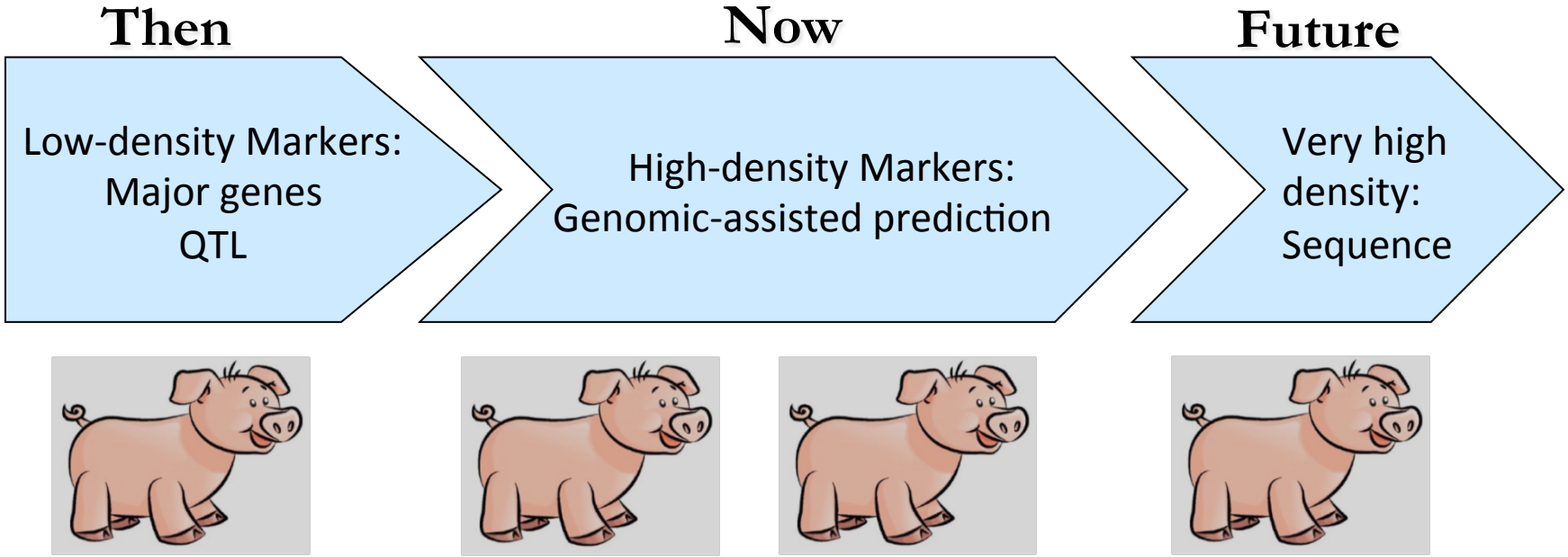
Application of genomic-assisted selection in swine breeding

Selma Forni, Matthew A Cleveland* and Nader Deeb

Genus plc, Hendersonville, TN USA

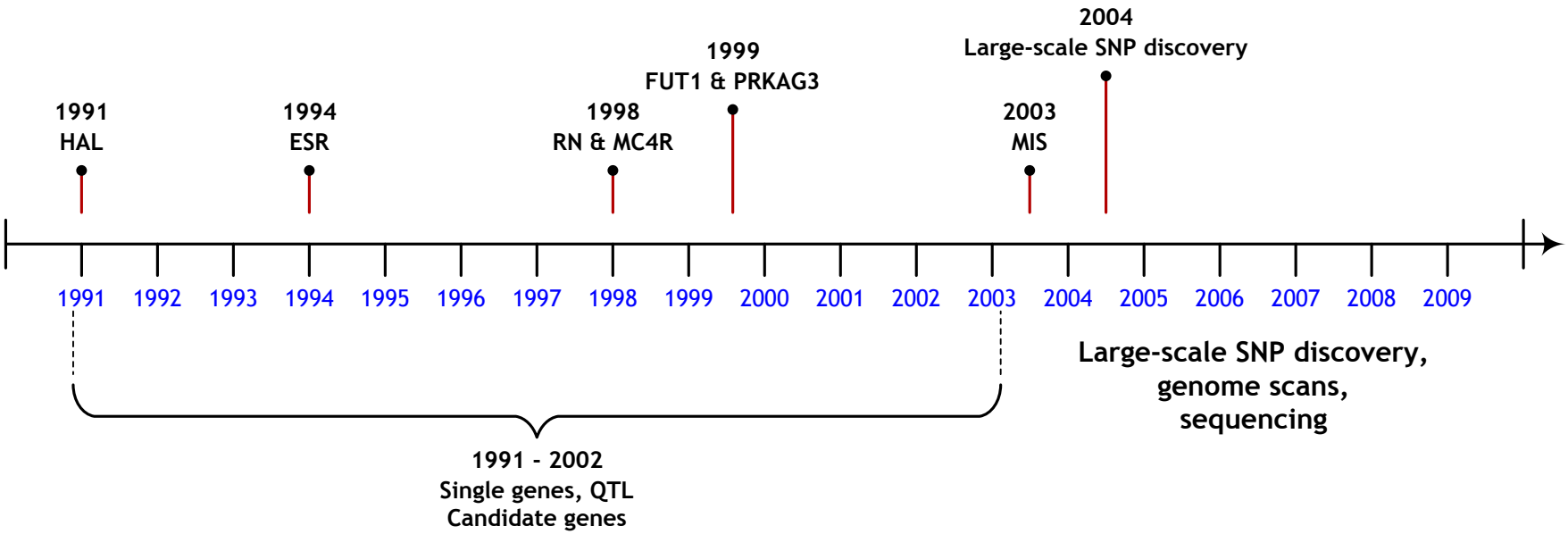
Session 17

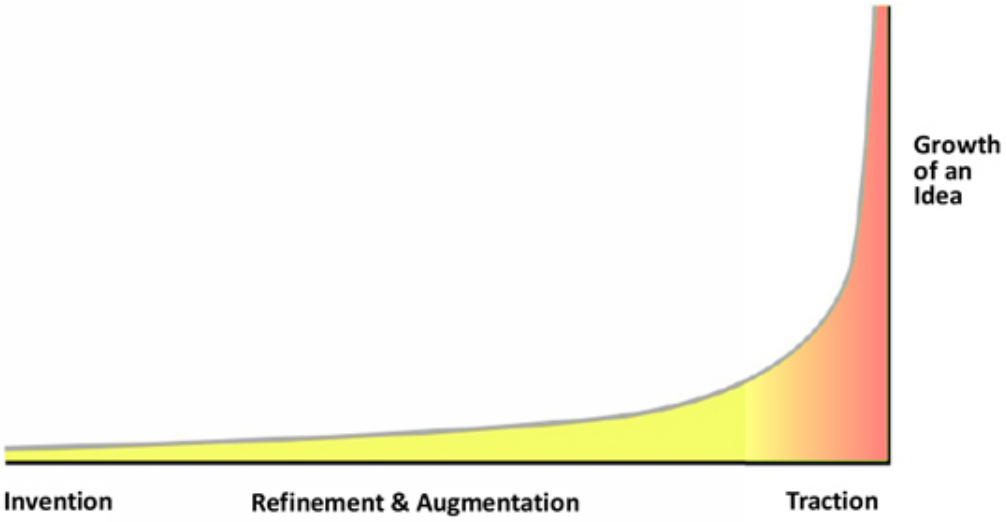
Genomic Technologies

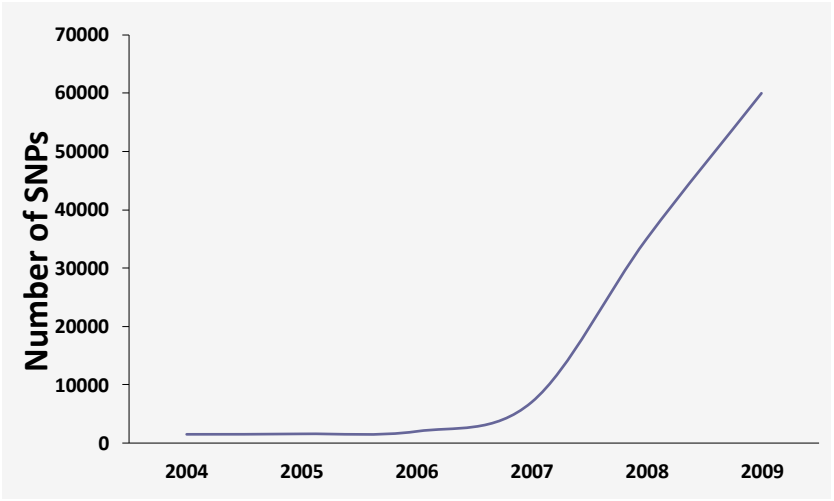


2

DNA marker use in commercial pig breeding: The long nose of innovation







Single-gene and QTL discovery

- Pig breeders have been using genetic marker technology since the early 1990s:
  - The halothane gene (HAL) - porcine stress syndrome
  - Napole gene (RN-) - low pH and water holding

- There are several examples of QTL discoveries in swine populations:
  - Growth, meat quality, body conformation, feed intake, disease resistance, birth defects (Dekkers, 2004)

Application of Marker-Assisted Selection (MAS)

- **The extension of MAS application in commercial breeding programs is not clear.**
  - **Increase in response up to 30% in litter size using ESR in selection index (Rothschild and Plastow, 1999).**
  - **Commercial lines specifically marketed based on fixation of genes RYR and RN (Knap et al., 2002).**

Marker-assisted Selection at PIC

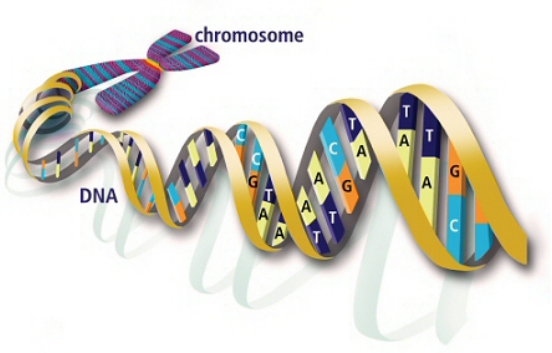
- Significant markers added to EBV (1-7k SNP scans)

Trait	# SNPs used in routine evaluations across lines
Loin depth	43
Backfat	33
Avg lifetime hot carcass daily weight gain	28
Leg score	27
pH24	14
Test average daily feed intake	11
Total number born	7
Stillborn	7
Pre-weaning mortality (piglet trait)	7
Piglet survival (sow trait)	2
Marbling	1

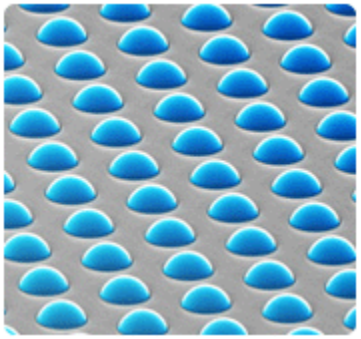
157 markers discovered before high-density genotyping

High-density Single Nucleotide Polymorphisms

Availability of the Illumina PorcineSNP60 BeadChip in January 2009 was a key enabling tool



- 64,000 SNP marker tests
- Cost for genotyping one animal ~~≈$150~~ $100



illumina

PorcineSNP60 BeadChip

The beadchip holds probes that simultaneously identify the alleles present for every marker on the chip in the DNA sample applied

PRRS Host Genetics Consortium (PHGC)

- Nursery pig model to assess resistance to PRRSV infection (Lunney et al., 2011).
- 7 groups of 200 pigs infected with PRRSV and genotyped 60k SNP.
- Viral load and weekly weights were recorded during 42 weeks.
- A 33 SNP region explained 15.7% of genetic variance for viral load and 11.2% for weight gain (Boddicker et al., 2011).

Genetic Markers Against Boar Taint

- **Boar Taint components (Merks et al., 2010):**
  - **Ansdrostenone ($h^2 = 0.64 \pm 0.08$)**
  - **Skatole ($h^2 = 0.36 \pm 0.07$)**
  - **Indole ($h^2 = 0.26 \pm 0.06$)**

- **Traditional selection can result in reproductive problems (Zamaratskaia and Squires, 2009).**

- **SNP explained from 2.5% to 16.3% of the total variation of boar taint components (Moe et al., 2009).**

- **Use of markers would decrease skatole levels from 20% to 53% and androstenone from 26% to 61% (Squires and Schenkel, 2010).**

Genome-wide Association Studies

- Sow reproductive traits (Uimari et al., 2011)
- Body composition and structural soundness (Fan et al., 2011)
- Scrotal hernia (Stinckens et al., 2011)
- Sow productive life (Onteru et al., 2011)
- Roan coat color (Cho et al., 2011)

Other Genome-wide Applications

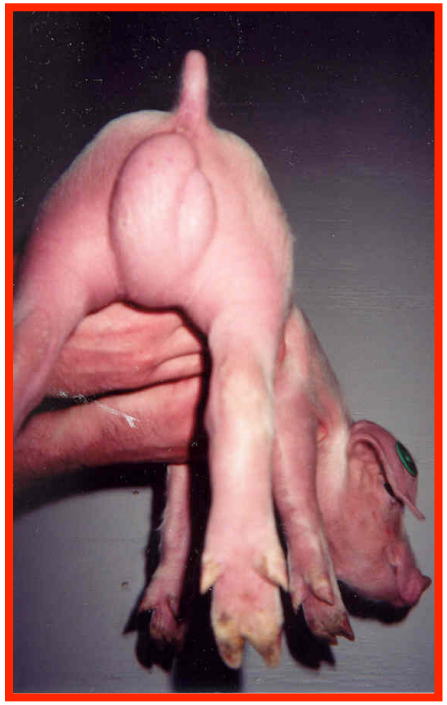
- Estimating LD decay in a commercial pig population (Deeb et al., 2010)
- Population genetic diversity and comparison to humans (Zhang and Plastow, 2011)
- Estimating LD, effective population size and persistence of phase in four domestic swine breeds (Badke et al., 2012).

11

Application of GWAS by PIC

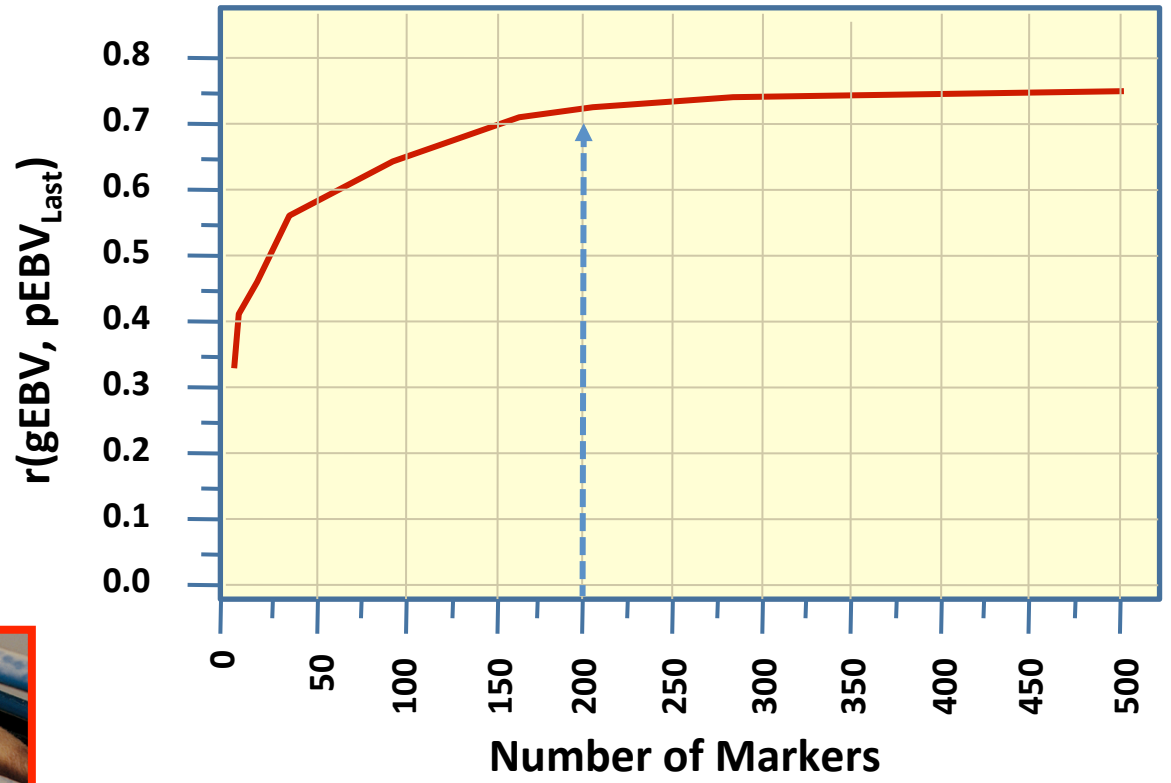
- Beginning in September 2010
- Trait-line specific

  - Scrotal hernia
  - Grow-finisher mortality
  - Total number born



Low-density Panel Development

Litter Size



Deeb et al. (2010)

Incorporate Panel into Genetic Evaluation

$$\Sigma[Prob_{11}x\text{-}\alpha + 0 + Prob_{22}x\alpha]$$

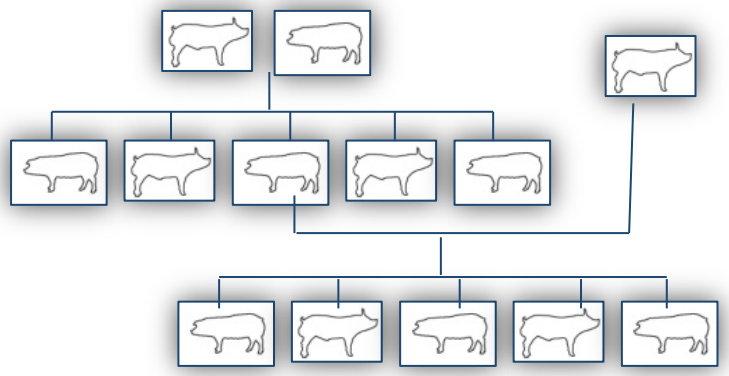
$$\Sigma(\alpha_i)$$

Genomic EBV

Genomic Enhanced Breeding Values (Blended BV)

$$bEBV = b_p \times pEBV + b_g \times gEBV$$

Polygenic EBV



Low-density SNP Panels

PIC has used ~100-200 trait-specific small panel markers for economically important traits in key lines

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%

A large number of selection candidates have been genotyped for each of the two panels

Non-parametric methods to select markers against scrotal hernia

Probability of a method identifying an individual with genetic susceptibility above the population average.

	TBA	BL	RF	L_2B	L_hB
Line A	0.64	0.65	**0.67**	0.55	0.60
Line B	0.70	0.69	**0.73**	0.60	0.72
Line C	0.62	0.62	**0.67**	**0.67**	0.66

TBA = Threshold Bayes A

BL = Bayesian Lasso

RF = Random Forest

LB = Boosting

Gonzalez-Recio & Forni (2011)

High-density Genotyping

- Possibility of scanning the complete genome to search for SNP associated to QTL.
- Possibility of using a large number of markers for prediction under different assumptions.

Computing GEBV for several lines weekly

- "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.

- "Training" would require very high computer power.
  - PIC weekly evaluations: 65 traits
    38,000,000 EBV stored
    (8% of all EBV computed)

Single-Step Genomic Evaluation

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

Misztal et al. (2009), Aguilar et al. (2009)

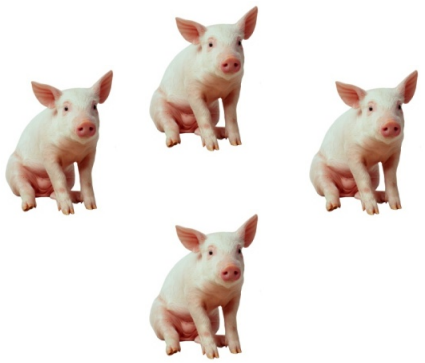
- ✓ Easily integrated into current systems for routine BLUP.
- ✓ Estimate EBV for genotyped and non-genotyped.
- ✓ It can be applied in any model (multiple traits).
- ✓ The number of parameters do not increase with the number of markers.

Single-Step Genomic Evaluation

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

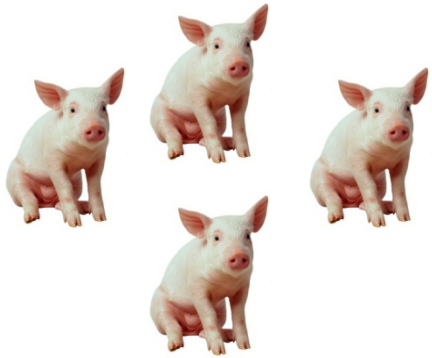
Same (co)variance components

Not genotyped



Relationships are defined with respect to a base population

Genotyped



Allele frequencies in the base population??

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

Single-Step Genomic Evaluation

	Average diagonal	*Average off-diagonal*	*Additive variance (se)*	*Accuracy GEBV (PEV)*
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!!

Forni et al. (2011)

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

Vitezica et al. (2011)

$$\mathbf{G}^{*} = \left(1 - \frac{1}{2}\alpha\right)\mathbf{G} + \mathbf{11}'\alpha$$

Christensen et al. (2012)

$$G_a = \beta G + \alpha,$$

where β and α solved the system of equations

$$\text{Avg}(\text{diag}(G))\beta + \alpha = \text{Avg}(\text{diag}(A_{11})),$$
$$\text{Avg}(G)\beta + \alpha = \text{Avg}(A_{11}).$$

We have to recognize when a strategy is safe to be implemented, but be aware that it will change quickly.

Single-Step Genomic Evaluation

- Phenotypes collected until 2009
- EBV accuracy for progeny: parents were genotyped 60k

TRAIT	Progeny of sires and dams born before 2007 (n=2,180)				Progeny of sires and dams born in 2007 and 2008 (n=227)			
	# of progeny	ACC EBV	ACC GEBV	increase	# of progeny	ACC EBV	ACC GEBV	increase
Total Number born	34,233	0.39	0.42	7%	4,881	0.35	0.39	11%
Stillborn	30,967	0.41	0.43	5%	3,839	0.33	0.37	12%
Survival birth - weaning	5,285	0.30	0.33	10%	1,096	0.23	0.29	26%
Litter weaning weight	5,285	0.40	0.44	10%	1,096	0.32	0.40	25%
Interval weaning - mate	29,703	0.38	0.39	2%	4,232	0.32	0.35	9%

Downside: same EBV for full-siblings

July 2012

Single-Step Genomic Evaluation

- Phenotypes collected until 2009
- Progeny genotyped(60k): animals born in 2009 and after (n=2,023)

TRAIT	ACC EBV	ACC GEBV	increase
Total Number born	0.25	0.42	68%
Stillborn	0.26	0.43	65%
Survival birth - weaning	0.17	0.26	53%
Litter weaning weight	0.23	0.35	52%
Interval weaning - mate	0.17	0.30	76%

July 2012

Imputation of HD from LD Panel

- **AlphaImpute**[1]
  - Combines simple phasing rules, long-range phasing, haplotype libraries, segregation analysis and recombination modeling

Imputation accuracy for two low-density panels

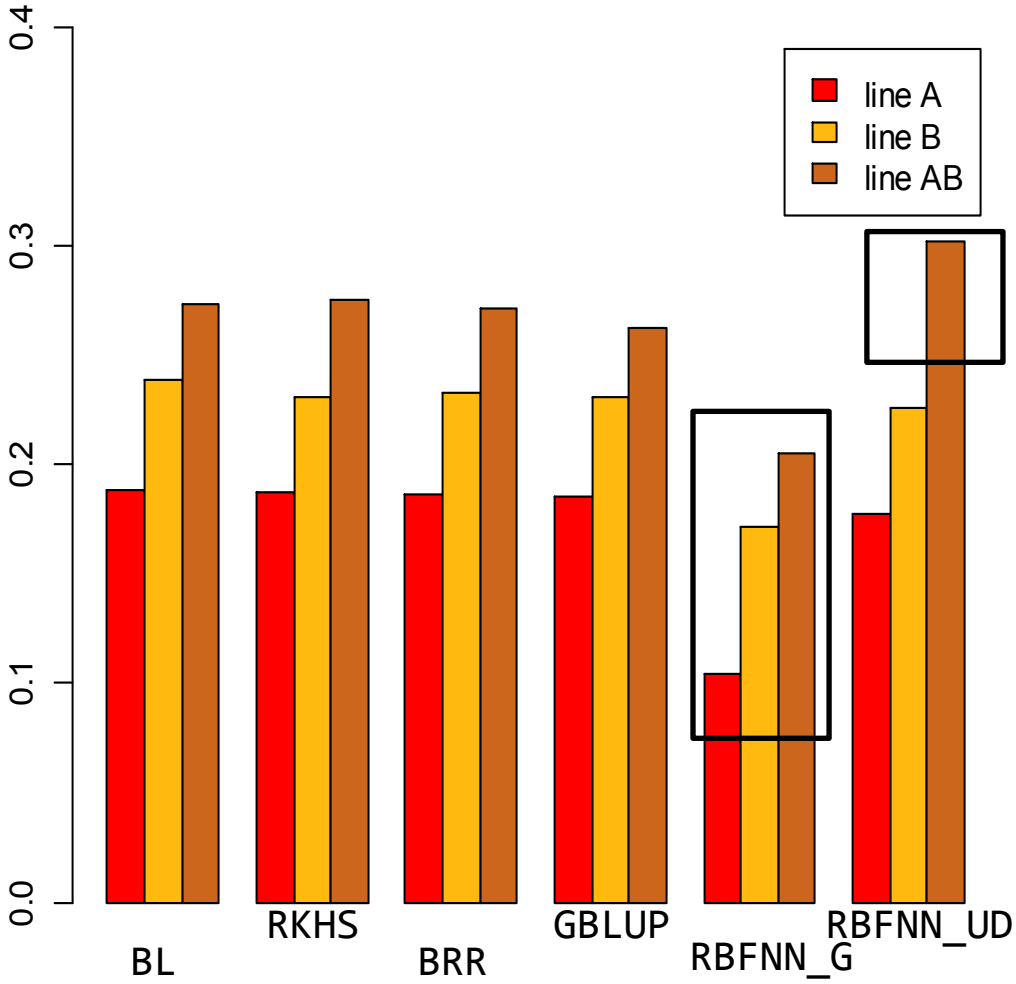
3k	384	
0.99	0.96	Both Parents
0.98	0.89	Sire and MGS
0.99	0.94	Dam and PGS
0.98	0.87	Sire
0.97	0.87	Dam
0.95	0.81	Other

PIC will use between 400 and 450 markers for a low-density equally spaced panel

[1]Hickey et al., 2012

A comparison of methods for predicting litter size in commercial pig lines
Tusell et al. (2012)

Average correlation between observed and predicted phenotypes in testing sets (**r**)



Within line, methods yielded similar predictive ability, except for RBFNN models.

1) Bayesian Lasso (**BL**)

2) Reproducing Kernel Hilbert Spaces with kernel averaging (**RKHS**)

3) Bayesian Ridge Regression (**BRR**)

4) Genomic BLUP (**GBLUP**)

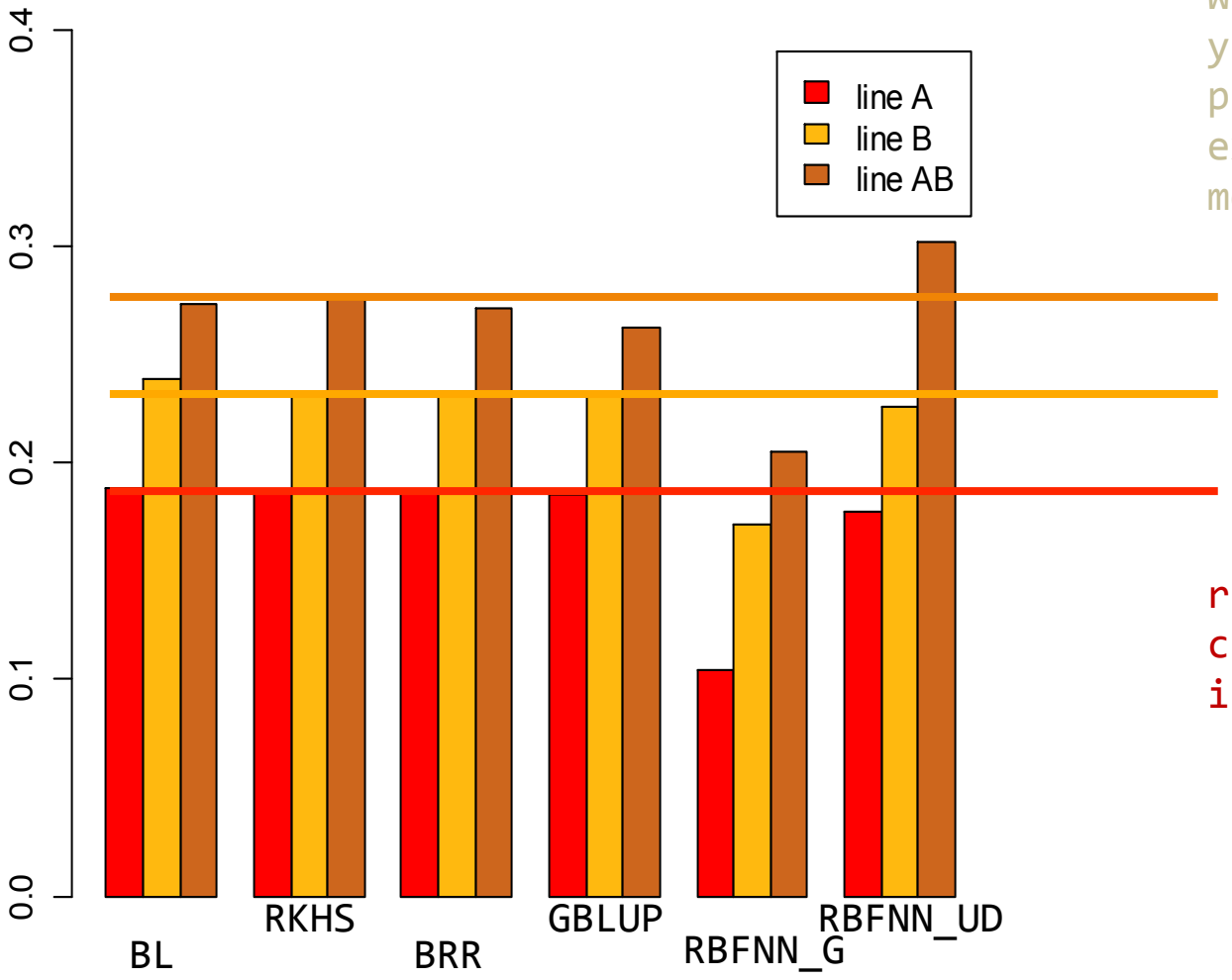
Radial Basis Functions Neural Networks using:

5) The additive genomic relationship matrix (**RBFNN_G**)

6) Principal component scores of the SNP matrix (**RBFNN_UD**)

A comparison of methods for predicting litter size in commercial pig lines
Tusell et al. (2012)

Average correlation between observed and predicted phenotypes in testing sets (**r**)



Within line, methods yielded similar predictive ability, except for RBFNN models.

AB r ≈ 0.27

B r ≈ 0.23

A r ≈ 0.19

r higher in the crossbred line than in purebred lines

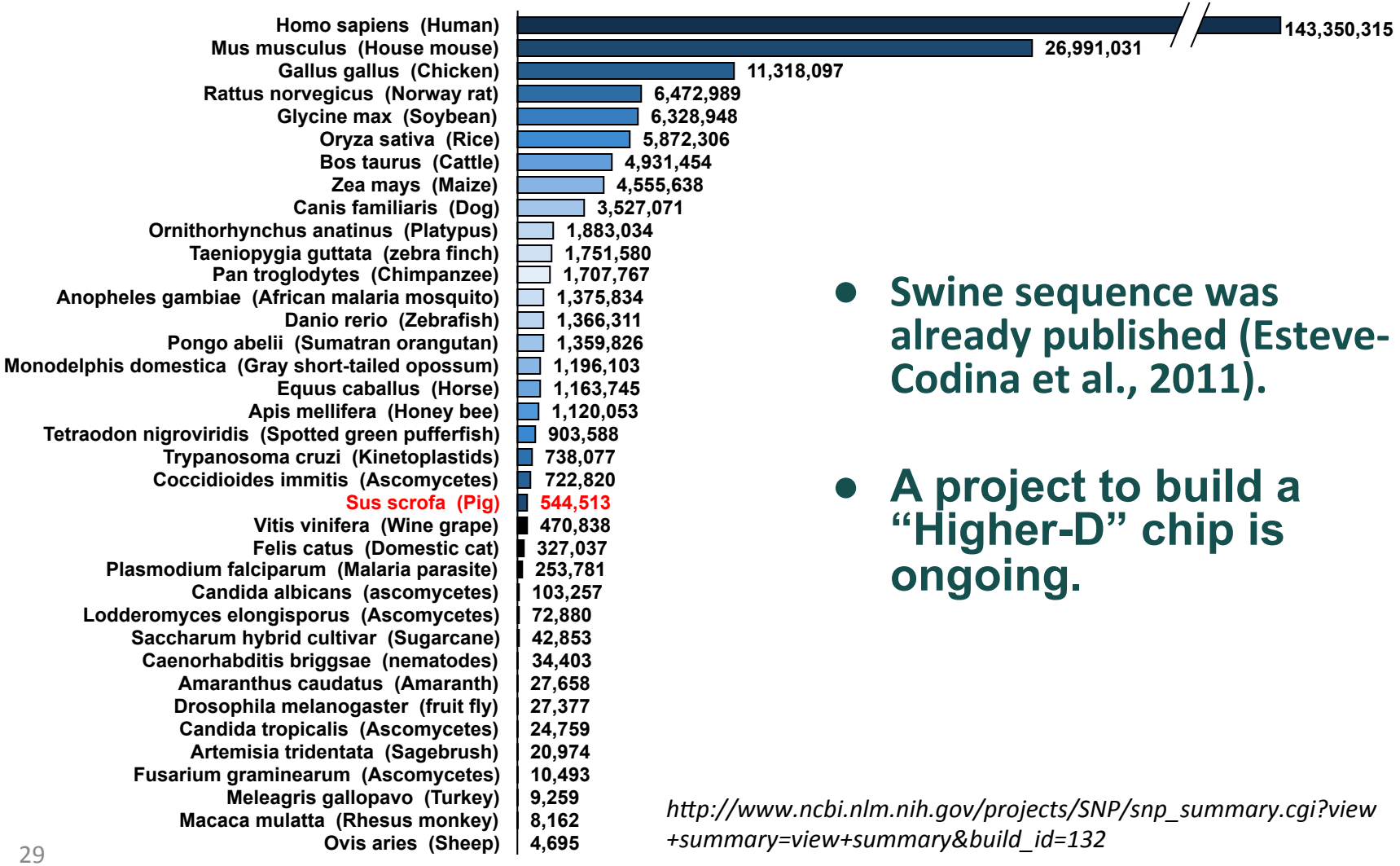
Overview

- Moved away from trait-specific LD panels that were enabled by HD genotyping in swine commercial breeding programs.

- HD genomic information of parents is used routinely for multiple trait evaluations.

- Imputation is being used to generate HD genomic information for all selection candidates.

Future



- Swine sequence was already published (Esteve-Codina et al., 2011).
- A project to build a “Higher-D” chip is ongoing.

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi?view+summary=view+summary&build_id=132

Thank you!