estimation**

    •   *Christensen et al. 2009, Mistzal et al 2009, Legarra et al. 2009*

Back blending challenge

Liu et al.

- EDP= "Eq_Daugh_Perform" = (DGV – average)/REL
- EDC = "Effective daughter contribution"
    = lamda * REL/(1-REL), lamda=$(4-h^2)/h^2$

    - Let REL=0.4, h2=0.25 → EDC=10
    - BUT: genotyped animal with progeny will get way too much DGV contributions

Ducrocq and Liu:

- EDC are iteratively discounted downwards
    - Animal with 3 genotyped sons → EDC= 5.3
    - But: Animal with 10 genotyped sons → EDC < 0.00

Bi-variate approach

- **Two traits "EBV-trait" and "DGV-trait"**

- **DGV has a h2 ~ 0.99 – 1.00**

$$r_{g,Trait*DGV} = \sqrt{REL_{DGV}}$$

- **Requires 2 consecutive evaluations**
  1. **EBV run (+ DGV computation)**
  2. **Bi-variate blending run**

  (same with Ducrocq and Liu…)

Example of blending

Combining genomic and classical information in national BLUP evaluations

V. Ducrocq[1], Z. Liu[2]

[1] *UMR1313 INRA, Génétique Animale et Biologie Intégrative, 78 352 Jouy-en-Josas, France*
[2] *VIT, Heideweg 1, D-27283 Verden, Germany*

vincent.ducrocq@jouy.inra.fr

Abstract

Blending genomic information with classical performances into a joint BLUP analysis has some appealing features, in particular its simplicity and its potential ability to account for genomic preselection of young sires. A simple approach consists in computing specific genomic equivalent contributions and genomic equivalent daughter performances. Two cases are presented here depending on the way genomic evaluations are performed: using prediction equations or BLUP with relationship matrix. It is shown through a small example that genomic EDC should be computed with caution to avoid double-counting, especially when closely related animals are genotyped. Otherwise, results and inflated reliabilities are obtained.
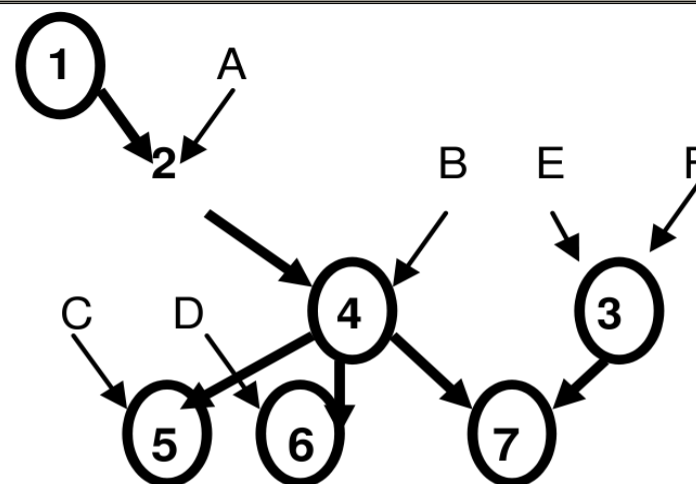
Figure 1: pedigree used for the numerical example (circled individuals are the genotyped ones).

Example:
Blending using genomic EDC method

File   Edit   View   Go   Help

Animal	Iteration 1		Genomic EDC Ψ_i^G	Final reliability (Harris & Johnson)
	True Reliability	Harris & Johnson		
1	0.411	0.414	9.56	0.400
2	0.203	0.224	0	0.191
3	0.439	0.438	8.78	0.400
4	0.508	0.508	5.33	0.400
5	0.445	0.445	8.51	0.400
6	0.445	0.445	8.51	0.400
7	0.487	0.487	6.61	0.400

Table 1 : Reliabilities of the animals from the numerical example, with naive EDC (columns 2 and 3) or with the strategy proposed (column 5). Column 4 indicates the final genomic EDC at convergence (starting value = 10).

Example:
Blending using genomic EDC method

:: Pour permettre de calculer l'ISU selon les mêmes modalités méthodologiques que les races Prim'Holstein, Normande et Montbéliarde, _ □ ×

File   Edit   View   Go   Help

Animal	Iteration 1		Genomic EDC Ψ_i^G	Final reliability (Harris & Johnson)
	True Reliability	Harris & Johnson		
1	0.411	0.414	9.56	0.400
2	0.203	0.224	0	0.191
3	0.439	0.438	8.78	0.400
4	0.508	0.508	5.33	0.400
5	0.445	0.445	8.51	0.400
6	0.445	0.445	8.51	0.400
7	0.487	0.487	6.61	0.400

Table 1 : Reliabilities of the animals from the numerical example, with naive EDC (columns 2 and 3) or with the strategy proposed (column 5). Column 4 indicates the final genomic EDC at convergence (starting value = 10).

Ducrocq and Liu (2009) added 7 more genotyped progeny to the animal 4
→ Accuracy of animal 4 → 0.49
→ EDC of animal 4 → 0.0

Example:
Blending using bi-variate method

Animal	Accuracy "DGV"	Accuracy "GEBV"	Reliability "GEBV"
1	0.99	0.630	0.396
2	0.63	0.398	0.159
3	1.00	0.629	0.396
4	1.00	0.629	0.396
5	1.00	0.629	0.396
6	1.00	0.629	0.396
7	1.00	0.629	0.396
8	0.99	0.629	0.396
9	0.99	0.629	0.396
10	0.99	0.629	0.396
11	0.99	0.629	0.396

Adding 7 more genotyped progeny to the animal 4
→ Accuracy of animal 4 → 0.40

Combining and back-blending simultaneously

Genomic EDC and bi-variate methods can do both combining and back-blending
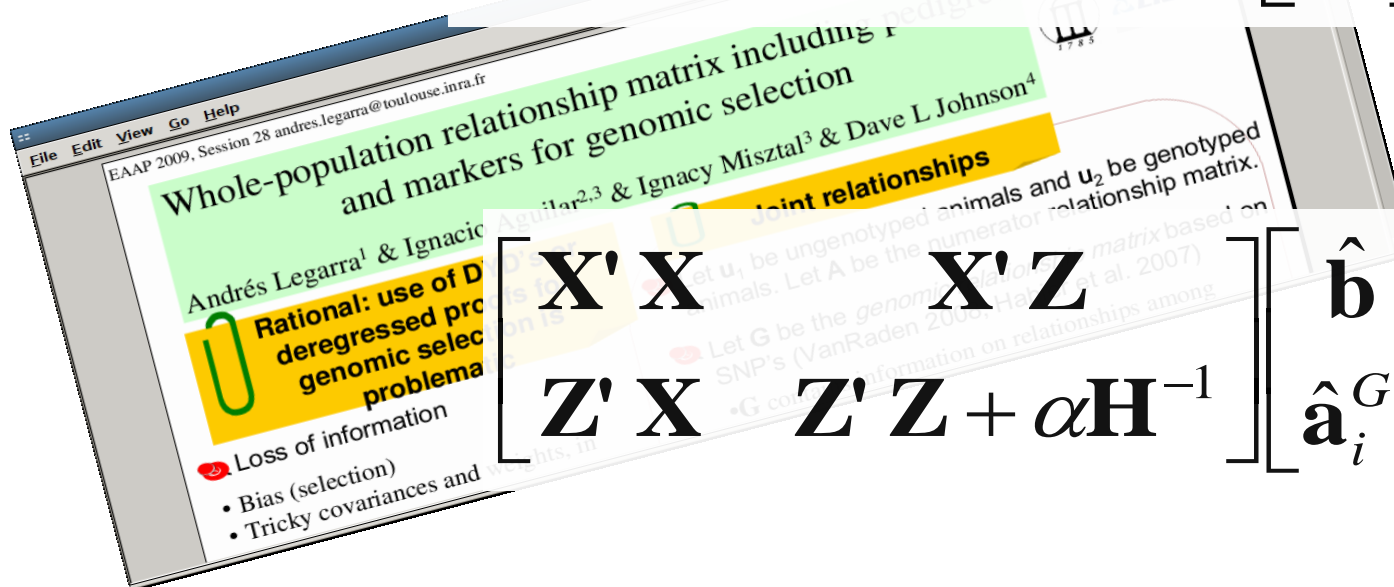
- The EBV from "pseudo" BLUP round has genomic information included

Combining G matrix and A matrix in BLUP breeding value estimation (*Christensen 2009*)

- Outmost elegant solution that combines features of complicated BLUP-models and GBLUP

HENDERSON'S
MME

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}}_i \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Whole-population relationship matrix including
and markers for genomic selection

Andrés Legarra[1] & Ignacio Aguilar[2,3] & Ignacy Misztal[3] & Dave L Johnson[4]

Rational: use of D...
deregressed pro...
genomic selec...
problema...
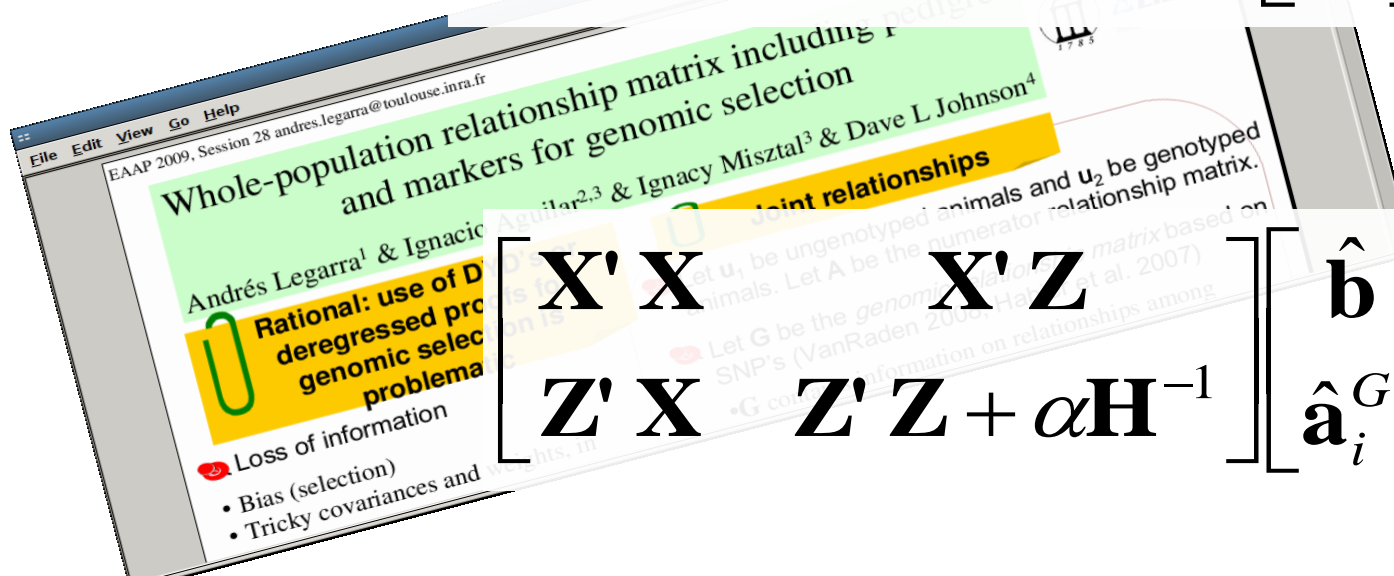
• Loss of information
• Bias (selection)
• Tricky covariances and ...

Joint relationships

...et $\mathbf{u}_1$ be ungenotyped animals and $\mathbf{u}_2$ be genotyped
...nimals. Let A be the numerator relationship matrix.

• Let G be the *genomic relationship matrix* based on
  SNP's (VanRaden 2008, Habier et al. 2007)
• G co... information on relationships among

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}}_i^G \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X'\,X} & \mathbf{X'\,Z} \\ \mathbf{Z'\,X} & \mathbf{Z'\,Z} + \alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}}_i \end{bmatrix} = \begin{bmatrix} \mathbf{X'\,y} \\ \mathbf{Z'\,y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X'\,X} & \mathbf{X'\,Z} \\ \mathbf{Z'\,X} & \mathbf{Z'\,Z} + \alpha\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}}_i^G \end{bmatrix} = \begin{bmatrix} \mathbf{X'\,y} \\ \mathbf{Z'\,y} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

NAV plan's

1. **Combine DGV and PA to get GEBV for genotyped young animals**

2. **Back-blend DGV information to all evaluated animals**

3. **Test possible better methods**

Time table:

- **Ongoing, ~ end of 2009**

- **"Official GEBV" 2009-2010**
  May 2010 Interbull test

- **2009-2011**

Summary

- **Combining and back-blending will lead into best use of genomic information**

- **Methods are developing rapidly:**
  - **New methods are not yet tested thoroughly**

SMALL CHIP,

BIG CHANGE

IN THE DAIRY PRODUCTION

THANK YOU FOR

YOUR ATTENTION