Promotion of alleles by genome editing in livestock breeding programs

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

Growth in genotyped animals in USDA evaluation

Courtesy of George Wiggans

Overall hypothesis of GS2.0
• "GS is now a mature technology" - 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

True haplotypes
• 10x sequencing (5 individuals)
• 2x sequencing (10 individuals)

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

<table>
<thead>
<tr>
<th>Animal ID</th>
<th>Conditional footprint count</th>
<th>Conditional footprint percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>1,601,064</td>
<td>2.72</td>
</tr>
<tr>
<td>Bella</td>
<td>1,099,450</td>
<td>1.87</td>
</tr>
<tr>
<td>Flora</td>
<td>877,394</td>
<td>1.49</td>
</tr>
<tr>
<td>Oscar</td>
<td>1,167,786</td>
<td>1.98</td>
</tr>
<tr>
<td>Shaun</td>
<td>865,116</td>
<td>1.47</td>
</tr>
<tr>
<td>Timothy</td>
<td>829,297</td>
<td>1.41</td>
</tr>
<tr>
<td>Dolly</td>
<td>1,669,476</td>
<td>2.84</td>
</tr>
<tr>
<td>Polly</td>
<td>2,792,226</td>
<td>4.75</td>
</tr>
<tr>
<td>Molly</td>
<td>2,734,123</td>
<td>4.65</td>
</tr>
</tbody>
</table>

50 individuals = 40% of the unique haplotypes
100 individuals = 80% of the unique haplotypes

10,000 x 2 x (50,000/100) = 50,000,000 slots to be filled

Family based phasing of sequence data

<table>
<thead>
<tr>
<th>Family ID</th>
<th>Conditional footprint count</th>
<th>Conditional footprint percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polly</td>
<td>2,792,226</td>
<td>4.75</td>
</tr>
<tr>
<td>Billy</td>
<td>2,734,123</td>
<td>4.65</td>
</tr>
<tr>
<td>Molly</td>
<td>1,669,476</td>
<td>2.84</td>
</tr>
<tr>
<td>Oscar</td>
<td>1,167,786</td>
<td>1.98</td>
</tr>
<tr>
<td>Shaun</td>
<td>865,116</td>
<td>1.47</td>
</tr>
<tr>
<td>Timothy</td>
<td>829,297</td>
<td>1.41</td>
</tr>
</tbody>
</table>

Family based phasing of sequence data

AlphaFamSeq (Mara Battagin)

- 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 = 1/4 to 1/150 genomes x10

- Given sequencing coverage per family member, calculate phasing accuracy

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

Genome editing

- GE is the process of genome editing genome
- Nucleotides can be
  - added
  - deleted
  - replaced

Examples!

Hypothetical genetic architecture for coat color in cattle
Promotion of alleles by genome editing

- Detect favorable alleles and promote via editing
- Challenges
  - Quantitative traits = 10s of 1000s 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

- The simulation reflects a recurrent selection scheme in a breeding nucleus using genomic selection 2.0
- Trait controlled by 10,000 QTN!!

Comparison metrics

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

Genetic gain

- Editing all 25 selected bulls
- Genetic Gain (since generation 0)
- 20 edits
- No editing
- 10 edits
- 5 edits
- 1 edit

Genetic Gain (since generation -20)

- Generations
- Genetic gain
- GS only
- GS + 20 edits
- 20 QTN with largest effect

Number of distinct QTN edited per generation

- Across the 20 generations, 5144 distinct QTN edited (259 per generation)
- Fixed editing resources to 500 edits per generation

Inbreeding

- Inbreeding coefficient
- GS only
- All 20 edits
- Top 10, 20 edits
- Top 5, 100 edits

What happens to the allele frequencies?

- All the QTN
- GS only
- GS + 20 edits

Number of distinct QTN being edited per generation

- Generations
- Number of QTN
- GS only

GS only
- All 20 edits
- Top 10, 20 edits
- Top 5, 100 edits
Synthetic gene drives

Genetic gain

Efficiency of turning variance into gain

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 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?
- Accuracy x Selection intensity x Diversity
- Time
  - Response = Genetic gain = Response + PAGE

Allele testing schemes
- Progeny testing schemes were the backbone of classical animal breeding
- A cascade of technologies for Allele Testing may be the backbones of future breeding
- GWAS
- FAANG
- In vitro editing
- In vivo editing
- Leave or reverse alleles

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
- Further information
  - www.alphagenes.co.uk
  - @hickeyjohn
  - john.hickey@roslin.ed.ac.uk
- Vacancies
  - Two post-doc positions currently available
Funding to the research program

Disclaimer: Subsets of funders support subsets of the work presented today.