

**ROS LIN**

Promotion of alleles by genome editing in livestock breeding programs

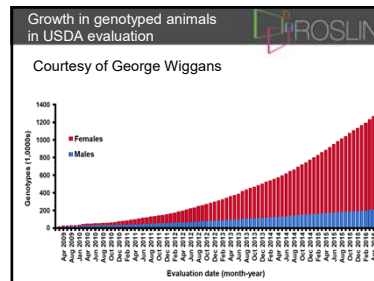
John M. Hickey, John Wooliams, Heidi Bergsø, Søren Sørensen, Matthew Cleveland, Alan Micham, William Herring, Bruce Whitelaw, Gregor Forgnone, Janusz Jankovic

www.alphagenes.roslin.ed.ac.uk  
@hickeyjohn

THE UNIVERSITY OF EDINBURGH BBSRC

Overview

- Generate and utilize large data sets with whole genome sequence for genetic improvement
- **Generate – Sequencing Strategy**
- **Utilize – Genome Editing**



Journal of Animal Breeding and Genetics

EDITORIAL

**Sequencing millions of animals for genomic selection 2.0**

Genomic selection (GS) has made animal breeding an extremely exciting field to have been part of in recent years. Breeding associations have been addressing... Generating sequence data for millions of individuals will require that the costs per individual be low, in order to maximize genetic gains across related channels, the...

Overall hypothesis of GS2.0

- "GS is now a mature technology" L. Misztal, JABG, 2016
- Sequence data has huge potential in breeding
- Huge volumes of sequence needed to realize potential (**because variants are correlated**)
- Breeding programs with 1 million animals with sequence information is normal (shortly!)
- Industrial scale fine mapping
  - X% of the variance mapped to causal variants
  - Which will lead to breeding opportunities

Breeding benefit

- **More persistent accuracy**
- **Commercial crossbred phenotypes**
- **Use of de-novo mutations**
- Manipulation of recombination / management of diversity
- **Part of a cascade of technologies to identify genome editing targets**

LCSeq for whole population whole genome sequencing

- Sequencing few individuals not that useful
- Sequence everybody at low-x & impute
- Make the population the target not the individual

For animals with bigger footprints

Algorithm 1 -> WHO to sequence

Algorithm 2 -> At which COVERAGE

Note: Animals already have genotype information

Algorithm 1 – WHO to sequence

- Divide chromosomes into  $n$  SNP long cores (e.g.,  $n=100$ )
- Build haplotype libraries for these cores across population
- Calculate haplotype population frequencies
- Find individual whose genome is most representative of population "Focal Individual"
- Mask focal individual's haplotypes in rest of population

### Conditional genomic footprints

$59,000 \times 2 \times (50,000/100) = 59,000,000$  slots to be filled

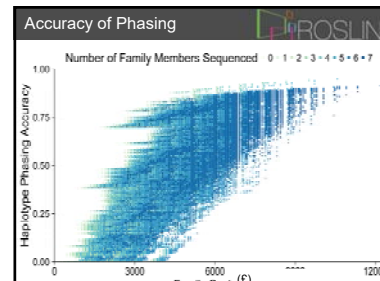
Animal ID	Conditional footprint count	Conditional footprint percentage
Molly	2,792,226	4.75
Polly	2,734,123	4.65
Dolly	1,669,476	2.84
Bob	1,601,064	2.72
Oscar	1,167,786	1.98
Bella	1,099,450	1.87
Bubbles	936,307	1.59
Flora	877,394	1.49
Shaun	865,116	1.47
Timothy	829,297	1.41

Sum of the top:  
 50 individuals = 40% of the unique haplotypes  
 100 individuals = 60% of the unique haplotypes

### Family based phasing of sequence data

Richardson (Mark Estillgen)

- Different combinations have different costs
  - Ancestors: 0 - 1 - 2 - 5 - 10x
  - Focal individual: 0 - 1 - 2 - 5 - 10 - 15 - 20 - 30 - 50 - 100x
  - £125 to >£20,000 per family
    - Library = £40; 1x whole genome = £85
- Given sequencing coverage per family member...
  - ... calculate PHASING ACCURACY

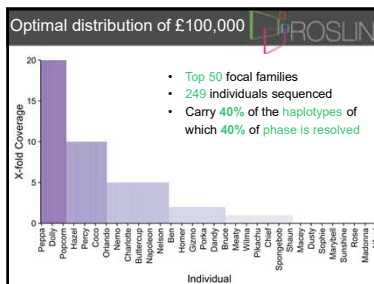


### Optimal distribution of £100,000

Account for:

- Conditional footprint
- Shared ancestry
- Budget
- Phasing accuracy

ID	Conditional footprint count	Conditional footprint %	Family sequencing cost (£)	Phasing accuracy
Polly	2,792,226	4.75	5,380	0.83
Billy	2,734,123	4.65	4,530	0.76
Molly	1,669,476	2.84	3,940	0.64
Oscar	1,167,786	1.98	2,200	0.58



### Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs

Janet Jenko, Gregor Gogjanc, Matthew A Cleveland, Rajeev K Varshney, C Bruce A Whitlow, John A Woolliams, and John M Hickey

Journal of Animal Breeding and Genetics (2015) 4:155  
 DOI 10.1186/s12271-015-0135-3

Open Access

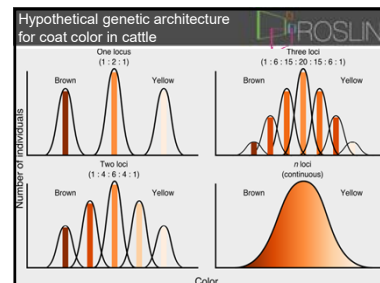
### Genome editing

GE is the process of precise editing genome

Nucleotides can be

- added
- deleted
- replaced

### Examples!



### PAGE

- Promotion of alleles by genome editing
- Detect favorable alleles and promote via editing
- Challenges
  - Quantitative traits = 10's of 1000's of favorable alleles
  - Millions of production animals
- Opportunities
  - Nucleus has only 25 to 500 sires per year
  - Huge genomic selection data sets to map causal variants
  - FAANG, Genome Editing, etc. to help prove causality
  - Dissemination structures in place

### Objective of study

- Develop a strategy to enable genome editing for quantitative traits in livestock breeding programs
  - PAGE
  - Promote alleles that already exist in the population
- Quantify the genome editing resources required
  - How many alleles per generation?
  - How many animals per generation?
- Quantify the benefit and risk
  - Extra genetic gain
  - Any impact on the genetic variance/long term response

### Overview of simulation

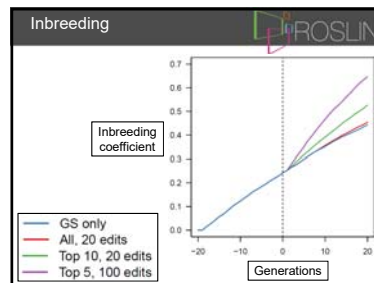
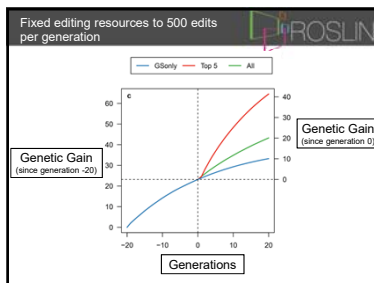
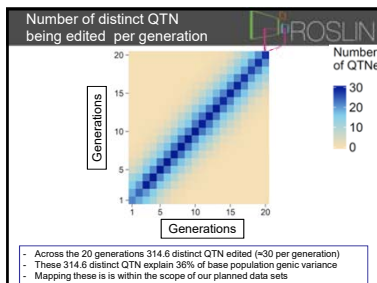
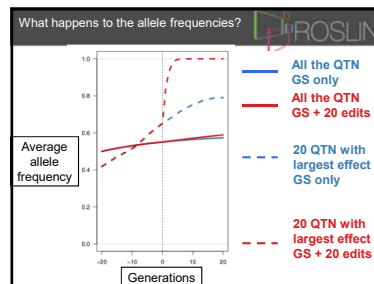
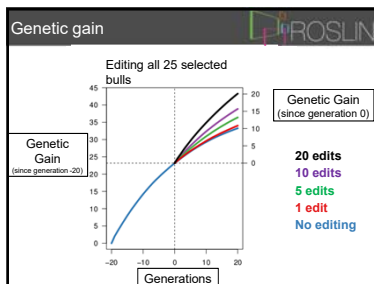
	Population	Generations	Mutation	Selection	PAGE
PAST	Ancestral	100,000 years	$2.5 \times 10^7$	Random selection	-
	Recent historical breeding	-20 to 0	-	Genomic selection 1000 candidates $10^4 \times 25 \mu$ selected parents	-
FUTURE	Future breeding	1 to 20 to 0	-	Genomic selection 1000 candidates $10^4 \times 25 \mu$ selected parents	Top 5 or 10, bottom 5 or 10 selected areas... G, 1, 5, 10, 20, 25, 50 or 100 edits per site

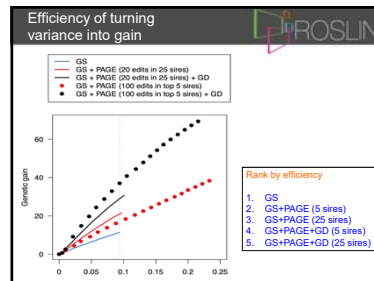
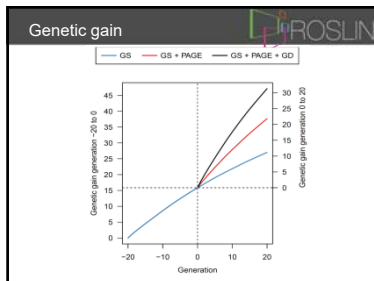
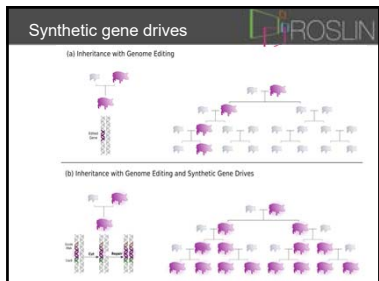
Trait controlled by 10,000 QTN!!!!

The simulation reflects a recurrent selection scheme in a breeding nucleus using genomic selection 2.0

### Comparison metrics

- Genetic gain
- Change in the allele frequency
- Number of distinct QTN edited
- Inbreeding





- ### Genome editing summary
- PAGE is very effective for increasing genetic gain**
    - 20 edits per sire
    - 25 sires per generation
  - Some risks if not managed properly**
    - Inbreeding
    - Targets
    - Epistasis (empirical results suggest this will be ok)
  - Practical use**
    - Huge data sets needed
    - Good targets
    - Costs and multiplexing need to be sorted!

- ### Genome editing summary
- Page works because of its precision**
    - Weakness of GS with perfect accuracy is that alleles do not segregate independently
    - With PAGE alleles behave as though they segregate independently
  - Can we find enough targets?**
    - 314.6 QTN edited that explain 36% of genic variance
    - Probably possible to find these with planned data sets
  - A big opportunity to protect genetic variance and efficiently turn it into gain**
  - However animal breeding "classic"**
    - will remain the cornerstone!!

### A new breeders equation?

Animal Breeding and Genetics

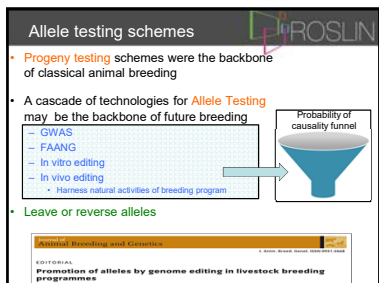
EDITORIAL: **Promotion of alleles by genome editing in livestock breeding programmes**

Genomic selection decouples selection from phenotyping  
Genome editing decouples gain from selection

$$\text{Response} = \frac{\text{Accuracy} \times \text{Selection intensity} \times \text{Diversity}}{\text{Time}}$$

PAGE = Promotion of alleles by genome editing

Genetic gain = Response + PAGE



- ### Final remarks
- Genome editing could work for quantitative traits
  - Likely next steps
    - Short term = focus on disease traits
    - Medium term = fix up recessive deleterious mutations
    - Long term = PAGE for quantitative traits

- ### Acknowledgements
- Roslin group members**
    - Gregor Gorjanc
    - Jarac Jentlo
    - Mara Battagin
    - Stefan Edwards
    - Serap Genen
    - Chris Gaynor
    - Roberto Artolin
    - Roger Ros Freixedes
    - Adriana Somavilla
    - Daniel Money
    - Masoud Shirazi
    - Paolo Gottardo
    - Diarmid de Burca
  - Former group members**
    - Anne-Michelle Faure
    - Jean Francois Dumasy
  - Further information**
    - www.alphagenes.roslin.ed.ac.uk
    - @hickeyjohn
    - john.hickey@roslin.ed.ac.uk
  - Vacancies**
    - Two post-doc positions currently available

