

What's next for dairy cattle breeding?

Gregor Gorjanc

John M Hickey, Janez Jenko, Ino Čurik, Ingelin Steinsland, Finn Lindgren,
Marija Špehar, Jana Obšteter, Owen Powell, Maria Lie Selle, Vladimir Brajković

@GregorGorjanc

A dramatic landscape photograph of a mountain range. The foreground shows a grassy slope with patches of brown and green. In the middle ground, a large, dark, jagged mountain peak is partially obscured by thick, grey, swirling clouds. The sky is filled with heavy, overcast clouds, creating a moody and atmospheric scene. The overall tone is somber and intense, reflecting the 'storm' metaphor in the text.

What's next from me?

vs.

What's the next BIG thing?

Roadmap

What's next from me?

Opening out the breeding bottleneck

Developing world

Multi-breed & sequence modelling

What's the next BIG thing?



Opening out the breeding bottleneck

- Genomic selection increased turnover of germplasm
- Breeding programs should monitor trends of
 - Genetic means
 - Genetic variances and covariances
- Genomic markers make this computationally tractable and more informative (genome regions)

→ Do you need to change breeding strategy, implement optimal contribution selection, ...?



Variance parameter and Genetic variance

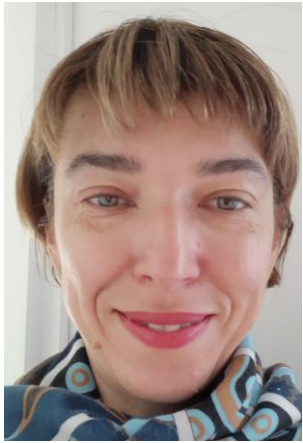
- Pedigree-based model
 - Variance parameter = base population (additive) genetic variance
 - Is base population a “coherent” time point?
- Marker-based model (SNP-BLUP & GBLUP)
 - Variance parameter = variance of allele substitution effects
 - Which time point? (= the genotype centring value)
- **We want genetic variance for a specific group of animals!**
(by year of birth, breeders vs. producers, different breeders, ...)

Genetic variance of a subset of animals

- Sampling and exact approaches
 - Pedigree-based model (Sorensen et al., GR 2001)
 - Marker-based model (Lehremeier et al., JABG 2017; Schreck et al., BioRxiv, 2019)
- Tedious with pedigree-based model
- Easy with marker-based model
 - Sample marker effects $\mathbf{m}^i | \mathbf{y} \sim N(E(\mathbf{m} | \mathbf{y}), \text{Var}(\mathbf{m} | \mathbf{y}))$
 - For a subset of animals j
 - Calculate breeding values $\mathbf{a}_j^i = \mathbf{W}[\text{animals}(j), :] \mathbf{m}^i$
 - Variance $\sigma_j^{2,i} = \text{Var}(\mathbf{a}_j^i)$

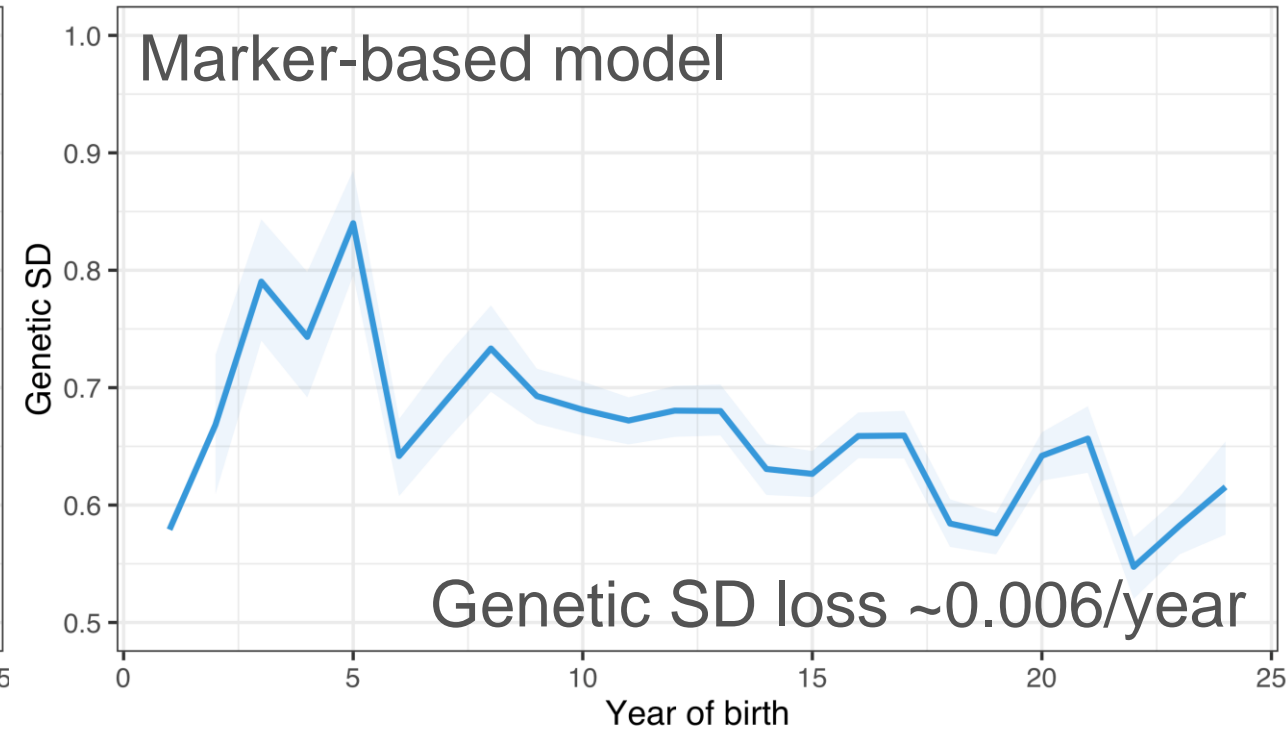
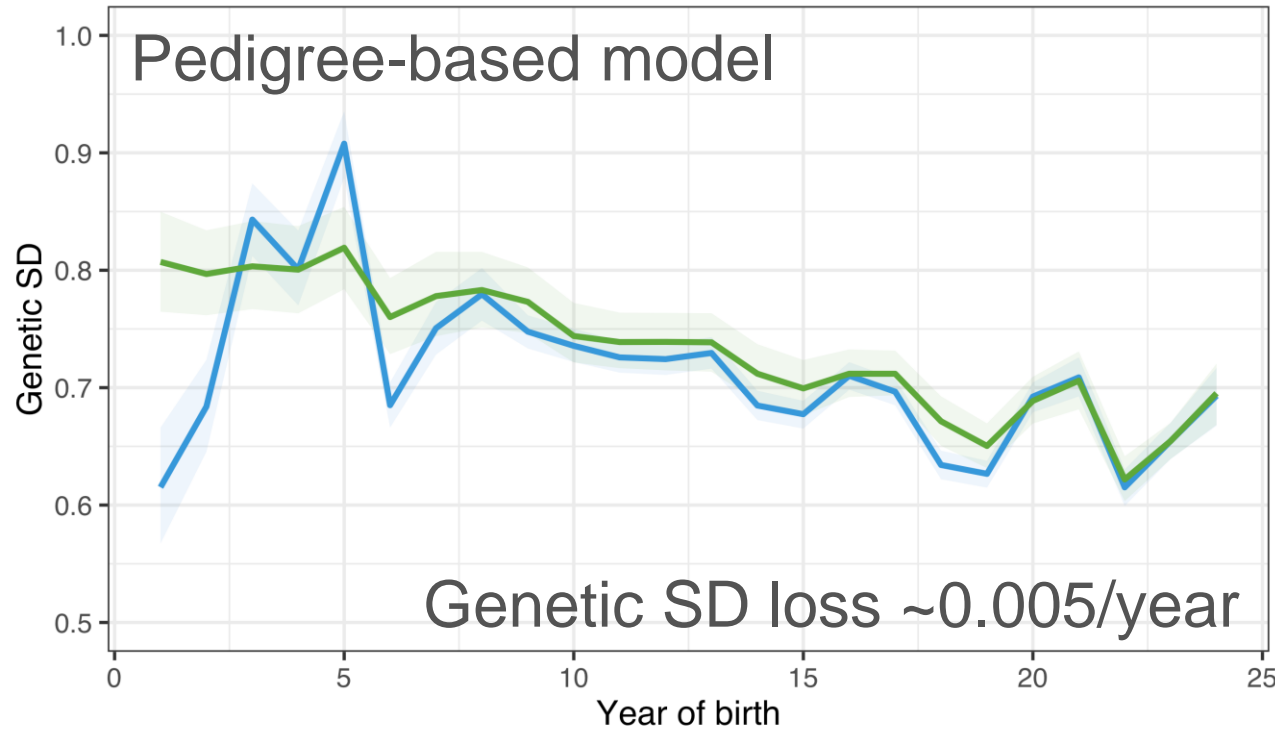
An example – trend for genetic stand. dev.

~9K bulls with DYD, ~60K pedigree, ~40K SNP markers



Group — Bulls — Cows & Bulls

Group — Bulls — Cows & Bulls



Trait specific effective population size for bulls: 33.8 (30.0, 39.5)

Genetic variance of a subset of genome!

- Easy with marker-based model

- Sample marker effects $\mathbf{m}^i | \mathbf{y} \sim N(E(\mathbf{m} | \mathbf{y}), \text{Var}(\mathbf{m} | \mathbf{y}))$

- For a subset of genome k

- Calculate breeding values $\mathbf{a}_{-k}^i = \mathbf{W}[:, \text{markers}(-k)] \mathbf{m}^i [\text{markers}(-k)]$

- Variance $\sigma_k^{2,i} = \text{Var}(\mathbf{a}_k^i)$

- **NOTE: Variance of a sum!**

- Genetic variance

- Genic variance $2p_i q_i m_i^2$

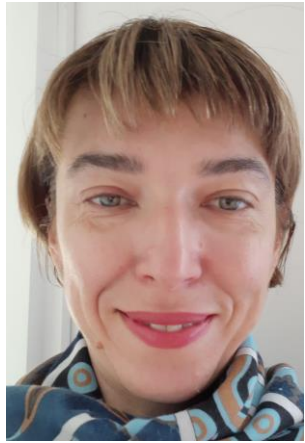
- Linkage-Disequilibrium

$$\text{Var} \left(\begin{bmatrix} \mathbf{a}_k^i \\ \mathbf{a}_{-k}^i \end{bmatrix} \right) = \begin{bmatrix} \sigma_k^{2,i} & \sigma_{k,-k}^i \\ \sigma_{k,-k}^i & \sigma_{-k}^{2,i} \end{bmatrix}$$

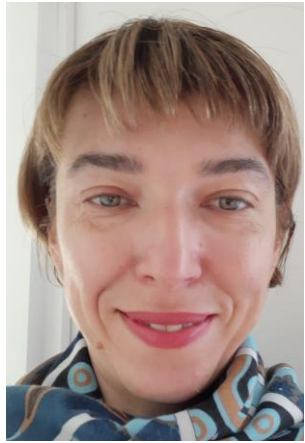
$$\text{Var}(\mathbf{a}^i) = \sigma_k^{2,i} + \sigma_{-k}^{2,i} + 2\sigma_{k,-k}^i$$

Genomic analysis of genetic correlation

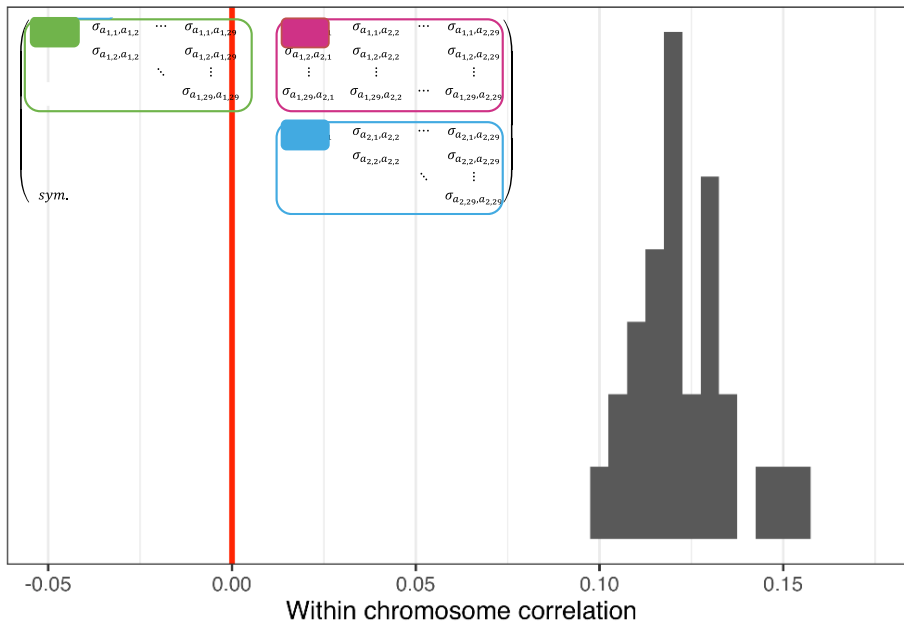
- Two traits
 - Genetic correlation: -0.06
 - Allele substitution effects correlation: +0.15



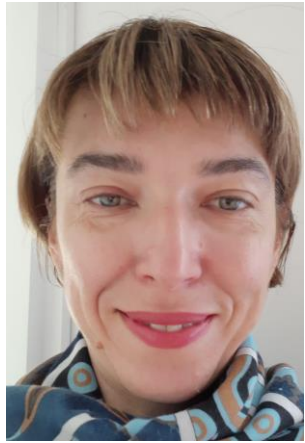
Genomic analysis of genetic correlation



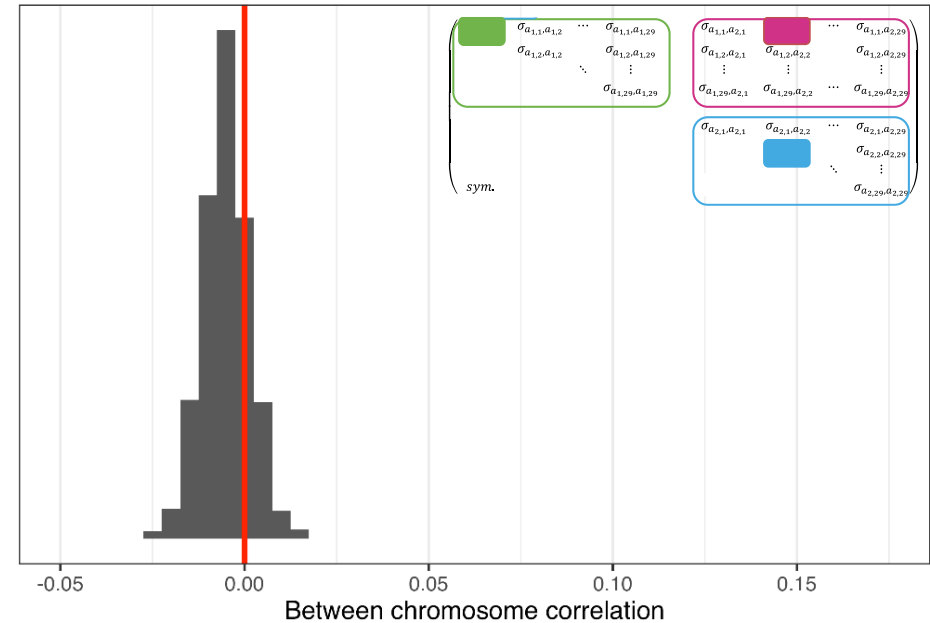
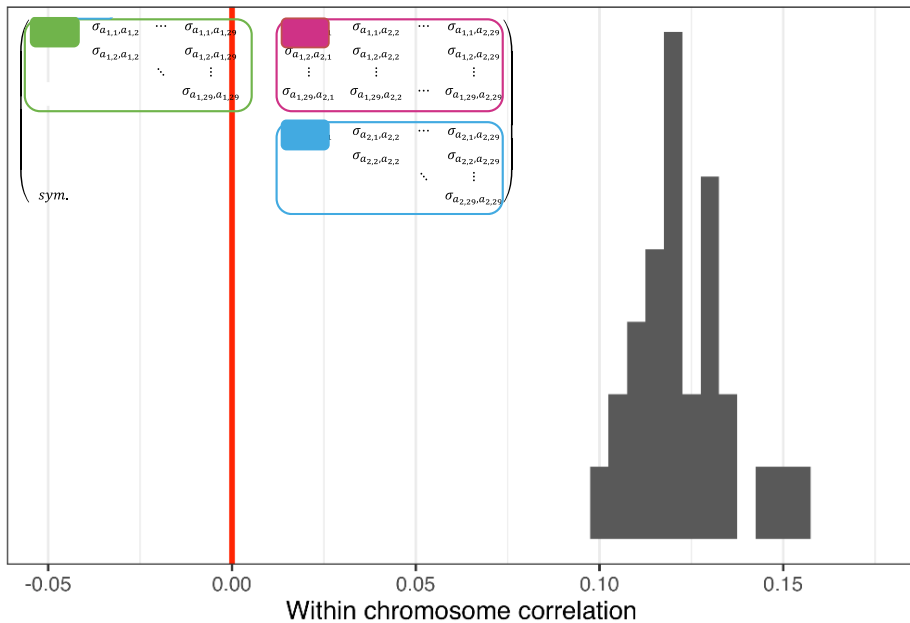
- Two traits
 - Genetic correlation: -0.06
 - Allele substitution effects correlation: +0.15
 - Individual chromosomes (n=29): +0.120 [0.099, 0.154] → +0.41



Genomic analysis of genetic correlation



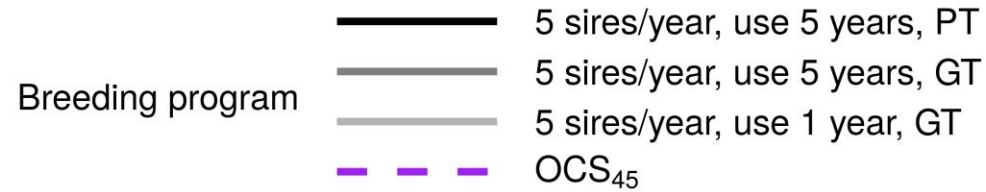
- Two traits
 - Genetic correlation: -0.06
 - Allele substitution effects correlation: +0.15
 - Individual chromosomes (n=29): +0.120 [0.099, 0.154] → +0.41
 - Pairs of chromosomes (n=812): -0.005 [-0.024, 0.016] → -0.47



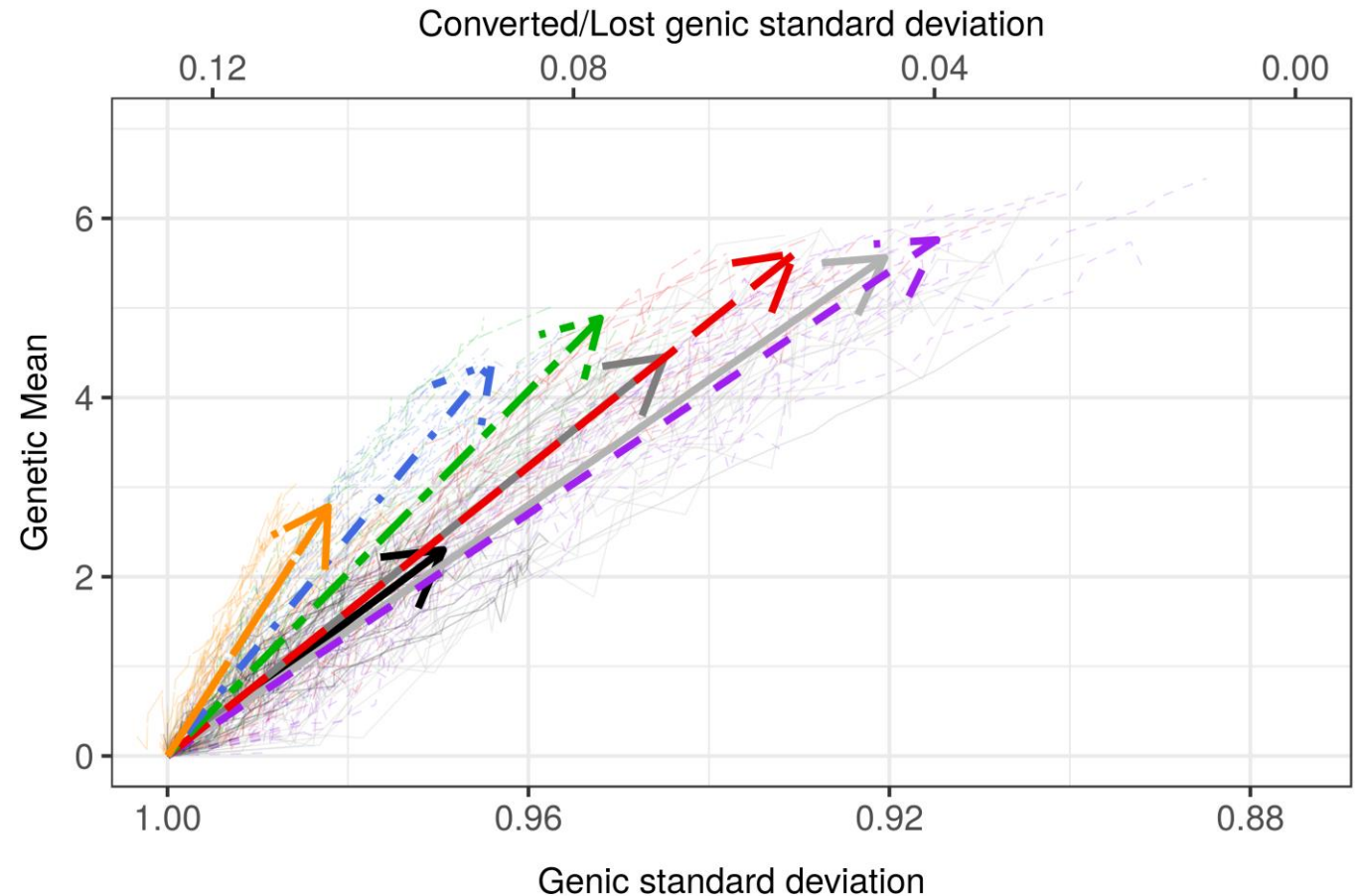
Opening out the breeding bottleneck



- Monitor genetic means & (co)variances



- Actively manage it
 - Optimal contribution selection!!!!
 - Mate allocation
(↓inbreeding depression,
↑within-family variance)



What about dominance?

- Data often supports “infinitesimal dominance”
(small variance, but significant inbreeding depression)
- Directional dominance model works very well!
Varona et al. (2018) GSE: 10.1186/s12711-018-0374-1
- Many genotyped & phenotyped cows
→ accurate a & d marker parameters
- Upgrade SNP-MACE to a directional dominance model
→ population specific allele sub. effects: $m = a + d * (q - p)$

Roadmap

What's next from me?

Opening out the breeding bottleneck

Developing world

Multi-breed & sequence modelling

What's the next BIG thing?

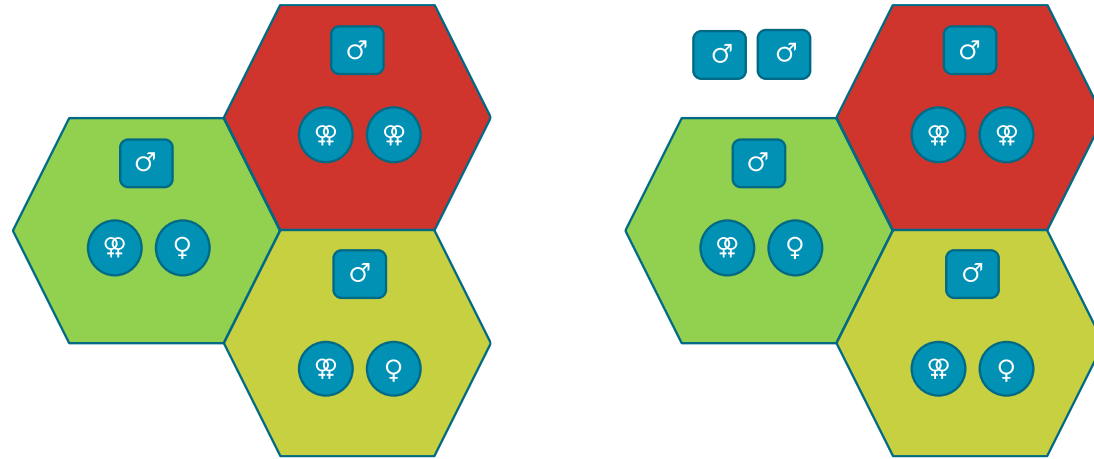


Genomic selection in the developing world

- “Phenotype and genotype some cows and you can do genomic selection”
- Small-holder farms are small & very variable
→ hard to separate genetics from environment
- A solution?
 - borrow information from neighbours (spatial model) or
 - measure key environmental indicators

Simulation

- Connectedness scenarios

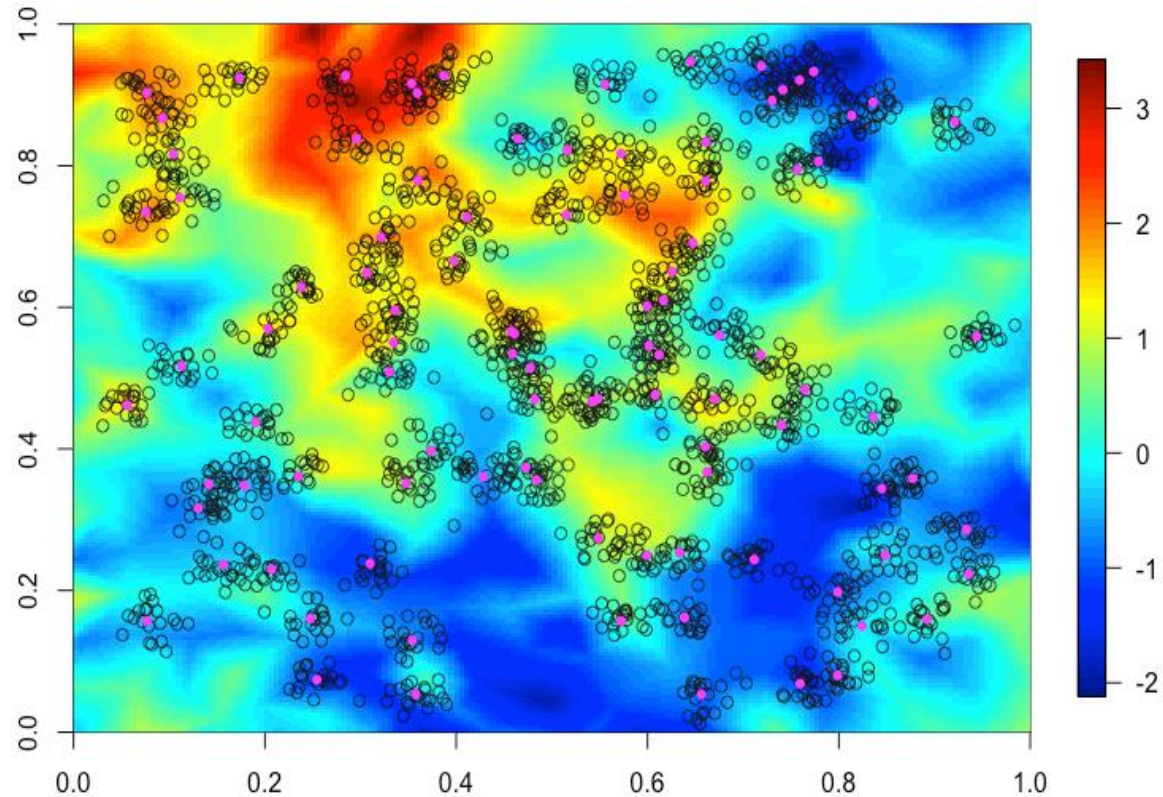


- Phenotype = Location + Herd + Genetics + Noise
0.40 0.25 0.10 0.25

- Location = sum(10 spatially varying covariates)
- Observe 5 covariates with noise and 2 as binary indicators

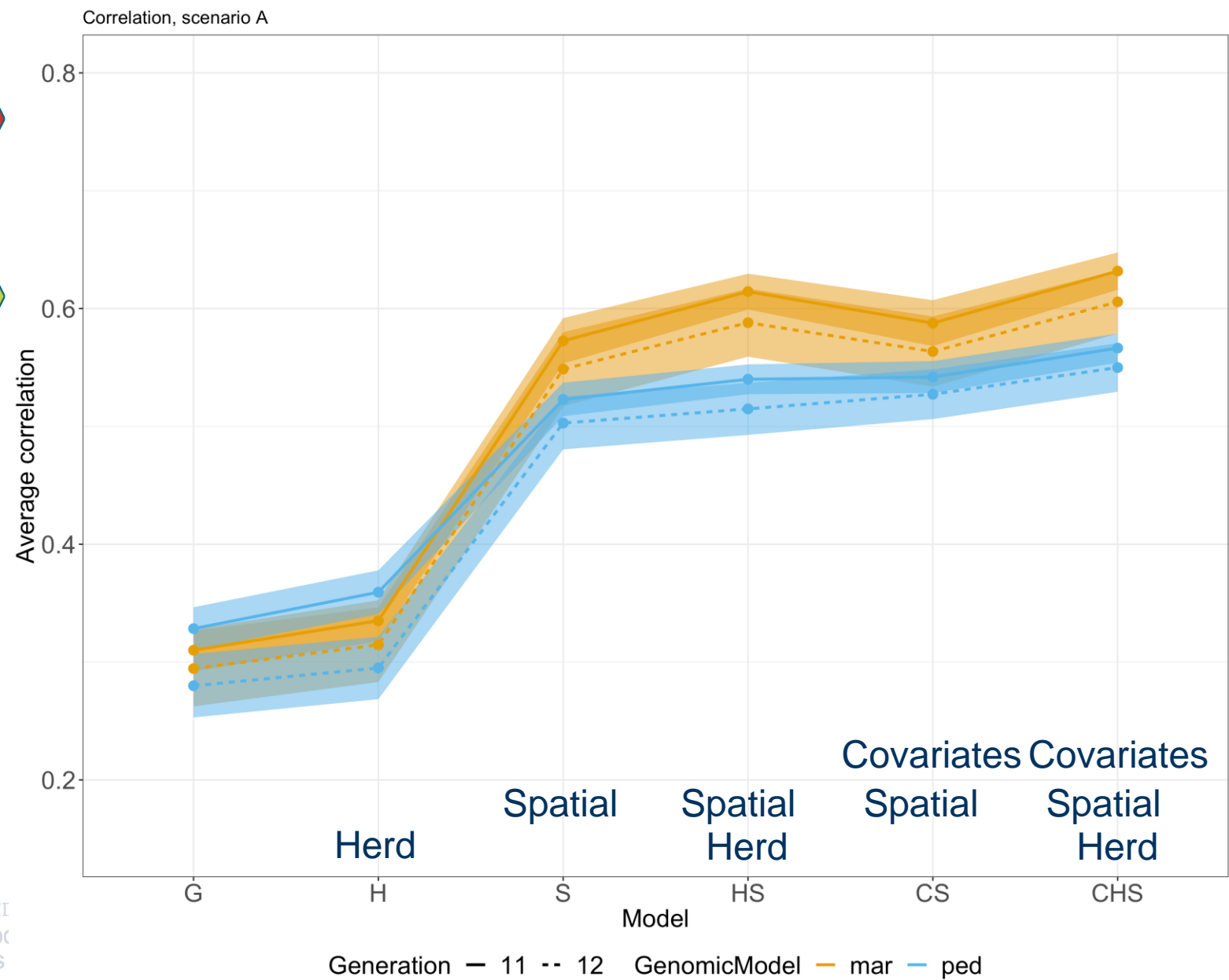
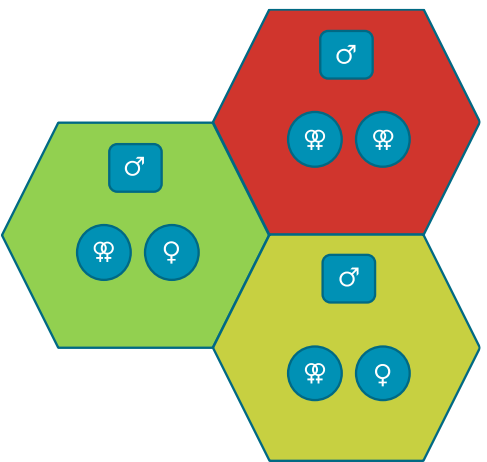


Spatial data and modelling

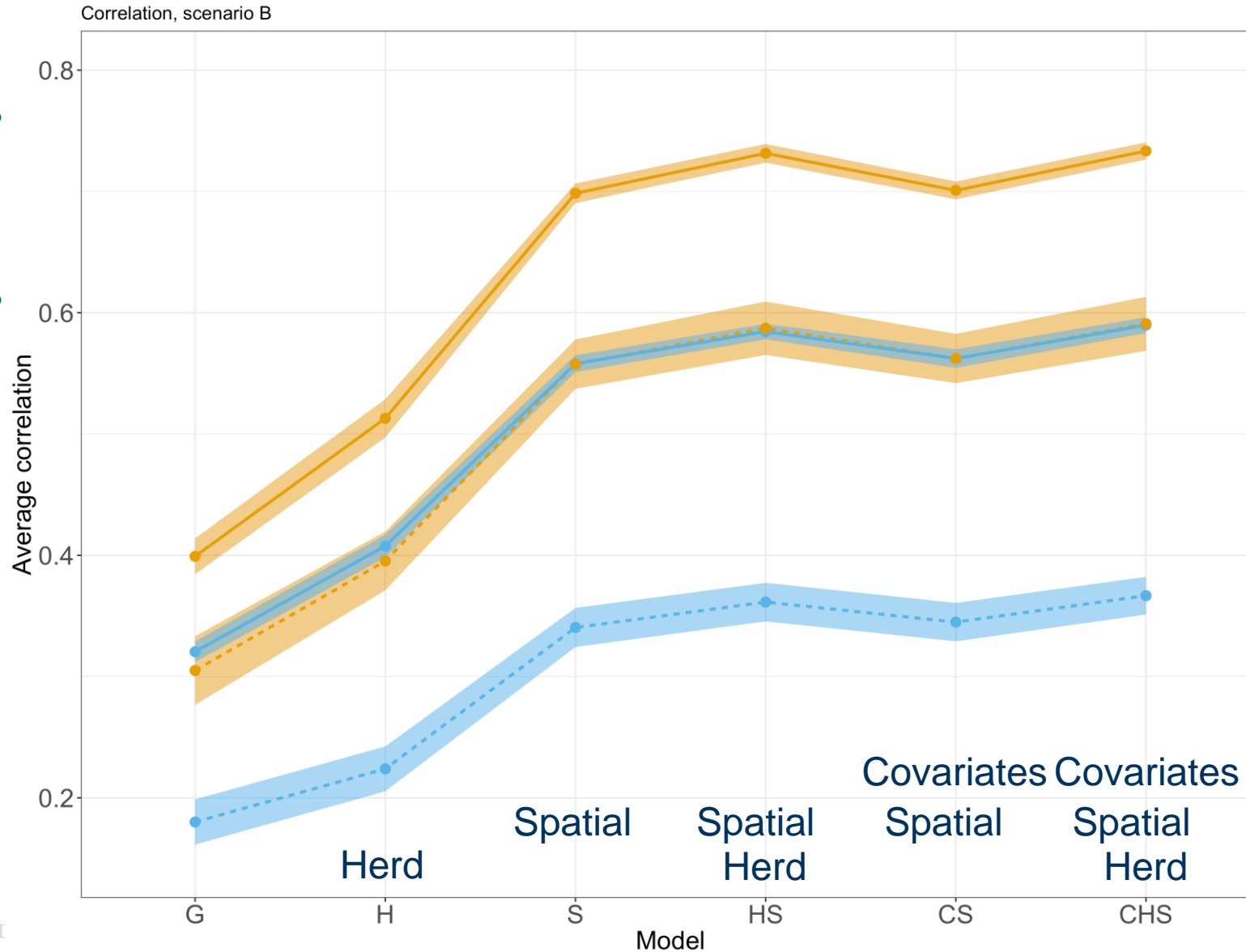
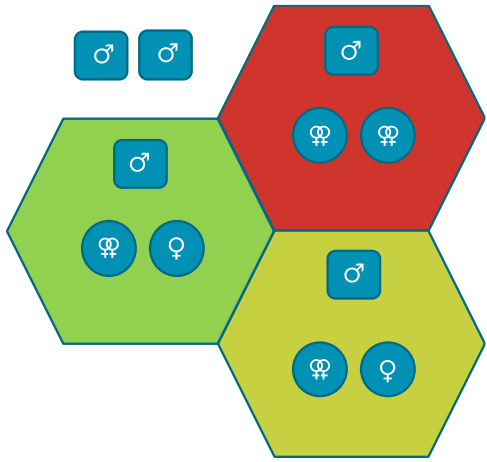


- Location-based geostatistical model
$$\mathbf{l} \sim N(\mathbf{0}, \text{Matérn}(\kappa, \sigma_l^2))$$
- SPDE approach of the INLA program

Accuracy of evaluation & prediction



Accuracy of evaluation & prediction



Roadmap

What's next from me?

Opening out the breeding bottleneck

Developing world

Multi-breed & sequence modelling

What's the next BIG thing?



Genomics research trends

- Whole-genome sequencing on the rise
 - Roslin & Genus/PIC sequenced ~8,000 pigs at low-coverage and accurately imputed ~20-60M SNP into ~375,000 pigs
 - Some dairy programs routinely sequence AI bulls
- Pan-genomes
 - Roslin/CTLGH pan-genome of *Bos Taurus* and *Bos Indicus* breeds
- We need algorithms & models that handle
 - the scale of the data (#animals & #variants)
 - SNPs, indels, CNVs, ... on autosomal, sex & mitochondrial chr.

Genomics research trends – succinct trees

Kelleher et al. (2018) BioRxiv: 10.1101/458067



A Conventional data storage

Sample of size n

site	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
1				0	0
2	0	0	0		
3	0	0	0	0	
4		0	0	0	0
5				0	0
6	0			0	0
7	0	0	0		

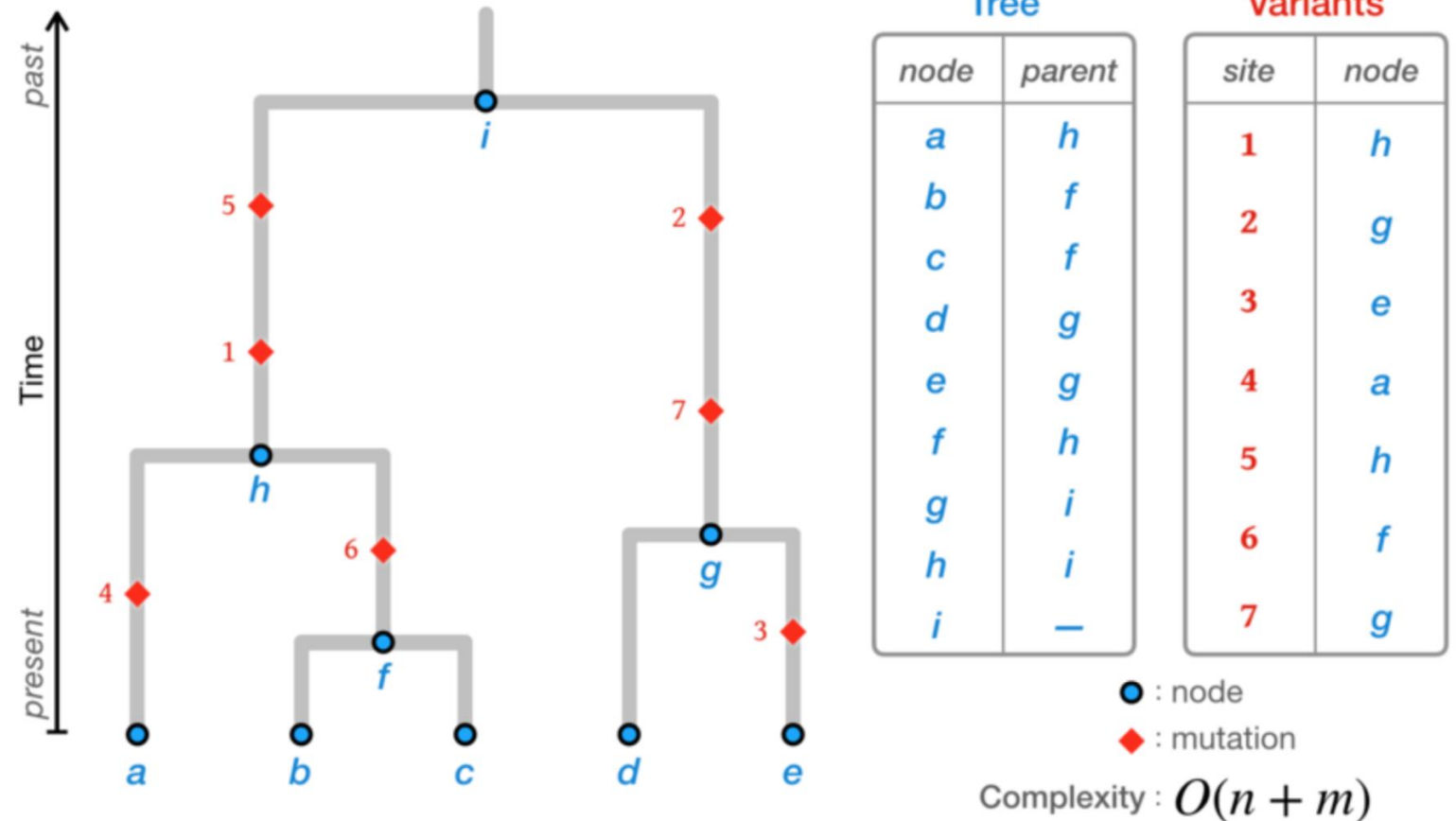
Variant sequence of length m

0 : ancestral allele

| : derived allele

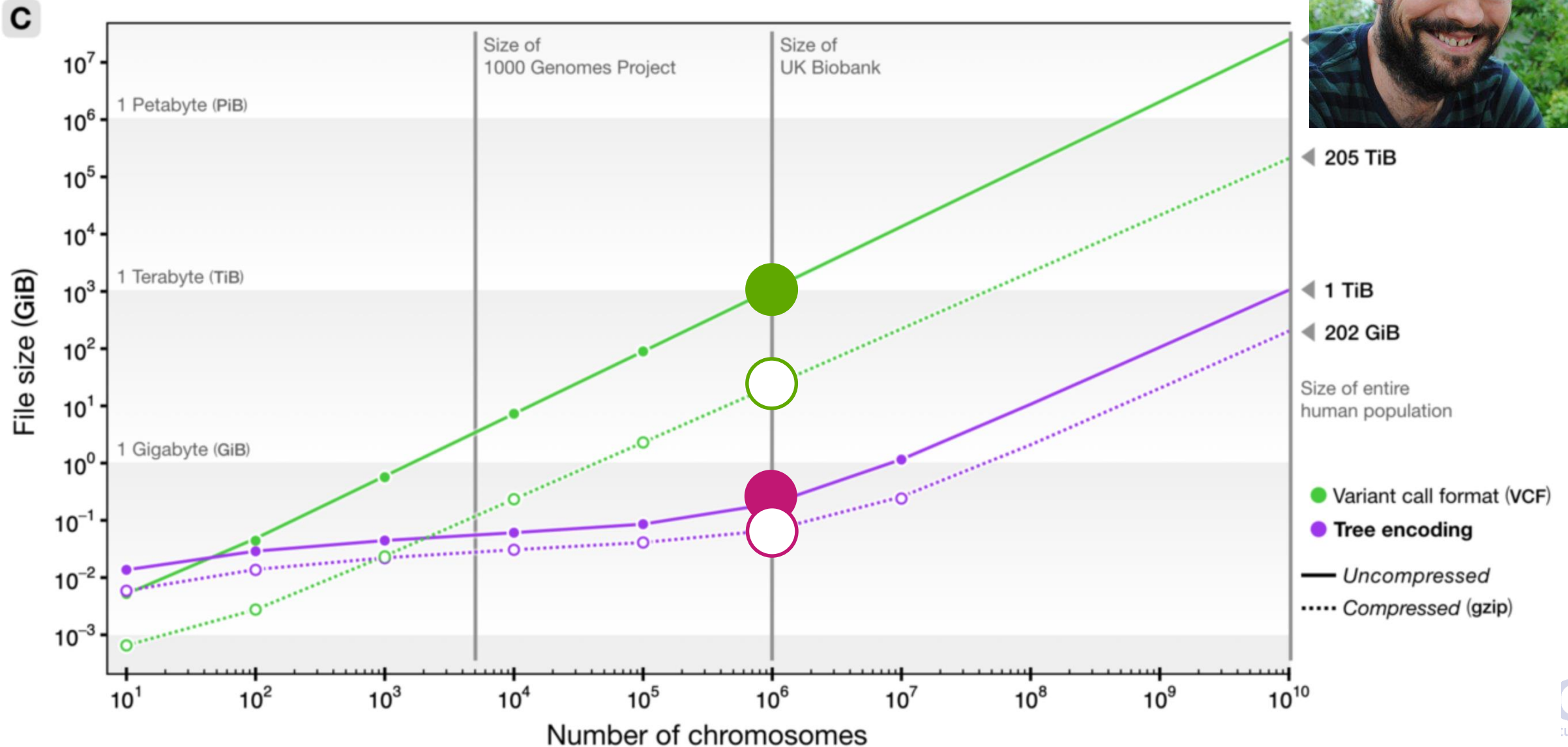
Complexity : $O(n \times m)$

B Tree encoding



Genomics research trends – succinct trees

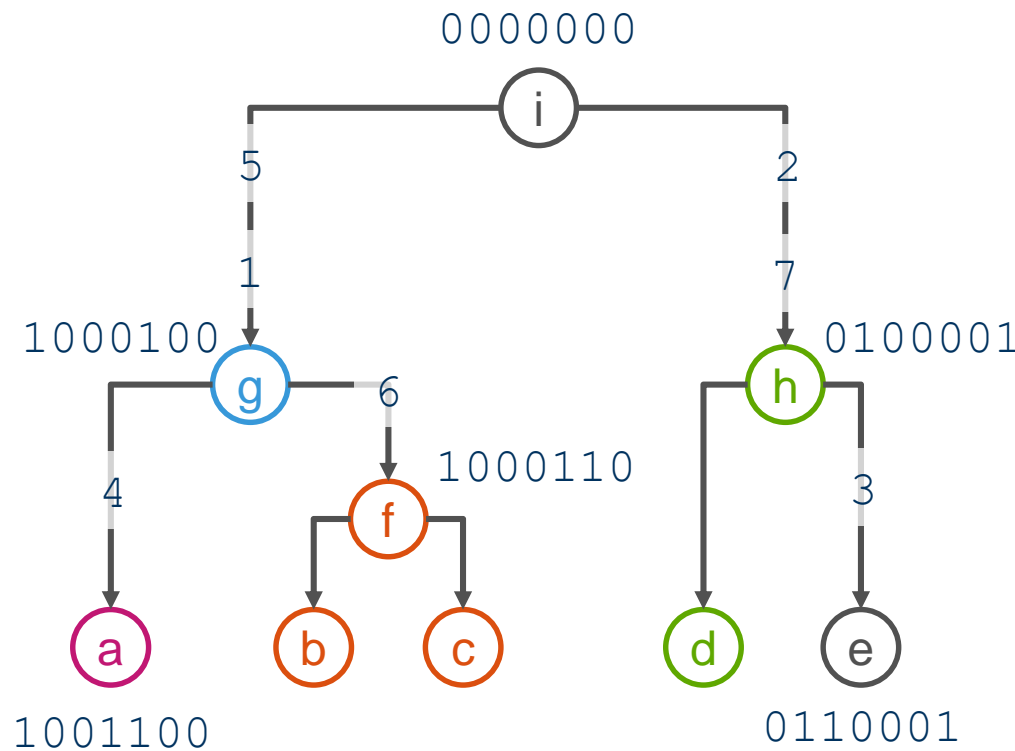
Kelleher et al. (2018) BioRxiv: 10.1101/458067



Tree-Auto-Regressive model



- Progeny/child haplotypes likely have similar effect
- Parameters: average correlation α and variance v



$$h_i \sim N(0, v)$$

$$h_{g'}|h_i \sim N(\alpha h_i, v_c)$$

$$h_g|h_{g'} \sim N(\alpha h_{g'}, v_c)$$

$$h_a|h_g \sim N(\alpha h_g, v_c)$$

$$h_f, h_b, h_c|h_g \sim N(\alpha h_g, v_c)$$

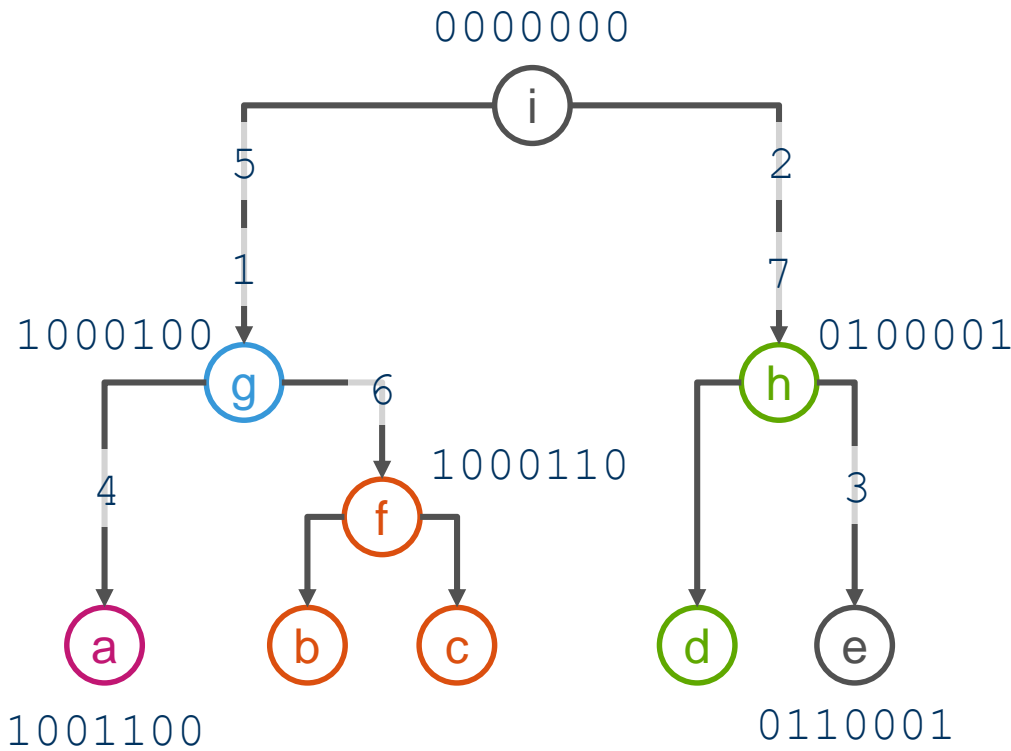
$$h_{h'}|h_i \sim N(\alpha h_i, v_c)$$

$$h_h, h_d|h_{h'} \sim N(\alpha h_{h'}, v_c)$$

$$h_e|h_h \sim N(\alpha h_h, v_c)$$

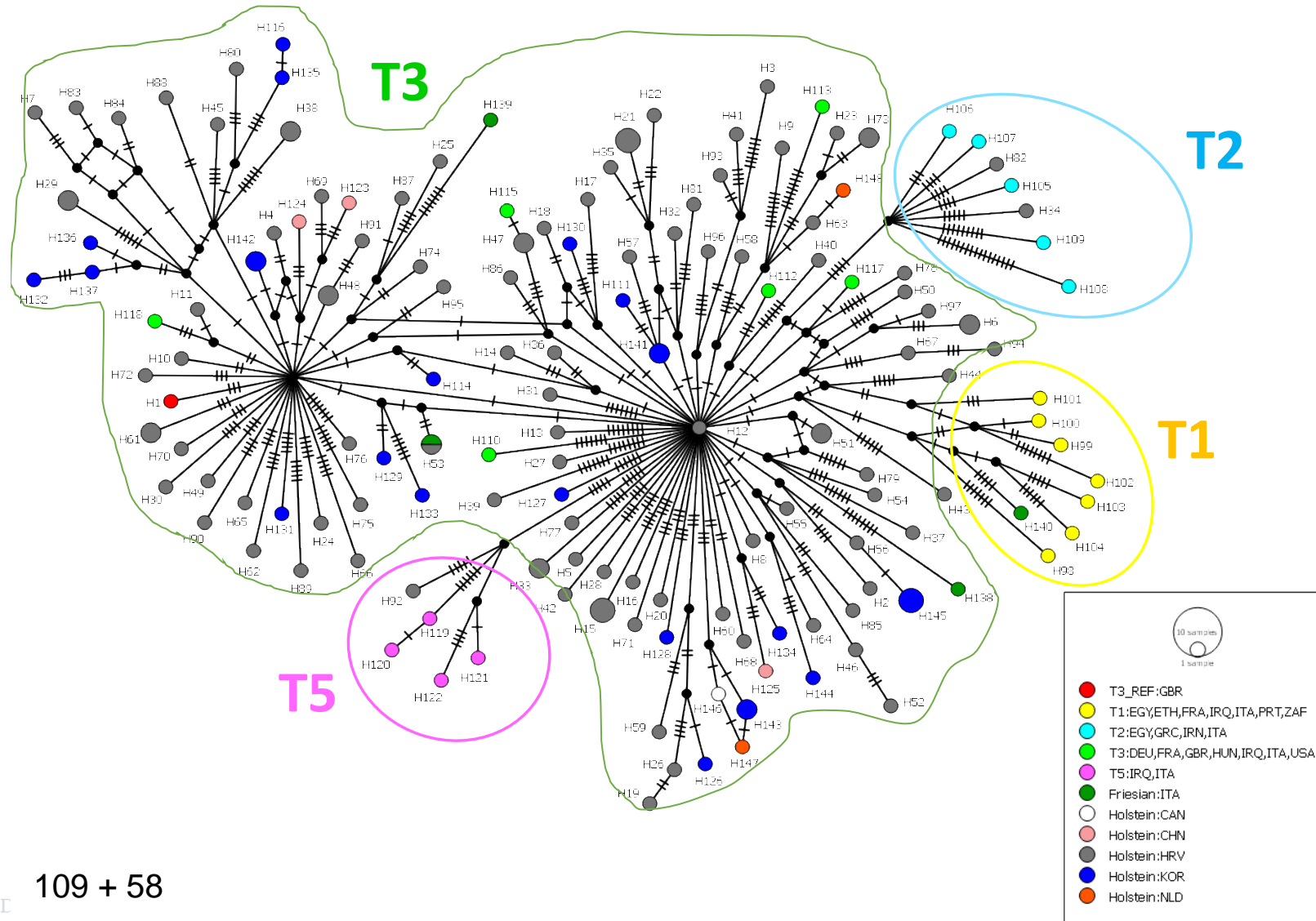
Tree-Auto-Regressive model

- Sparse and simple matrix inverse



$$\begin{matrix}
 h_i & h'_g & h_g & h_a & h_f & h'_h & h_h & h_i \\
 \left(\begin{array}{cccccccc}
 1 + \alpha^2 + \alpha^2 & -\alpha & & & & & & \\
 -\alpha & 1 + \alpha^2 & & & & & & \\
 & -\alpha & 1 + \alpha^2 + \alpha^2 & -\alpha & -\alpha & & & \\
 & & -\alpha & 1 & & & & \\
 & & -\alpha & & 1 & & & \\
 & -\alpha & & & & 1 & -\alpha & \\
 & & & & & -\alpha & 1 & -\alpha \\
 & & & & & & -\alpha & 1
 \end{array} \right)
 \end{matrix}$$

Tree-Auto-Regressive model



109 + 58



THE UNIVERSITY of ED
Royal (Dick) School of
Veterinary Studies



Tree-Auto-Regressive model



- Extensions

- Different types of mutations with more corr. parameters
→ use of biological information

- Account for divergence times between haplotypes
→ Ornstein–Uhlenbeck process

- Recombination!?



A dramatic landscape photograph of a mountain range. The foreground shows a grassy slope with patches of brown and green. In the middle ground, a large, dark, jagged mountain peak is partially obscured by thick, grey, swirling clouds. The sky is filled with heavy, overcast clouds, creating a moody and atmospheric scene. The overall tone is somber and intense, reflecting the 'storm' mentioned in the text.

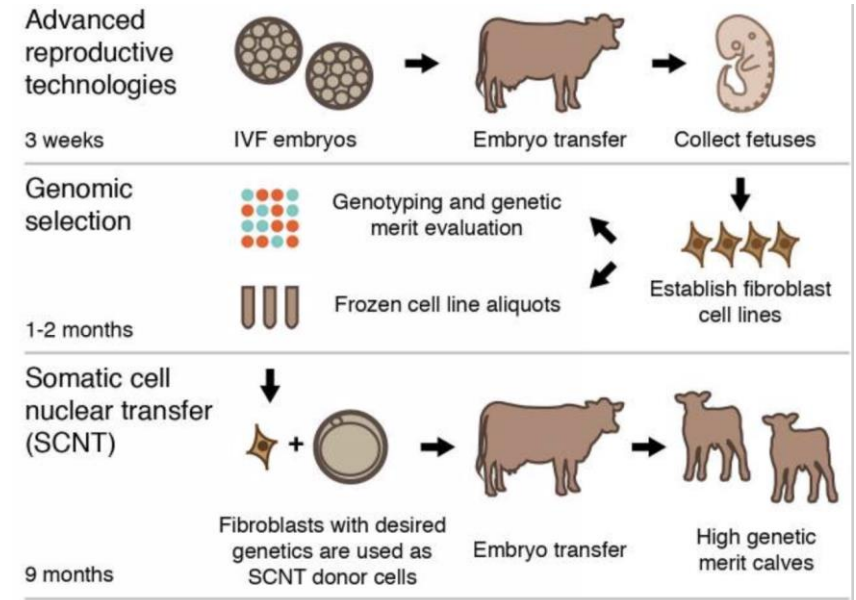
What's next from me?

vs.

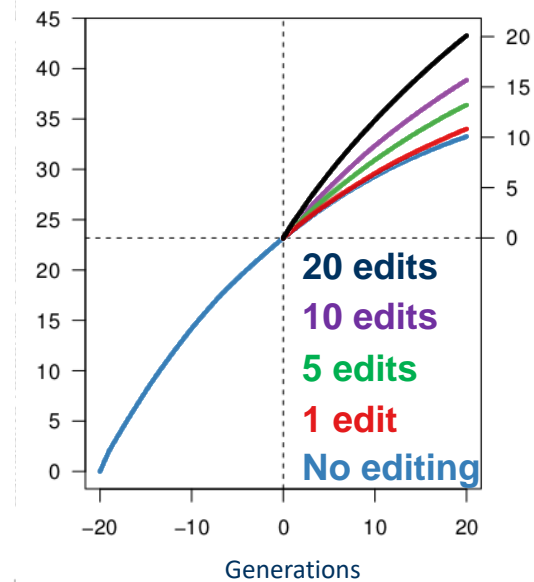
What's the next BIG thing?

What's the next BIG thing?

- Rapid in-vitro embryo tech. with accurate genomic selection on whole-genome sequence
- Genome editing (perhaps even gene drives!)



Kasinathan et al. (2015) SciRep: 10.1038/srep08674



Roadmap

What's next from me?

Opening out the breeding bottleneck

Developing world

Multi-breed & sequence modelling

What's the next BIG thing?



What's next for dairy cattle breeding?

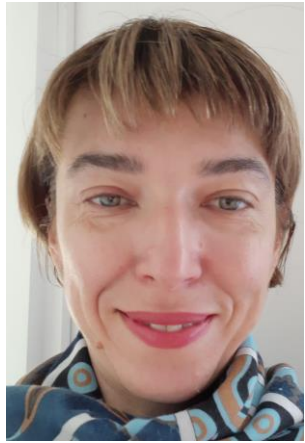
Gregor Gorjanc

John M Hickey, Janez Jenko, Ino Čurik, Ingelin Steinsland, Finn Lindgren,
Marija Špehar, Jana Obšteter, Owen Powell, Maria Lie Selle, Vladimir Brajković

@GregorGorjanc

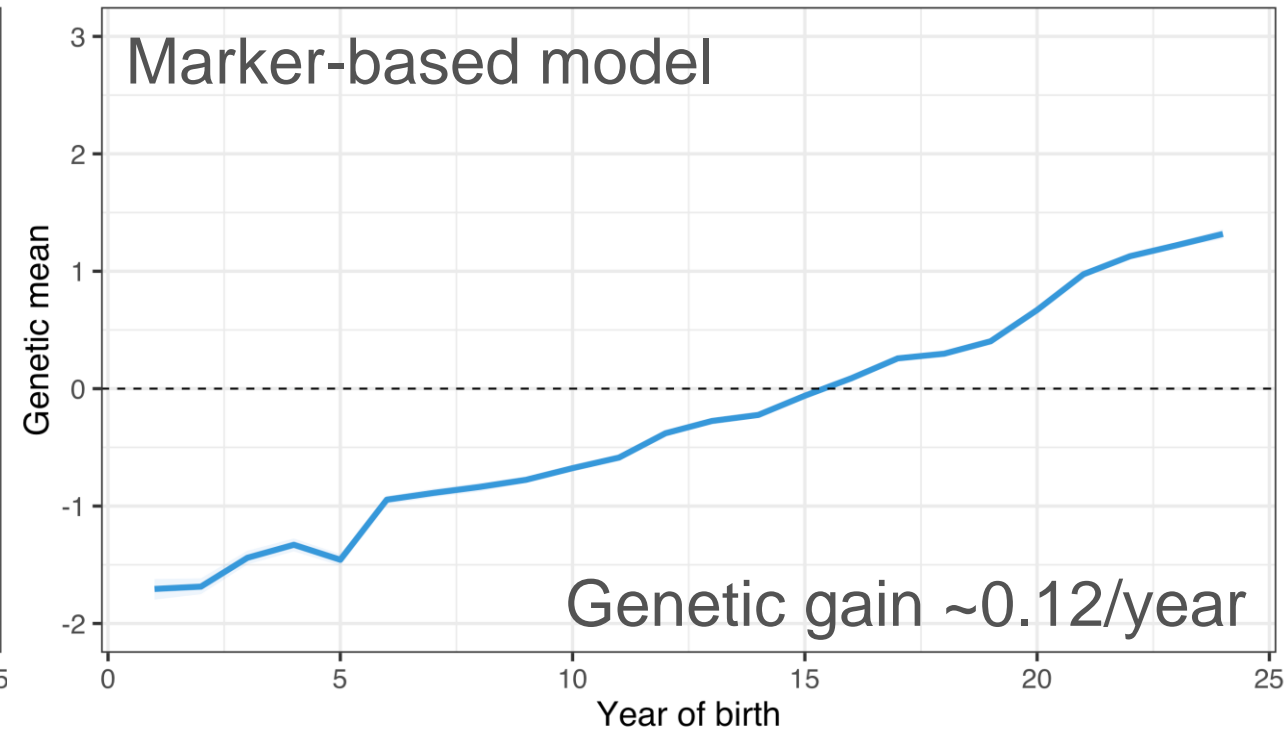
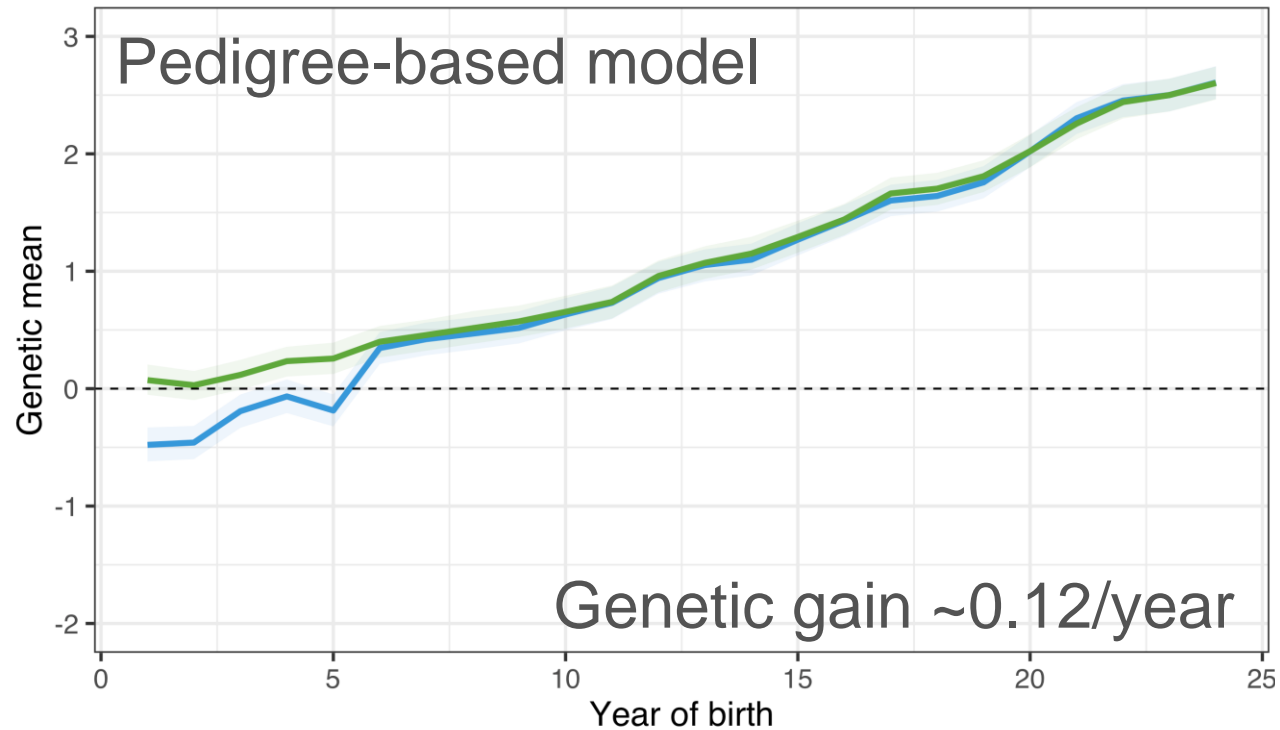
An example – trend for genetic mean

~9K bulls with DYD, ~60K pedigree, ~40K SNP markers

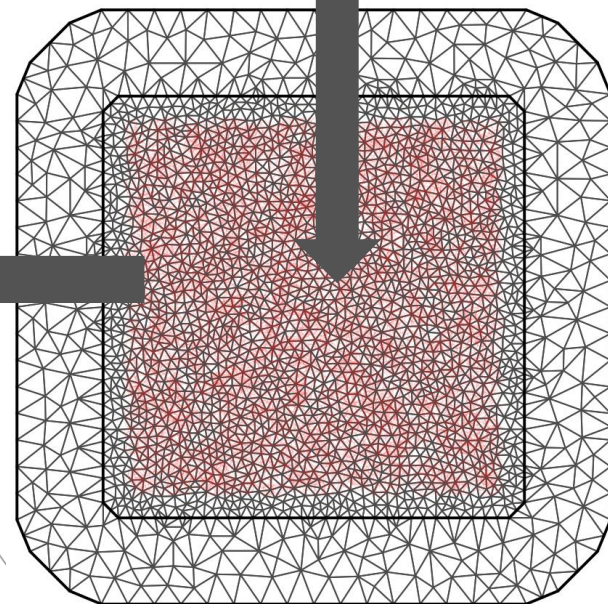
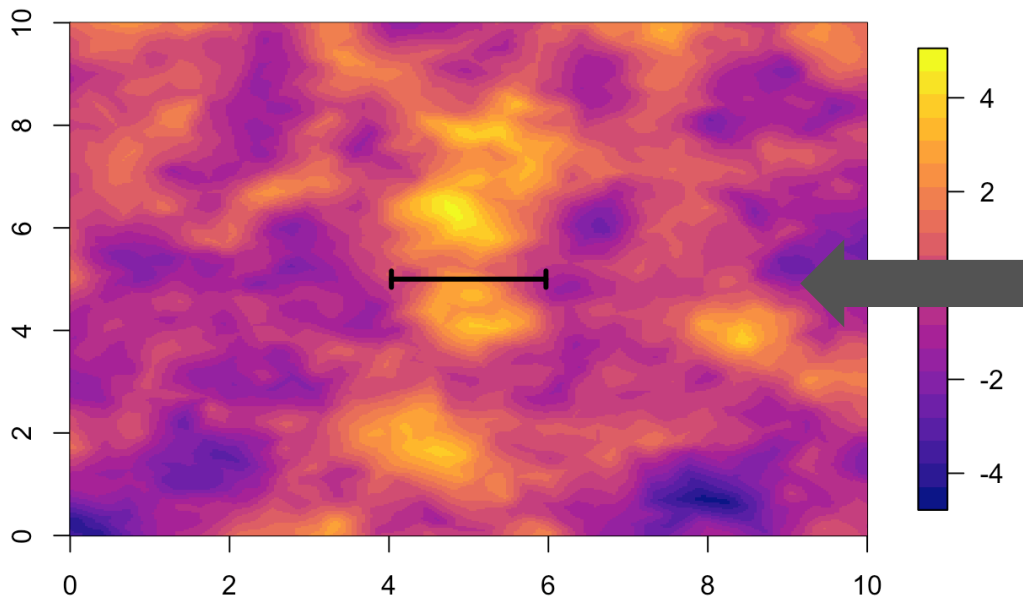
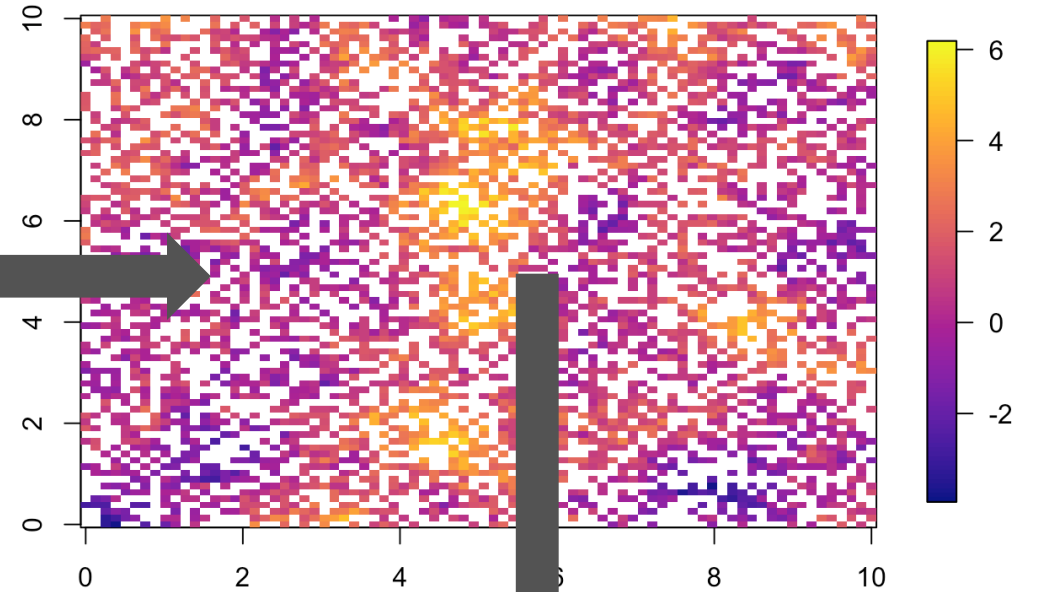
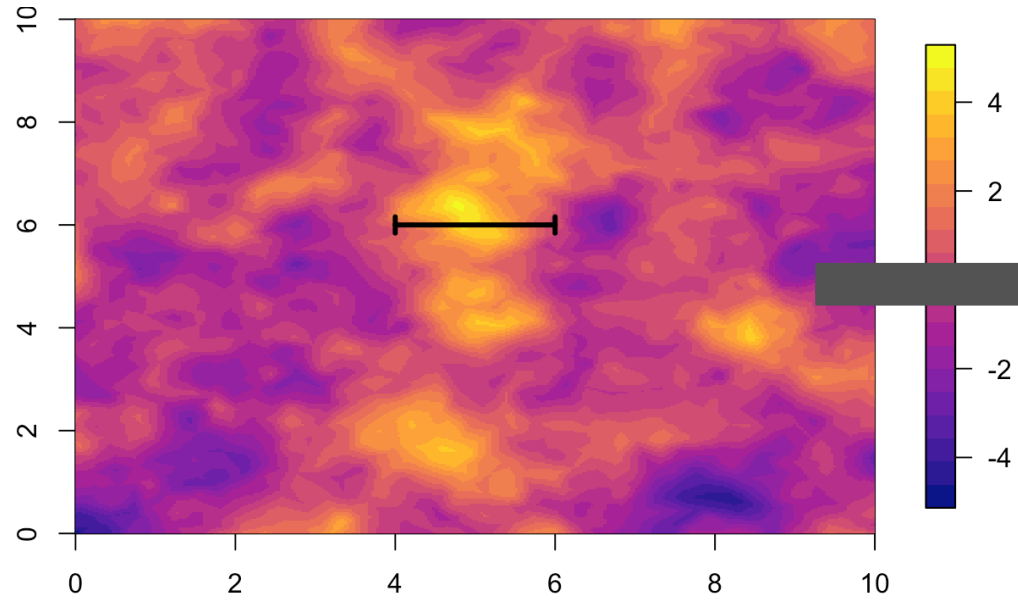


Group — Bulls — Cows & Bulls

Group — Bulls — Cows & Bulls

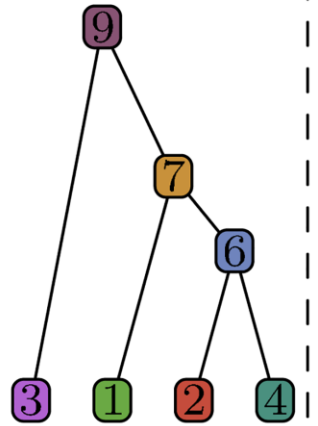
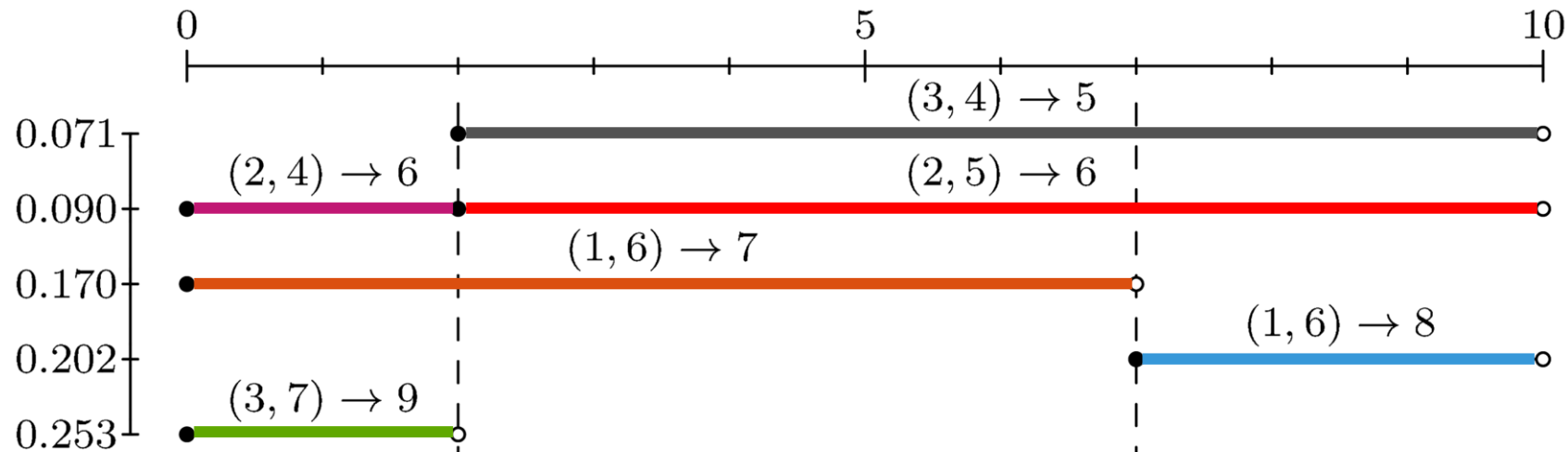


SPDE approach

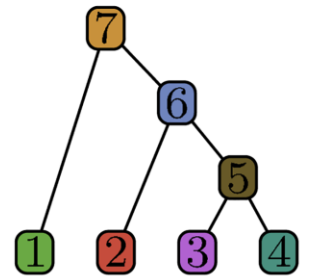


Genomics research trends – succinct trees

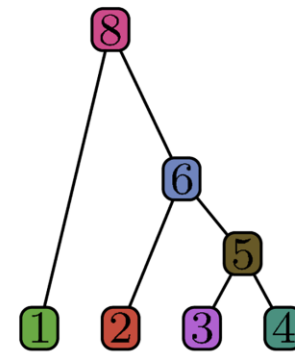
Kelleher et al. (2016) PCB: 10.1371/journal.pcbi.1004842



$\pi = 769607900$



$\pi = 765567000$



$\pi = 865568000$

	S	E	P	C	T
1	0	2	9	3,7	.25
2	7	10	8	1,6	.20
3	0	7	7	1,6	.17
4	0	2	6	2,4	.09
5	2	7	6	2,5	.09
6	2	10	5	3,4	.07

Multi-breed & sequence modelling

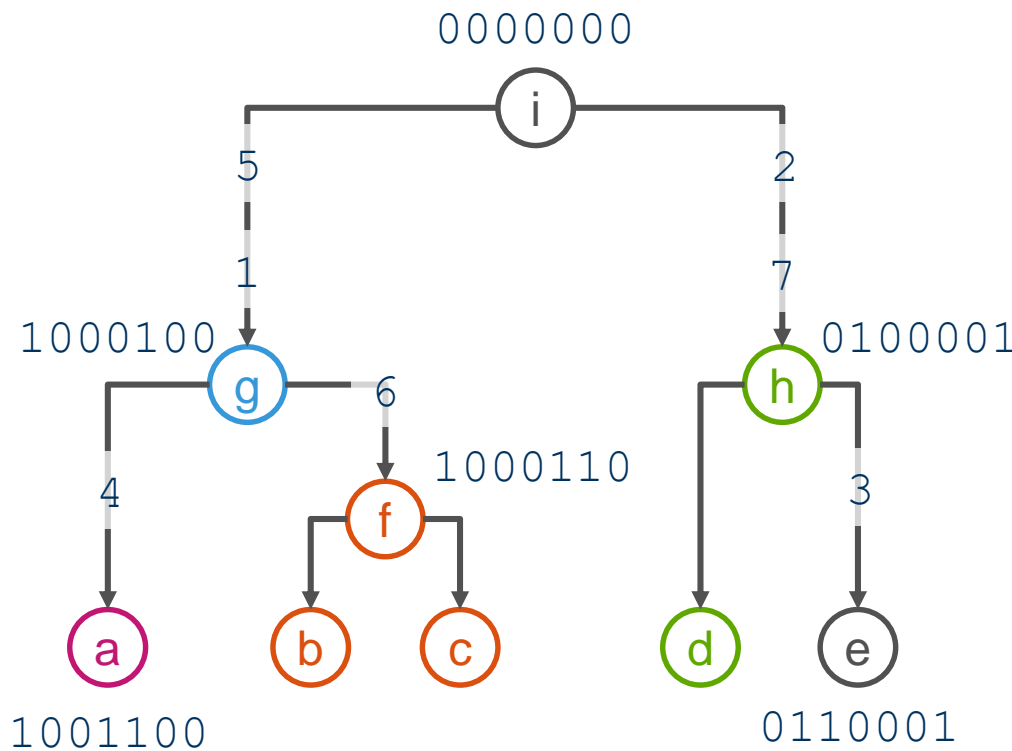
- Multi-breed predictions don't work as well as we want
- Marker-based models assume the same allele substitution effect irrespective of “sequence context”
→ data pulls estimates to the “LD context” of large breed(s)
- Can we model allele substitution effects in the “sequence context”?

Tree-Auto-Regressive model



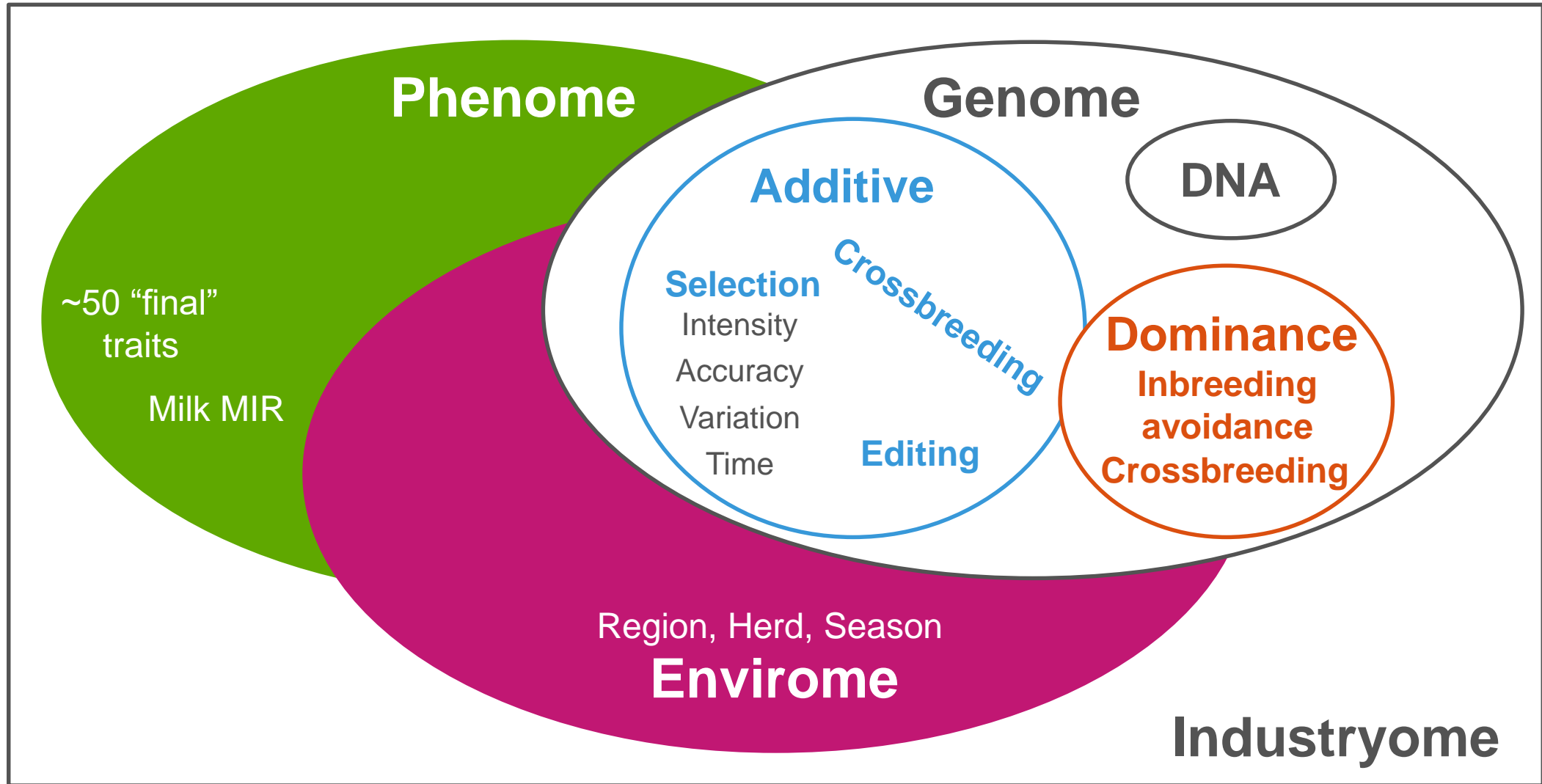
$$y = Xb + ZWh + e$$

W quite ~dense



- $h_i \sim N(0, v)$
- $h_{g'}|h_i \sim N(\alpha h_i, v_c)$
- $h_g|h_{g'} \sim N(\alpha h_{g'}, v_c)$
- $h_a|h_g \sim N(\alpha h_g, v_c)$
- $h_f, h_b, h_c|h_g \sim N(\alpha h_g, v_c)$
- $h_{h'}|h_i \sim N(\alpha h_i, v_c)$
- $h_h, h_d|h_{h'} \sim N(\alpha h_{h'}, v_c)$
- $h_e|h_h \sim N(\alpha h_h, v_c)$

Breeder's framework



Phenomics research trends

- New traits/data

- Pedometers, images, IR cameras, MIR, rumen sensors, microbiome, methylation, gene-expression, ...
- High-dimensional data (feature-wise and time-wise)
- Sub-phenotype definitions!? → Breeding objectives

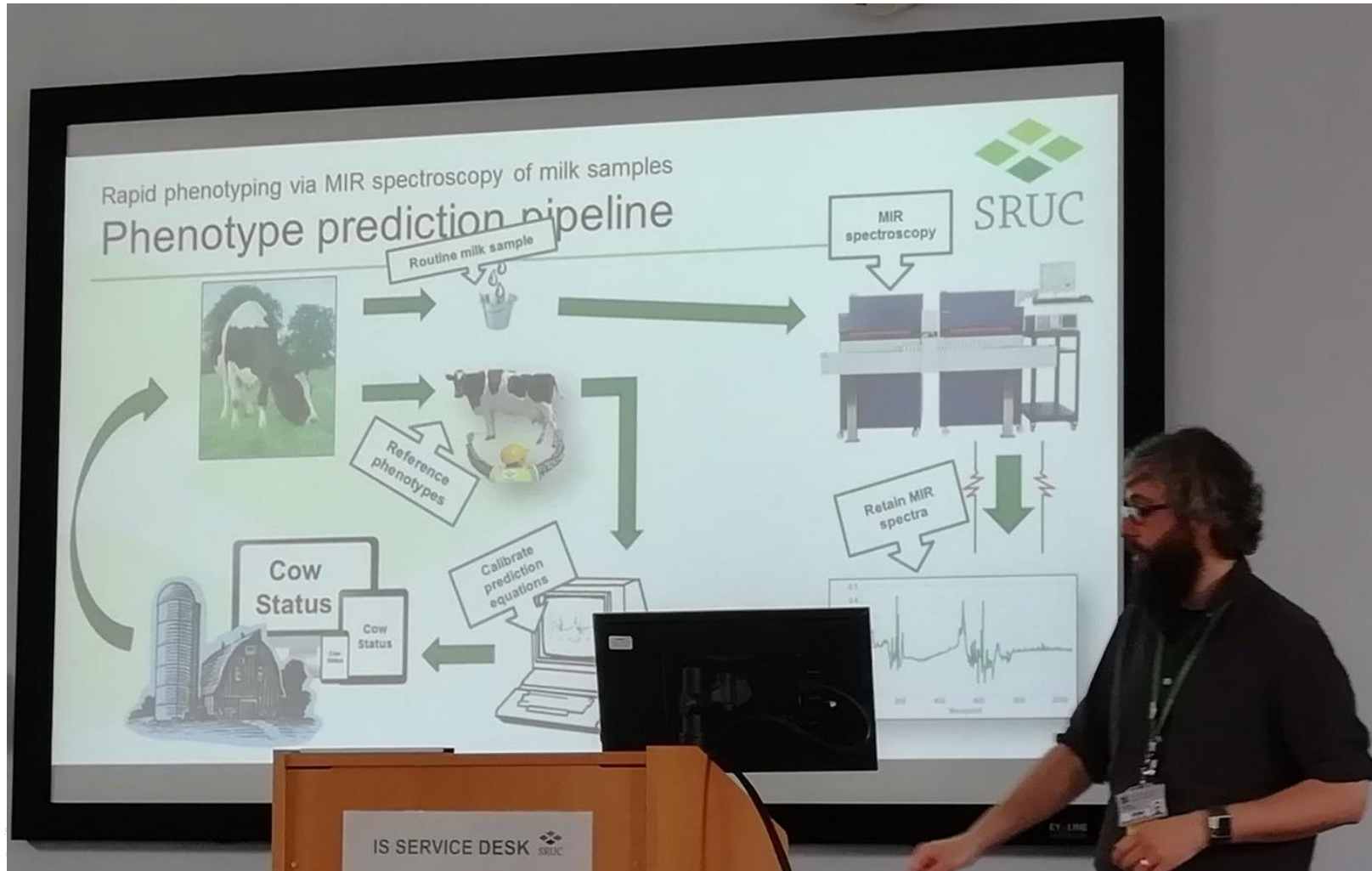
Precision Management

Phenotype generation
(machine learning)

Phenomics research trends – application in UK

- MIR → cow status (pregnancy, bTB, ...)

Coffey et al.



@dadyo32