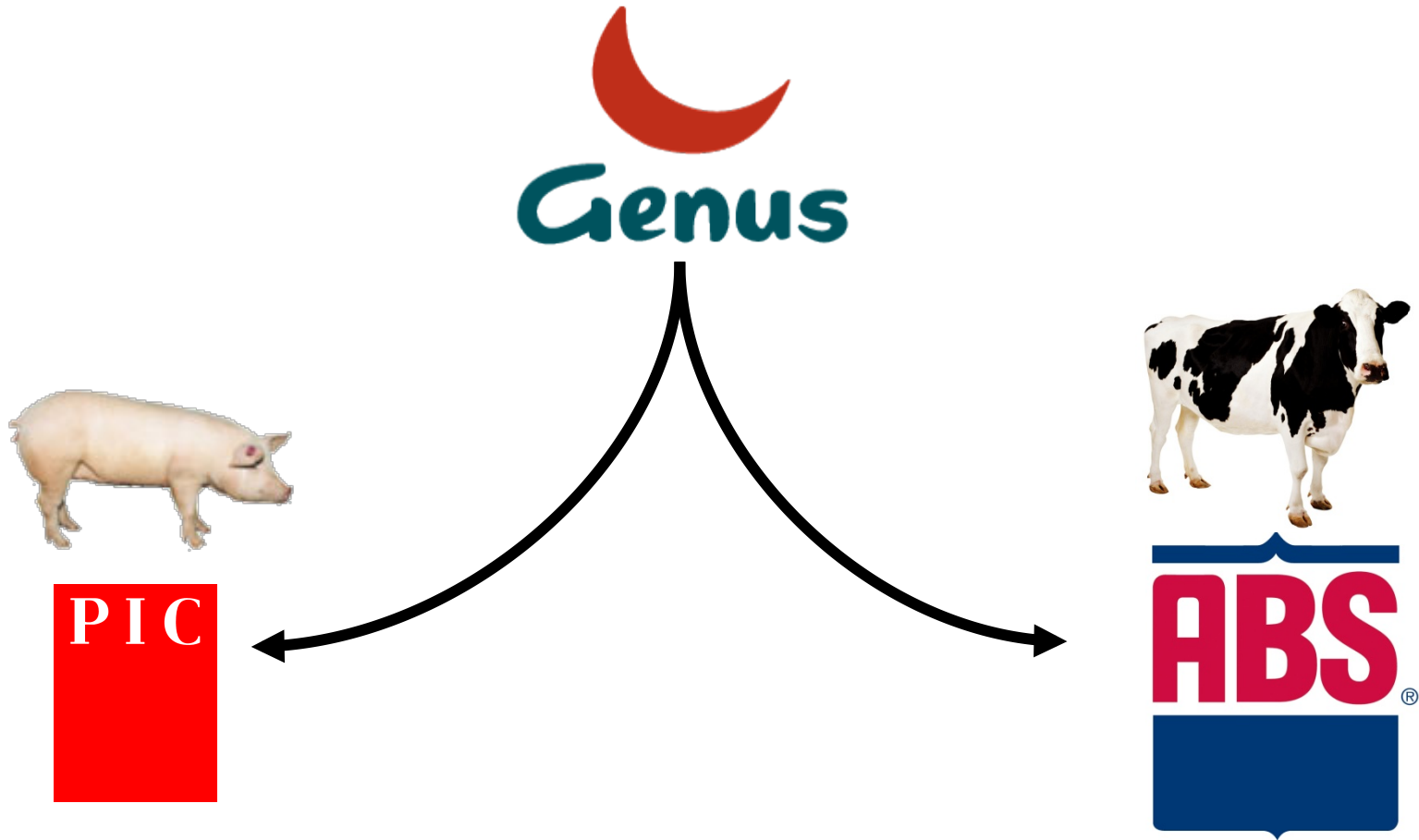


# Implementation of Genomic Selection in Commercial Pig Breeding

Selma Forni

June 1<sup>st</sup>, 2012

## Where I work



# Genomic Selection in the beginning

- Bayesian Regressions to identify markers associated with specific traits – too expensive HD genotyping all selection candidates.
- Estimate marker effects for each trait in each line.
- Estimate BV combining marker and pedigree information of selection candidates.
- Update marker effects when new genotypes/phenotypes are available.

Line	Trait	Accuracy		
		Pre-Genomics	Post-Genomics	% Increase
Sire Line	Scrotal Hernia	0.239	0.332	38.9%
Sire Line	Mortality	0.215	0.340	58.1%
Dam Line (LR)	Total Born	0.560	0.787	40.5%

# But we want to implement Genomic Selection for all traits in all lines ...

- Bayesian Regressions ~~to identify markers associated with specific traits – too expensive HD genotyping all selection candidates.~~
- Estimate marker effects for each trait in each line.

Training” requires a large number of individuals with genotypes, and phenotypes or progeny recorded.

This is problematic in swine populations that are usually much smaller than other species such as dairy cattle.

- Estimate BV combining marker and pedigree information of selection candidates.

every day – automated system and accessed in 36 countries

- Update marker effects when new genotypes/phenotypes are available.

every week : ~60 traits ~8 lines

# Much simpler if we use Single-Step Genomic Evaluation

- ✓ Easily integrated into systems for routine BLUP that we have been using for years.
- ✓ Estimate BV for genotyped and non-genotyped all at once.
- ✓ Automatically links newest phenotypes and genotypes with pedigree information.
- ✓ It can be implemented in any model (**multiple traits**, maternal effects).
- ✓ The number of parameters do not increase with the number of markers.

## Necessary change:

A priori distribution for marker effects: Normal  
Very small impact for most traits

Tue class

We can revisit that in the future

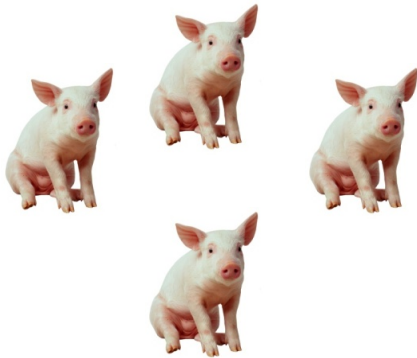
Wed class

# Single-Step Genomic Evaluation

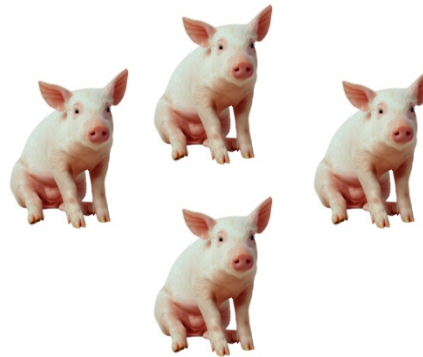
$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + H^{-1} \otimes G_0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Same (co)variance components

Not genotyped



Genotyped



$$G = \frac{(M - P)(M - P)'}{2 \sum p_i (1 - p_i)}$$

Relationships are defined with respect to a base population

Allele frequencies in the base population??

# Single-Step Genomic Evaluation

	<i>Average diagonal</i>	<i>Average off- diagonal</i>	<i>Additive variance (se)</i>	<i>Accuracy GEBV (PEV)</i>
A	1.000	0.032	2.27 (±0.52)	0.22
G05	1.253	0.595	3.43 (±0.56)	0.37
GMF	1.697	1.022	3.43 (±0.56)	0.49
GOF	0.936	0.000	2.41 (±0.39)	0.30
GOF*	0.505	0.000	4.46 (±0.73)	0.43
GN	1.002	0.000	2.25 (±0.36)	0.28

**A** = Pedigree-based Relationship

**G05** = Genomic Relationship with allele frequency equal to 0.5

**GMF** = Genomic Relationship with allele frequency equal to average MAF

**GOF** = Genomic Relationship with observed allele frequency

**GOF\*** = Genomic Relationship with allele frequency following a Beta distribution

**GN** = Normalized Relationship Matrix

$$GN = \frac{(M - P)(M - P)'}{\left\{ \text{trace} \left[ (M - P)(M - P)' \right] \right\} / n}$$

Forni et al. (2011)

# The algorithm evolved!!

Thur class

Vitezica et al. (2011)

$$\mathbf{G}^* = \left(1 - \frac{1}{2}\alpha\right) \mathbf{G} + \mathbf{1}\mathbf{1}'\alpha$$

Christensen et al. (2012)

$$\mathbf{G}_a = \beta \mathbf{G} + \alpha,$$

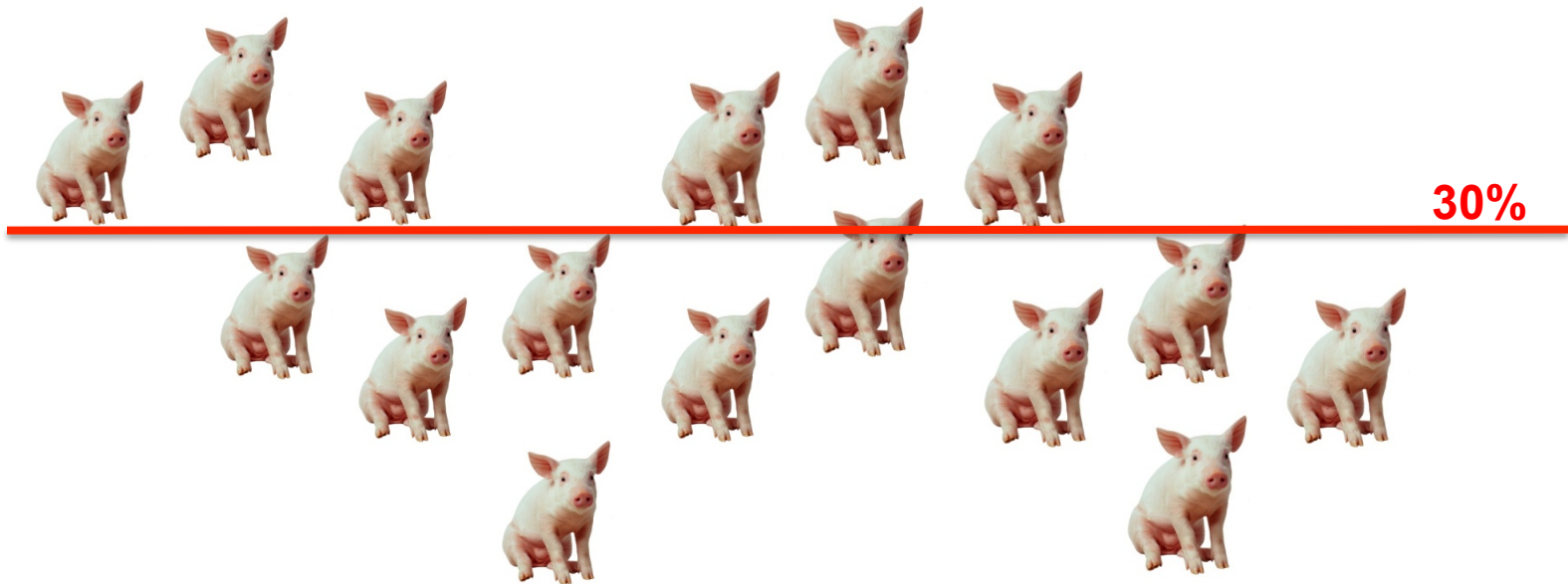
where  $\beta$  and  $\alpha$  solved the system of equations

$$\text{Avg}(\text{diag}(\mathbf{G}))\beta + \alpha = \text{Avg}(\text{diag}(\mathbf{A}_{11})),$$

$$\text{Avg}(\mathbf{G})\beta + \alpha = \text{Avg}(\mathbf{A}_{11}).$$



# Does all that really matter ... \$\$\$\$?



GN is the “best job” that we can do: 597 animals selected / year

## How many animals would not had been selected?

A	G05	GMF	GOF*	GOF
141	30	29	43	10

# What we have learned

- Pedigree and genomic relationships need to be calculated with respect to the same base population.
- To achieve that one needs to use the best guess of allele frequencies in the base population and adjust G (tunning).
- Otherwise:
  - Bias in GEBV
  - Underestimation of PEV
  - Bias in variance components
  - Less than optimal selection decisions

# Single-Step Genomic Evaluation

- ACC improvement even for sires and dams that already have progeny data

TRAIT	# Daughters	ACC EBV	ACC GEBV	increase
Total Number born	high	0.57	0.66	16%
Stillborn	high	0.57	0.66	16%
Litter weaning weight	high	0.45	0.52	15%
Interval weaning - mate	high	0.45	0.53	18%
Survival birth - weaning	high	0.33	0.36	9%

- 4,200 animals genotyped 60k

# Single-Step Genomic Evaluation

- Greater ACC improvement for young sires and dams

TRAIT	# Daughters	ACC EBV	ACC GEBV	increase
Total Number born	low or zero	0.44	0.56	27%
Stillborn	low or zero	0.44	0.56	27%
Litter weaning weight	low or zero	0.38	0.46	21%
Interval weaning - mate	low or zero	0.31	0.42	35%
Survival birth - weaning	low or zero	0.26	0.30	15%

- 1,000 animals genotyped 60k

# Single-Step Genomic Evaluation

- ACC improvement for progeny of genotyped sires and dams – **these animals were not genotyped**

TRAIT	Progeny of <b>all</b> sires and dams genotyped			Progeny of <b>young</b> sires and dams genotyped		
	ACC EBV	ACC GEBV	increase	ACC EBV	ACC GEBV	increase
Total Number born	0.39	0.42	7%	0.36	0.41	14%
Stillborn	0.39	0.41	5%	0.36	0.40	11%
Litter weaning weight	0.41	0.43	5%	0.31	0.34	10%
Interval weaning - mate	0.37	0.39	5%	0.25	0.29	16%
Survival birth - weaning	0.29	0.30	3%	0.41	0.43	5%

- But we still have the same BV for full siblings

# ACC improvement for the entire population

Table 2 Daily gain

	All		Genotyped		Non-genotyped	
	$Cor(GBV, y_d)$	Reg	$Cor(GBV, y_d)$	Reg	$Cor(GBV, y_c)$	Reg
Univariate						
Ped	0.193 <sup>a</sup>	0.91	0.179 <sup>a</sup>	0.72	0.193 <sup>a</sup>	0.90
1-step	0.226 <sup>b</sup>	0.92	0.345 <sup>b</sup>	0.94	0.217 <sup>b</sup>	0.90
1-step-a	0.229 <sup>c</sup>	0.93	0.353 <sup>c</sup>	0.97	0.219 <sup>c</sup>	0.91
GBLUP			0.351 <sup>b,c</sup>	0.80		
Bivariate						
Ped	0.193 <sup>a</sup>	0.90	0.177 <sup>a</sup>	0.72	0.193 <sup>a</sup>	0.90
1-step	0.225 <sup>d</sup>	0.92	0.344 <sup>b</sup>	0.94	0.216 <sup>d</sup>	0.90
1-step-a	0.228 <sup>e</sup>	0.93	0.352 <sup>c</sup>	0.97	0.218 <sup>e</sup>	0.91
GBLUP			0.352 <sup>b,c</sup>	0.806		

Christensen et al. (2012)

# Imputation

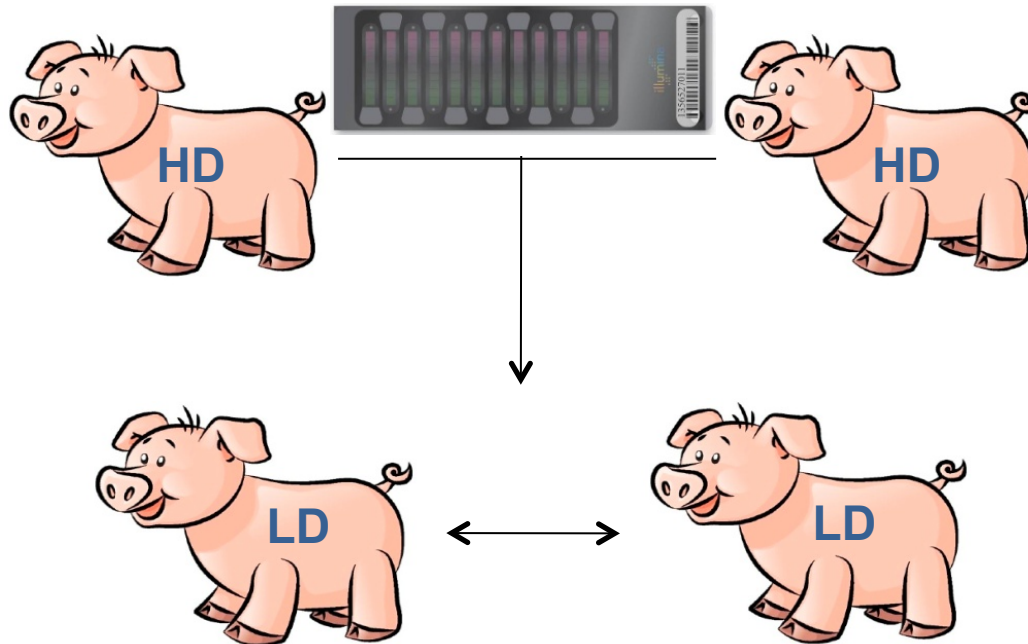
Imputation\*: using well-spaced LD genotypes on selection candidates to 'fill-in' missing HD genotypes

GN parents, grandparents, ...

All HD genotyped

GN progeny

HD genotype imputed from LD panel and pedigree HD



# AlphaImpute Results


## Imputation accuracy for two low-density panels

3k	384	
0.99	0.96	Both Parents
0.98	0.89	Sire and MGS
0.99	0.94	Dam and PGS
0.98	0.87	Sire
0.97	0.87	Dam
0.95	0.81	Other



# Accuracy will be improved by imputing 60k genotypes on selection candidates

TRAIT	Have their own record	ACC EBV	ACC GEBV	Increase
Total Number born	<b>NO</b>	0.36	0.62	<b>71%</b>
Total Number born xbred	<b>NO</b>	0.22	0.39	<b>75%</b>



Phenotype measured in commercial farms  
Expensive to measure  
How we can generate \$\$ for our costumers  
Genetic improvement depends on **multitrait analysis**

# Pedigree is not always “fair”

Thur class

	Animals in the pedigree	Parents without records or genotypes	Animals related to genotyped	Time to converge
PED_full	605,046	154,075	6,987	38h 56min
PED_18	473,825	22,854	6,987	30h 20min
PED_10	473,789	22,818	6,951	33h 20min
PED_3	471,234	20,236	5,366	4h 20min
PED_2	468,057	17,086	5,058	2h 40min
PED_1	463,609	12,638	4,691	2h 20min

# Pedigree is not always “fair”

Thur class

	Diagonals	Off-diagonals	All
PED_full	0.31	0.72	0.78
PED_18	0.31	0.72	0.78
PED_10	0.31	0.72	0.78
PED_3	0.37	0.70	0.74
PED_2	0.39	0.72	0.70
PED_1	0.39	0.75	0.69

# Large fluctuations in GEBV when young genotyped animals were added

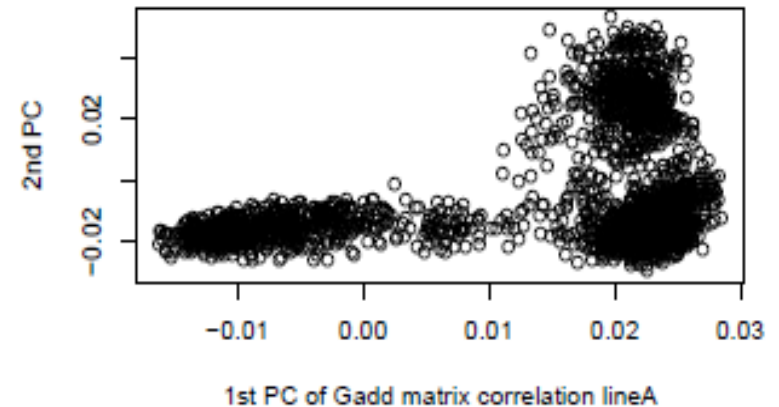
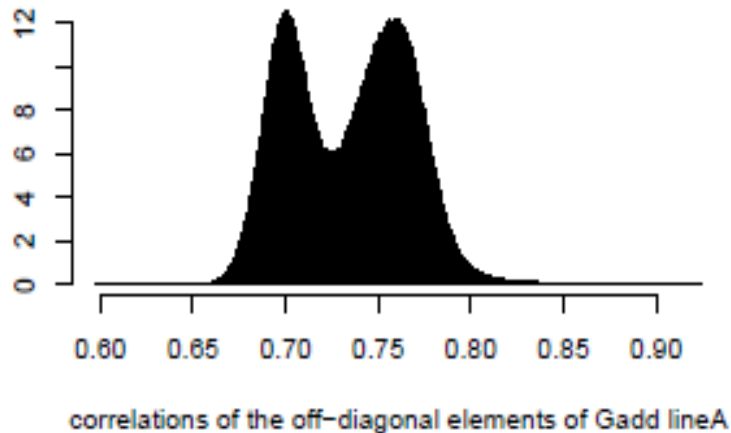
## Correlations between GEBVs from consecutive months

TRAIT	G_MATRIX_DEC	G_MATRIX_JAN
TOTAL BORN	0.99	0.95
STILLBORN**	0.99	0.08
TOTAL BORN XBRED	0.98	0.98
STILLBORN XBRED**	0.35	0.06
SURVIVAL**	0.93	0.75
LITTER WEANING WEIGHT	0.99	0.99
DAYS FROM FARROWING TO NEXT MATING	0.99	0.99

\*\* Traits with low  $h^2$  ( $< 0.10$ ) and recorded in few herds

# Principal Components of G: population stratification

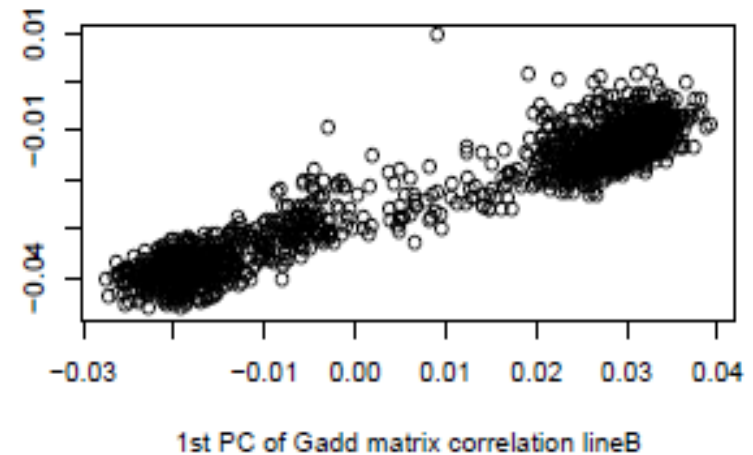
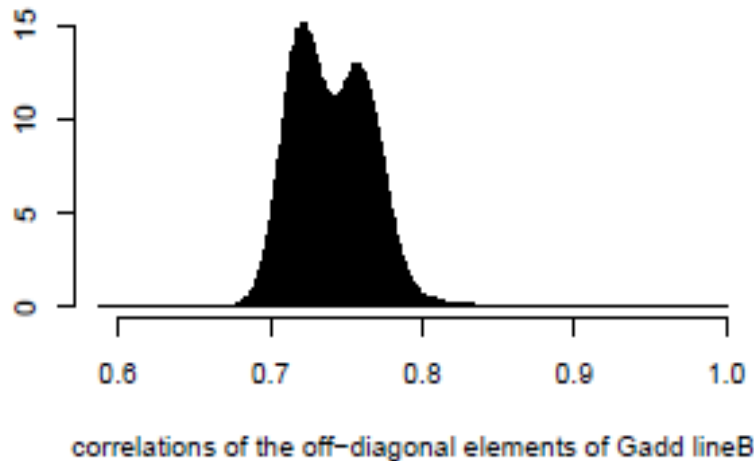
Line A



Groups of animals more related between themselves than to the rest of the population.

# Principal Components of G: population stratification

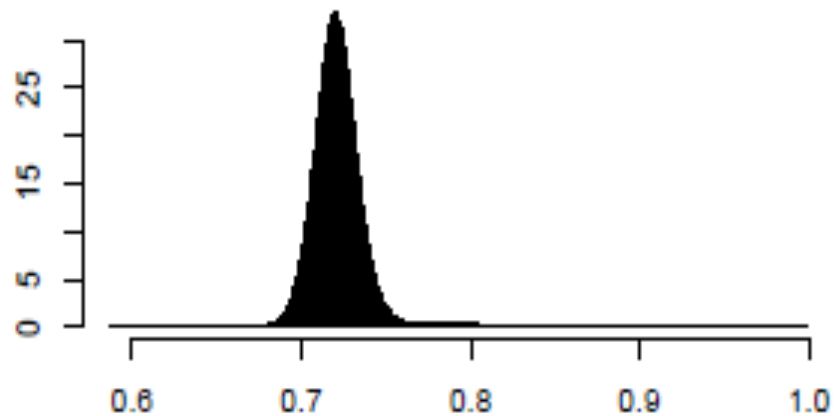
Line B



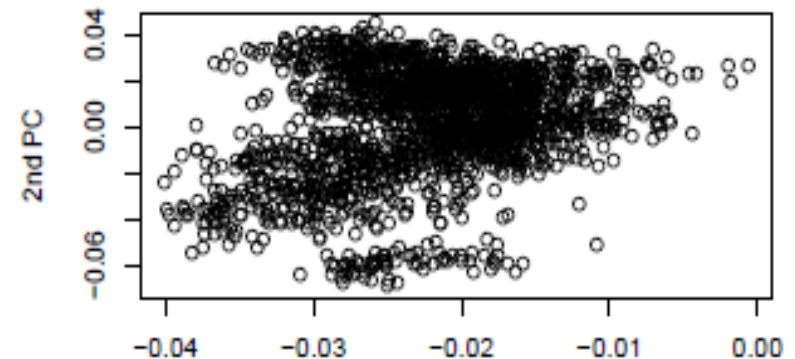
Groups of animals more related between themselves than to the rest of the population.

# Principal Components of G: population stratification

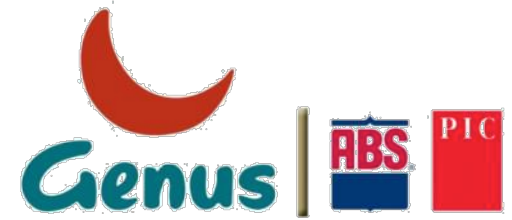
Line AB



correlations of the off-diagonal elements of Gadd lineAB



1st PC of Gadd matrix correlation lineAB



**Questions?**

**THANK YOU**