Developing DNA Tests for Improved Fertility and Reduced Embryonic Loss in US Cattle Breeds

Jerry Taylor, Troy Rowan, Tamar Crum, Jesse Hoff, Bob Schnabel, Jared Decker and Dave Patterson

BIF Research Symposium and Convention, June 18-21, 2019
Brookings, SD

GGPF250

Contents

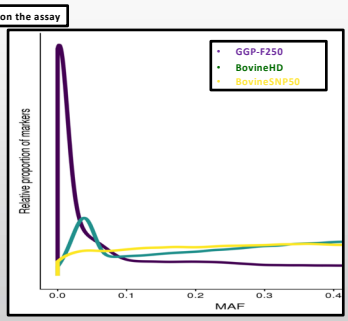
- GGP-F250
- Genotype Imputation
  - What is it?
  - How do you do it?
  - Why does it matter?
- Haplotypic Diversity in US Beef Breeds
  - What is it?
  - How do you measure it?
  - Why does it matter?
- Detecting Early Embryonic Lethals
  - Haplotypes that are never found in animals with 2 copies
  - Within and across breeds

GGP-F250

- GGP-F250 is a new genotyping assay developed in a collaboration between GeneSeek and MU
- Differs DRAMATICALLY from existing genotyping assays
  - GeneSeek BOVG50v1, GGP-90KT, GGP-HDV3, GGP-LDV1, GGP-LDV3 and GGP-LDV4
  - Illumina BovineHD and BovineSNP50
  - Zoetis i50K
  - Irish Cattle Breeding Federation IDBv3
- GGP-F250 is focused on genes and rare potentially functional variants
  - 227,233 targeted loci of which 175,135 variable in 18,786 genotyped animals
  - 32,428 common to BovineHD and BovineSNP50
  - >68% are in genes

GGP-F250

Proportion of SNPs on the assay



← Rare Variation Common Variation →

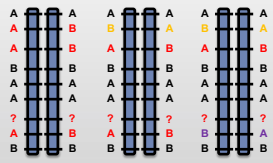
Rare Variants

- Most variation in the genome of cattle is rare
- Most variants that create variation in traits of cattle are rare
- The variants used on cattle genotyping chips are mostly common
  - Probably do not capture well the effects of rare alleles
  - Makes animals appear to be more similar to each other than they actually are
  - EPDs less accurate than they could be
- Lethal variants are usually rare!
  - The lethal allele is usually at low frequency at each locus but can be driven to relatively high frequencies by AI
  - The number of lethal loci in a breed is not known
  - About 10% of all genes (2,400) are essential for life

Genotype Imputation

When we genotype an animal the data look like

SNP	Chromosome	Position	Genotype
1	1	16947	AA
2	1	36337	AB
3	1	78655	AB
4	1	83412	BB
5	1	89725	AA
6	1	100260	AA
7	1	135098	-
8	1	146011	AB
9	1	147231	BB



Imputation is the process of:
1. Sorting alleles onto each chromosome (phasing)
2. Estimating missing genotypes

Genotype Imputation

Pair of Chromosomes

Variable positions on 50K

Variable positions in Sequence

If I know EXT's 50K genotypes

Can I estimate his genome sequence?

Genotype Imputation

- Routinely used to "fill in the blanks" when we genotype animals with a 50K assay for EPD analysis
- Used to impute 3K → 50K and 11K → 50K for Holstein PTAs
- Used to impute 50K → 800K → Whole Genome Sequence (10M SNPs)
- Genotypes for SNPs with MAF > 5% accurately imputed
- Doesn't have much much effect on accuracy of EPDs
  - So we don't impute to high density for national cattle evaluation
- But what about the rare variants that are difficult to impute?

Imputation Pipeline

bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

doi: https://doi.org/10.1101/517144

New Results

A Multi-Breed Reference Panel and Additional Rare Variation Maximizes Imputation Accuracy in Cattle

Troy N Rowan, Jesse L Hoff, Tamar E Crum, Jeremy F Taylor, Robert D Schnabel, Jared E Decker

doi: https://doi.org/10.1101/517144

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract Info/History Metrics Preview PDF

Quality Control (PLINK)

Merged filtered file

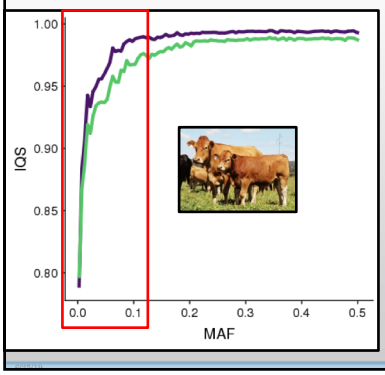
Phasing (Eagle)

Imputation (Minimac3)

850K imputed genotypes

50K → 850K (HD + F250) (835,947 autosomal)

Imputation Accuracy By SNP

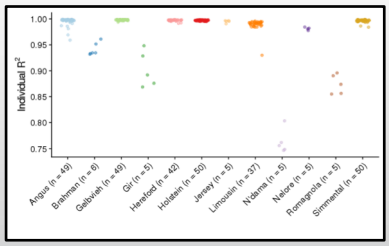


Gelbvieh Reference
265 F250
514 HD
IQS = 0.982

Composite Reference
28,183 F250
9,629 HD
IQS = 0.990

Imputation Accuracy By Breed

Imputation Accuracy



Reference Panel

Breed	Complete	HD	F250
Holstein	1932	3170	1944
Gelbvieh	257	265	514
Angus	132	2067	14454
Simmental	67	427	1759
Brahman	7	25	632
Romagnola	4	11	4
Nelore	3	855	4
Jersey	3	21	5
Gir	3	10	6
Ndama	3	7	4
Hereford	0	569	1834
Limousin	0	215	142

Poorly imputed breeds have low representation in Reference Panel

Feed Efficiency

Assay	No. of Variants	No. of Samples
BOVGv1	47,843	1
GGP-90KT	76,999	2
GGP-F250	227,234	4,463
GGP-LDV3	26,504	40
BovineHD	777,962	1,638
ICBF IDBV3	53,450	315
BovineSNP50	58,336	7,018
Total		13,477

Feed Efficiency

Breed	Average Ancestry %	SD %
Angus	31.87	± 24.67
Braunvieh	1.6	± 4.13
Brown Swiss	0.3	± 1.47
Charolais	6.91	± 13.66
Gelbvieh	6.69	± 13.64
Guernsey	0.56	± 2.01
Hereford	17.39	± 29.19
Holstein	2.03	± 3.79
Indicine	0.17	± 1.56
Japanese Black	0.22	± 4.34
Jersey	0.26	± 1.40
Limousin	3.08	± 11.40
N'Dama	0	± 0.17
Red Angus	16.02	± 20.06
Romagnola	0.19	± 1.20
Shorthorn	3.7	± 4.90
Simmental	9.01	± 14.51

MultiBreed Trait Analysis

Trait	h^2	Va	Ve
RFI	0.45	2.3029	2.7901
MMWT	0.56	112.73	88.334
DMI	0.5	5.1659	5.0817
ADG	0.42	0.1796	0.2462

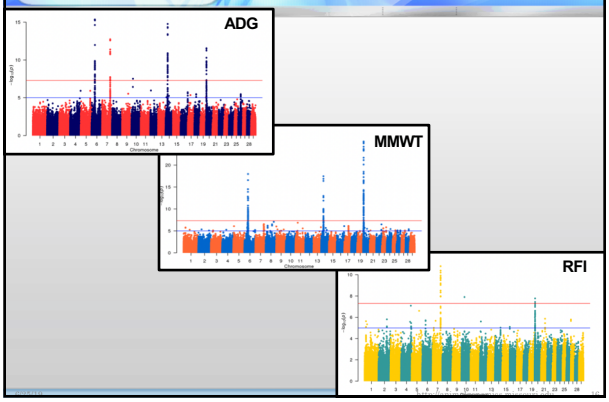

QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies

Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle

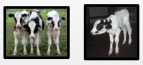
Trait	h^2 850K	h^2 HD	h^2 SNP50
RFI	0.45	0.34	0.4
MMWT	0.56	0.49	0.5
DMI	0.5	0.32	0.3
ADG	0.42	0.26	0.29
No. Animals	11,505	3,973	5,047
Av. No. Animals/Analysis		1,310	1,262
Populations x Replicates		3 x 2	4 x 1

Heritability is higher when you include rare variation!

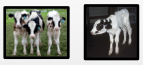
Large-Effect Genes



Bovine Respiratory Disease

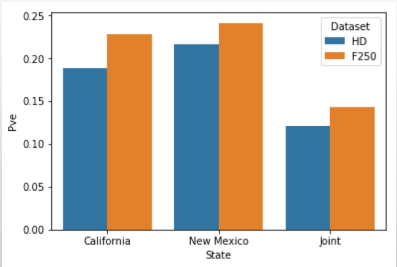


California:
Controls = 1,011
Cases = 1,003



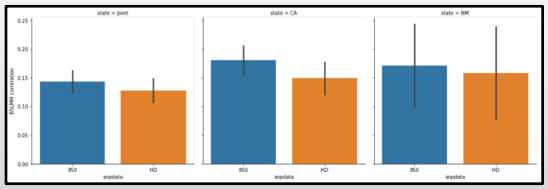
New Mexico:
Controls = 372
Cases = 376

Bovine Respiratory Disease



Heritability IS higher when you include rare variation!

Bovine Respiratory Disease



Prediction Accuracy increases when you include rare variation!

How Important Is This?

Genomic Diversity


1 Mb

Angus have 5 variable sites and 5 haplotypes (distinct chromosomes)

Simmental have 7 variable sites and 8 haplotypes

You are selecting for the best chromosome!!!

Haplotype Frequencies?


1 Mb

0.90 0.02 0.03 0.01 0.04

0.20 0.08 0.09 0.10 0.15 0.22 0.12 0.04

Genotypic Diversity

$$G = 1/\sum_{i=1}^{\#Hap} p_i^2$$

G ranges from 1 to # Haplotypes
"Effective" Number of Haplotypes in Population

Haplotype Frequencies?


1 Mb

0.90 0.02 0.03 0.01 0.04

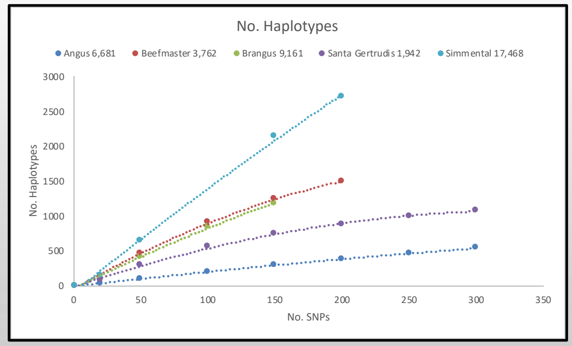
0.20 0.08 0.09 0.10 0.15 0.22 0.12 0.04

G = 1.23 6.61

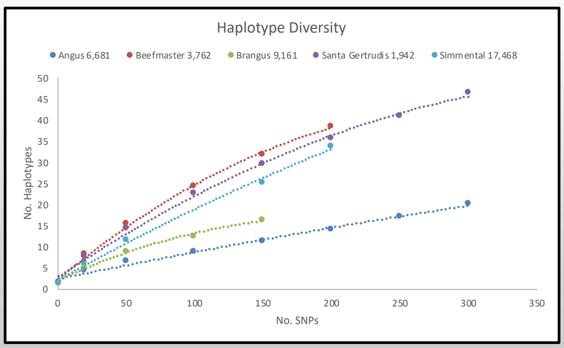
Breed Association Data

Breed	Genotypes	No. Animals	Haplotyped SNPs							
			1	20	50	100	150	200	250	300
Angus	50K-->850K	6,681	✓	✓	✓	✓	✓	✓	✓	✓
Beefmaster	50K-->850K	3,762	✓	✓	✓	✓	✓	✓		
Brangus	50K-->850K	9,161	✓	✓	✓	✓	✓			
Santa Gertrud	50K-->850K	1,942	✓	✓	✓	✓	✓	✓	✓	✓
Simmental	50K-->850K	17,468	✓	✓	✓		✓	✓		

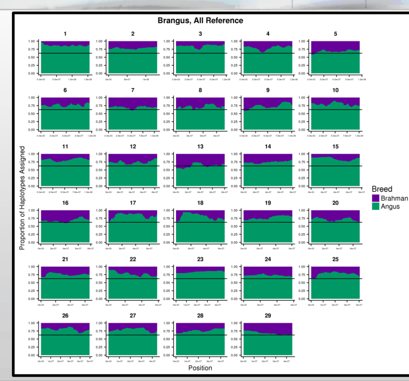
Number of Haplotypes



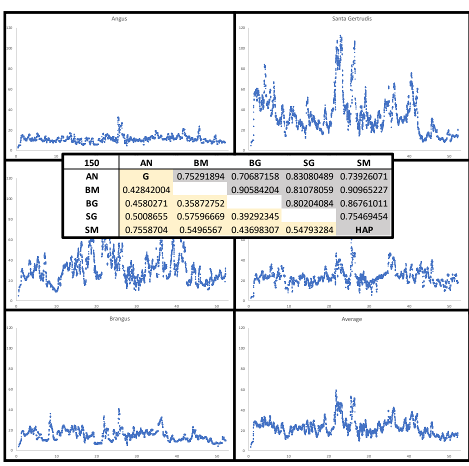
Haplotype Diversity



Brangus

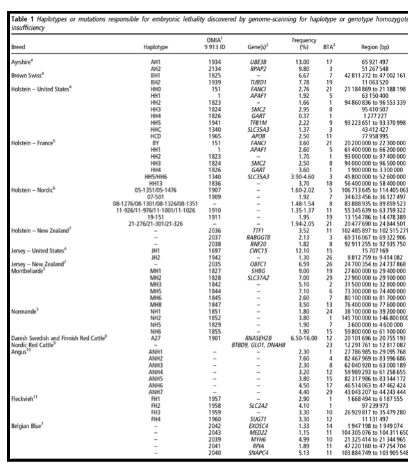


BTA23



150	AN	BM	BG	SG	SM
AN	G	0.75291894	0.70687158	0.83080489	0.73926071
BM	0.42842004		0.90584204	0.81078059	0.90965227
BG	0.4580271	0.35872752		0.80204084	0.86761011
SG	0.5008655	0.57596669	0.39292345		0.75469454
SM	0.7558704	0.5496567	0.43698307	0.54793284	HAP

Evidence For Lethals

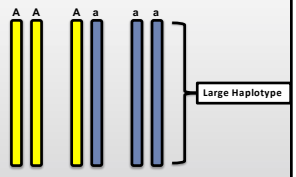


How Were These Detected?

Hardy-Weinberg Equilibrium principle indicates that:

Gametes	Female	
	Alleles	A	a
	Frequencies	p	q
Male			
A	p	p^2	pq
a	q	pq	q^2

Table : Punnett square for Hardy-Weinberg equilibrium



In a sample of N individuals we would expect to see	Np^2	2Npq	Nq^2	
So…if q=2% and N = 1,000,000 animals genotyped			400	Expected
But what if…			0	Observed?

How Large Should Haplotypes Be?

- If you had accurate WGS variants on a very large sample of animals (10s to 100s of thousands) you would simply analyze single markers
- Using 50K markers Hoff et al. 2017 estimated 20 markers
  - These haplotypes are about 1 Mb long
  - Long haplotypes will capture recent mutations
  - Smaller haplotypes will capture older mutations

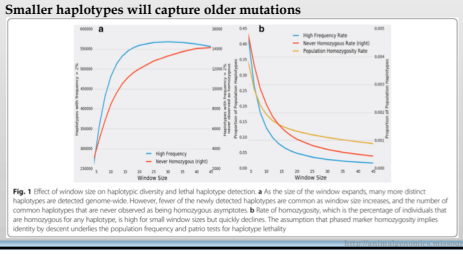


Fig. 1 Effect of window size on haplotypic diversity and lethal haplotype detection. **a** As the size of the window expands, many more distinct haplotypes are detected genome-wide. However, fewer of the newly detected haplotypes are common as window size increases, and the number of common haplotypes that are never observed as being homozygous asymptotes. **b** Rate of homozygosity, which is the percentage of individuals that are homozygous for any haplotype, is high for small window sizes but quickly declines. The assumption that phased marker homozygosity implies identity by descent underlies the population frequency and patrio tests for haplotype lethality

How Was Analysis Performed?

- Overlapping windows of N=1, 20, 50, 100, 150, 200, 250 or 300 SNPs
  - 8 separate analyses
  - Step along chromosome 1 SNP at a time
- Test every haplotype with no homozygotes
  - Function of haplotype frequency
  - Retain those with P < 0.10
  - Concatenate all overlapping regions – select largest frequency and smallest P-value to represent region

Angus 6,681

Haplotype Size (SNPs)	No. Haplotypes P<0.10	Genomic Regions
1	118	118
20	6,563	457
50	12,995	327
100	19,165	241
150	23,288	200
200	26,337	190
250	29,621	184
300	32,762	180

How Was Analysis Performed?

- Pool regions from analyses of N=1, 20, 50, 100, 150, 200, 250 or 300 SNPs
- Identify regions overlapping from different analyses
  - Retain regions identified in 2 or more analyses and for which at least one analysis had P < 0.05 (the rest are P <0.10 at worst)
  - Concatenate overlapping regions – Megaregion
  - Identify analysis with highest significance – most likely region
- Identify all genes within Megaregion
  - Identify those essential for life

Angus

- 122 regions identified by an average of 5.10 ± 2.11 analyses
  - Average size 1,490,585 bp (range 48,748 to 5,296,187 bp)
  - Total size 181,851,353 bp (7.31% of autosomal genome)
  - 1,350 genes (7.92% of autosomal genes)
  - 184 essential for life (13.63% of regional genes)
  - Embryonic loss = 8.8% (if all real)

Angus 6,681

No. Analyses	No. Regions	% Regions
2	22	17.07
3	16	14.63
4	12	9.76
5	9	7.32
6	16	13.01
7	35	29.27
8	12	8.94
	122	100.00

Angus

- 111 regions contain genes identified by 5.30 ± 2.07 analyses
- 76 regions have genes known to be essential for life in human (68%)
  - Identified by 5.87 ± 1.88 analyses
  - 36 of these regions have essential genes in most likely haplotype (47%)
  - Embryonic loss = 5.5%
- 2 regions contain BovineHD markers located in essential genes with largest association for all regional analyses
  - Probably not causal – but ARE diagnostic

Conclusions

- **GGP-F250 data from Heifer Fertility, Respiratory Disease, Feed Efficiency and Local Adaptation Projects**
  - Allows accurate imputation of 50K data to 850K in many breeds
  - Includes rare variation
  - Genomic Predictions based on 850K are more accurate than 50K!
  - Breed Association need to evaluate utility of rare variant imputation for national cattle evaluation
- **Strong evidence for 76 regions in Angus harboring lethals**
  - 5.5% of embryos are lost in U.S. registered Angus
  - Sequence carrier bulls
  - Identify candidate variants and include on industry assays
  - 2 variants are known – test for these now
- **Analyses underway for Simmental, Beefmaster, Brangus, Santa Gertrudis**

Acknowledgements

- **Breed Associations sharing data:**
  - American Angus Association
  - International Brangus Breeders Association
  - Beefmaster Breeders United
  - Santa Gertrudis Breeders International
  - American Simmental Association
- **10,000 heifers**
  - Missouri Show-Me-Select Replacement Heifer Program
  - Missouri Angus Association
  - Circle A Angus
- **USDA NIFA grants:**
  - 2011-68004-30214, 2011-68004-30367, 2013-68004-20364
  - 2015-67015-23183, 2016-68004-24827
- **GeneSeek for building the GGP-F250!**

The End