Genomic sequencing and selection strategy of livestock animals

Dr Steven Kappes, Associate Administrator, USDA, ARS, Office of National Programs

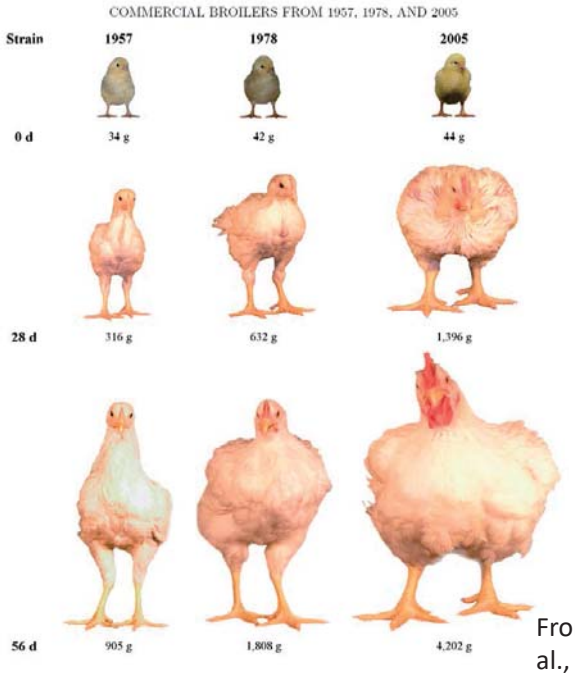
Animal agriculture is a critical component of the U.S. economy

- > 1 million farms
- > $182 billion in products (2011)
- > 89 million cattle, 66 million pigs and 500 million chickens
- > 2.3 M people employed
- > 63% of all farm income
- > $440 billion in economic output

-- National Agricultural Statistics Service, 2012

What has traditional genetic selection done for us?

Genetic Selection has been amazingly successful for highly heritable traits

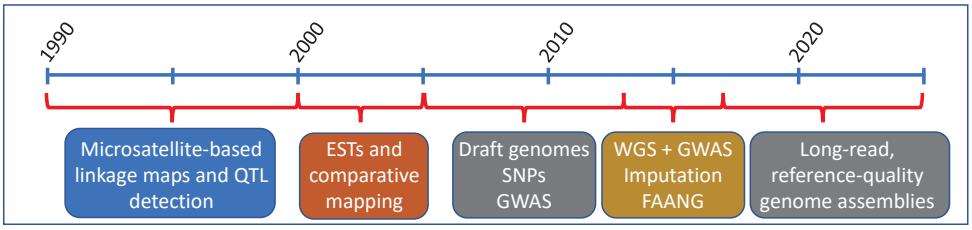


From Zuidhof et al., 2014

What could genomics do for us?

- Aid in selection :
  - Identify QTL
  - Develop markers to predict favorable allele(s)
  - Genotype individuals to determine selection

Technology driving genomics : a brief history of (cow) time

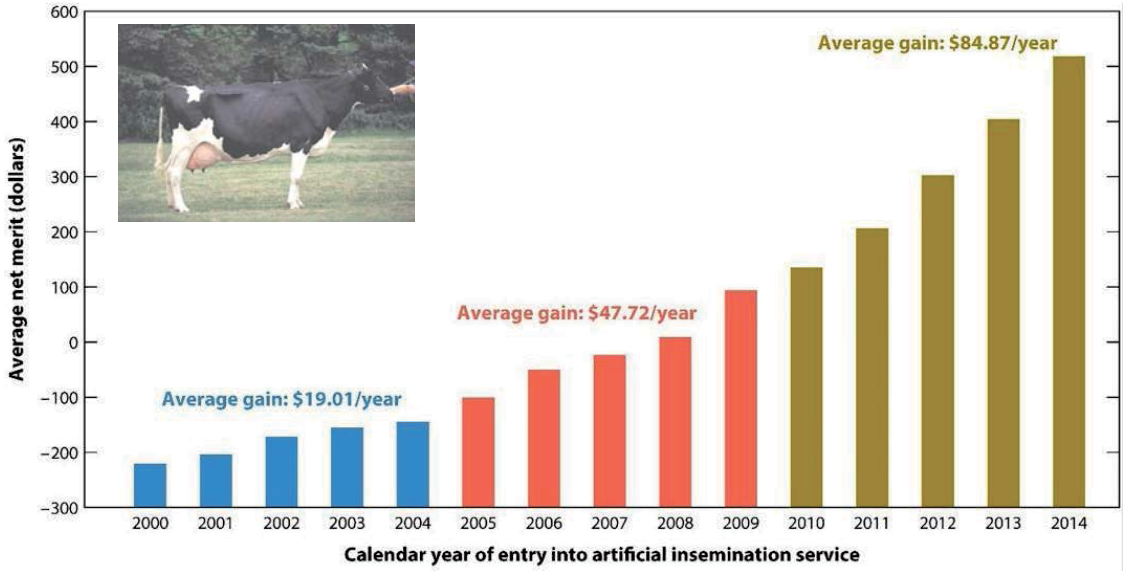


QTL :	Quantitative Trait Loci
EST :	Expressed Sequence Tag
SNP :	Single Nucleotide Polymorphism
WGS :	Whole Genome Shotgun (sequencing)
GWAS :	Genome-wide Association Study
FAANG :	Functional Annotation of Animal Genomes

Traditional Selection Programs

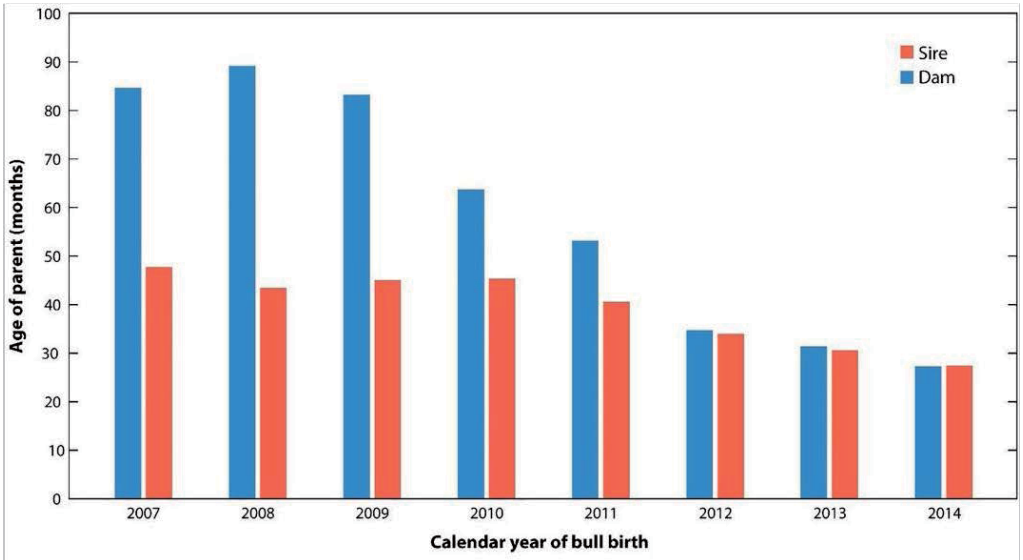
- Dairy data collected for >100 years
- Estimate genetic merit for animals and select
- Genetic improvement approached theoretical optimum (200+ lb milk per yr)

More importantly, genomics enabled genetic selection in dairy cattle effectively doubled genetic progress



Wiggans GR, et al. 2017.
Annu. Rev. Anim. Biosci. 5:309–27

Through the ability to effectively select young animals, genomic selection reduces generation interval and the expense of sire evaluations, leading to more rapid genetic progress



Wiggans GR, et al. 2017.
Annu. Rev. Anim. Biosci. 5:309–27

Why genomics works in dairy cattle

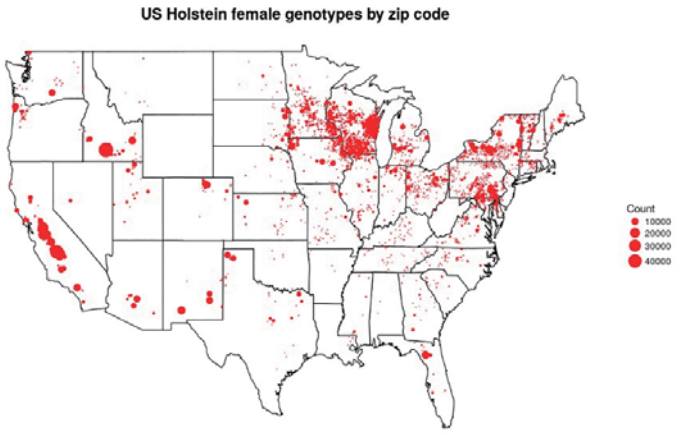
- Extensive historical data available
- Well-developed genetic evaluation program
- Widespread use of artificial insemination bulls
  - Historic DNA source
  - Large half-sib families
  - Progeny test programs – accurate genetic merit
- High value animals, justify genotyping expense
- Long generation interval associated with data collection

- Our best success, genomic selection in dairy cattle
  - ARS partnered with the Council for Dairy Cattle Breeding

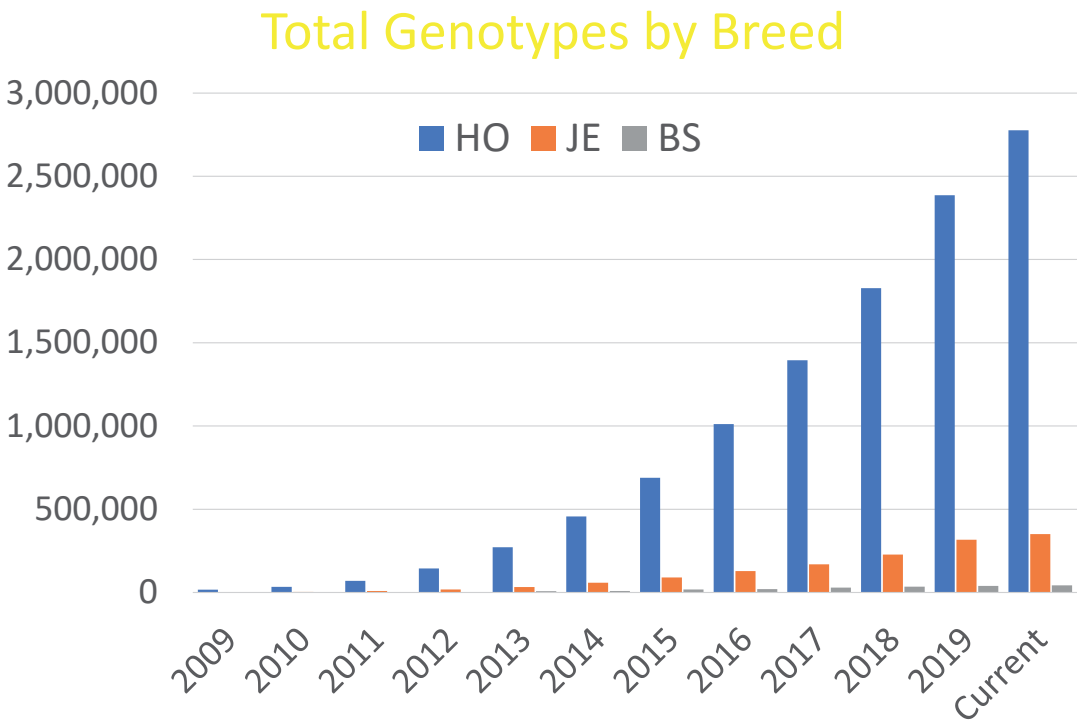
CDCB
COUNCIL ON DAIRY CATTLE BREEDING

CDCB Connection January 2018

CDCB and USDA Extend Data, R&D Agreement



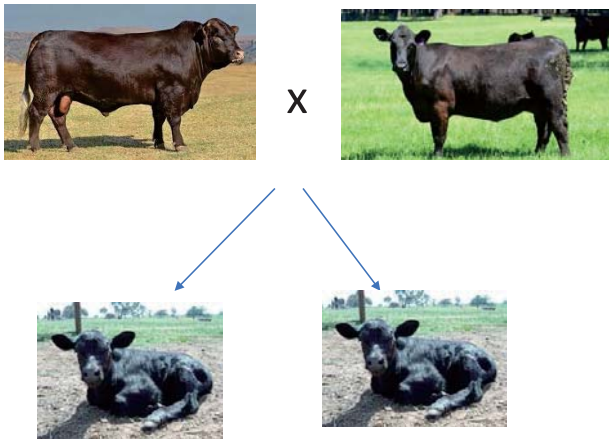
Troy Rowan, George Wiggans
1,136,252 animals with Known Zip Code



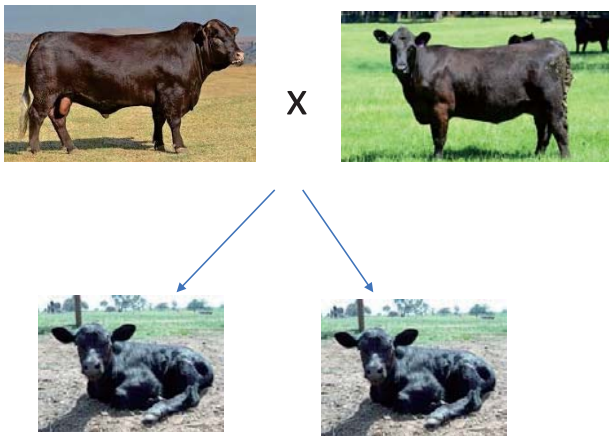
Similar methods are used in other livestock species but:

- Individual companies have separate populations, rather than a single national population in dairy cattle. Training populations are smaller and are specific to the population being predicted
- Commercial pigs and poultry (and 50% beef cattle) are typically crossbred animals and genomics applied to purebred lines. Some dilution of genetic progress in selected populations
- Pigs and poultry are not as valuable, so genotyping costs are harder to justify
- Generation intervals in pigs and poultry are shorter, so not as much advantage to shortening generation interval as in dairy cattle.

One reason genomic selection is effective is that it allows better tracking of inheritance from parent to offspring

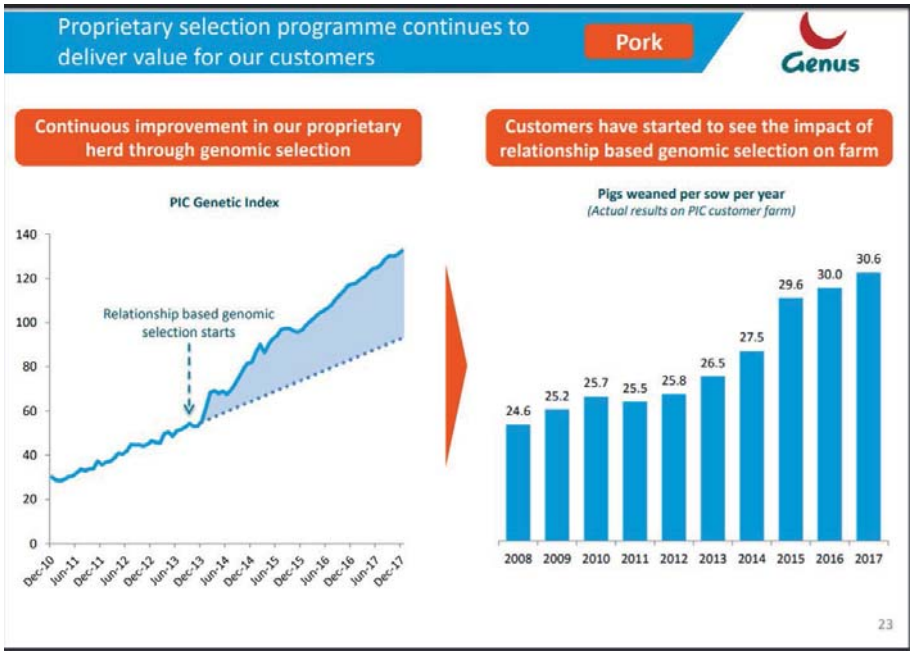


Full sib **average** genetic relationship .5
17% greater than or equal to .7
17% less than or equal to .3



Genomics provides the actual relationship, substantially improving the calculations

Genetic Improvement Swine



Porcine Genetic Improvement Index US$

Measures the genetic gain we achieve in our porcine nucleus herds.

Definition
The index measures the marginal economic value improvement in customers' US$ profitability, per commercial pig per year, on a rolling three-year average. Prior years' index ratings have been updated, to reflect the latest results from genomic selection and economic values of pork production.

Performance
A step change in genetic gain value improvement, up US$1.41 to US$3.64, as a direct result of implementing genomic selection technology over the last two years.

https://www.genusplc.com/media/1460/genus-interim-results-presentation-28feb2018.pdf

Similar improvements occur in poultry. Improved poultry genetics are distributed throughout the world

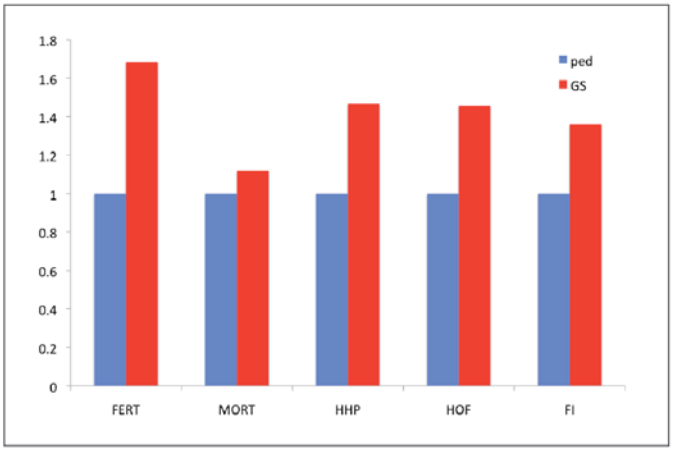


Figure 6. Relative improvement in prediction accuracy of genomic selection (GS) over pedigree-based (Ped) EBVs, measured as the correlation of EBVs with adjusted phenotype for 5 traits: fertility % (FERT), laying mortality (MORT), hen-housed egg production (HHP), hatchability % (HOF) and feed intake (FI).

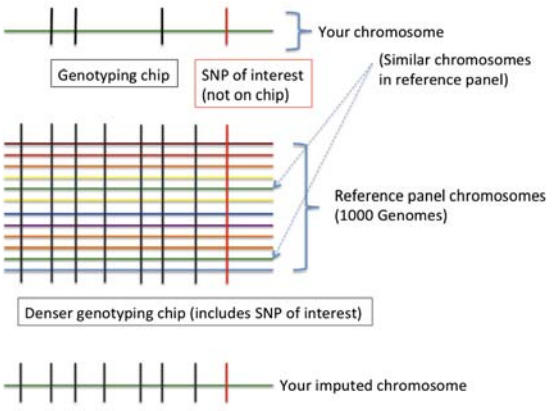
From Wolc et al., 2015

Worldwide impact of HyLine Genetics – sales of Hy-Line Birds

97 countries, and another 23 where distributors sell

Genomic approaches use to improve rate of progress for performance Traits

Imputation to reduce genotyping cost



Genotype founder animals at high density. Genotype offspring at lower density (less expensive) and impute the transfer of higher density genotypes. Must genotype some animals at high density to maintain relationships.

Strategy used in all species but particularly important in poultry

However, full implementation of Genomic Selection will require dealing with:

- Genomic predictions are highly dependent on the relationship between the training population and the selection candidates.
- Since genetic markers are not causal, the persistence of the genomic predictions will decline with each generation, and will be relatively ineffective for populations not related to each other.

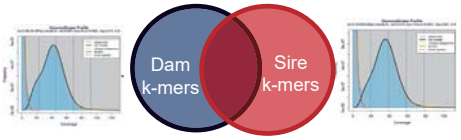
Ultimate Solution

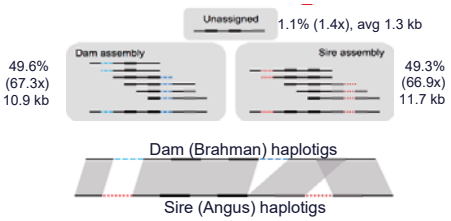
Identify the causal variants

Better genomes and functional genomics research is needed to make this a reality

Trio-based binning

• K-mer profiling of each parent (Illumina, 60x)





• Sequencing of the F1 (PacBio, 120x)

Angus x Brahman F1

Complete assembly of parental haplotypes with trio binning. Koren et al. (2018)

We get much better assemblies through the Trio binning approach to sequencing. Long range (PacBio) sequences from the offspring are used to extract short (Illumina) sequences originating from the parents. Maternal and Paternal long range sequences and their matching short sequences are then assembled, providing individual sequences of each chromosome from each parent. Sequence polymorphisms are used to match offspring to origin, so the more divergent the parents, the better this works

The trio-based method : sequence parents and offspring

short-read (300 bases)
sequencing of parents (Angus and Brahman)

CCTCAGATACCCGATCAACG
TTACGACCTCTCAAGCCCTA
TCCGAAATCATAGCGGGCTA

ATATTTCACTTTCTGAGACT
GGGCACATCATTTACGTACG
TATAAAGATATACCCTCTCG

long-read (15,000 bases)
sequencing of offspring (F_1 Angus x Brahman)

....TTAACCCTATCTAGATCTTTAGCGCAATTACGATCCGAAATCATAGCGGGCTATATGGCCTAGTCAGTCAATCGGACCTA....

Paternal (Angus) read "bin"

....GATACCCTCATTACCTTAGAGTCCTTACGGCATATAAAGATATACCCTCTCGGGGTCACATCCTAGATCTTACCGATTAA....

Maternal (Brahman) read "bin"

Separately assemble the Angus and Brahman
bins "haplotigs" (versus "contigs")

Interspecies crosses maximizes contrast between parental genome contributions

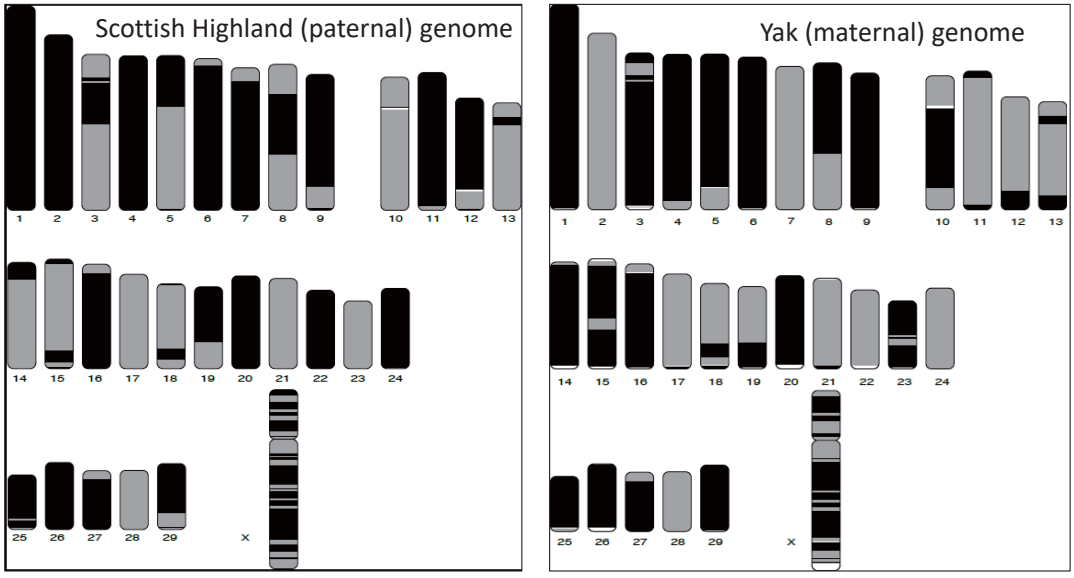
"Duke"
Scottish Highland

"Molly"
Imperial Yak

Taking this further, very divergent crosses should work really well

"Esperanza"
Yaklander cattle/yak F_1
heterozygosity = 1.3%

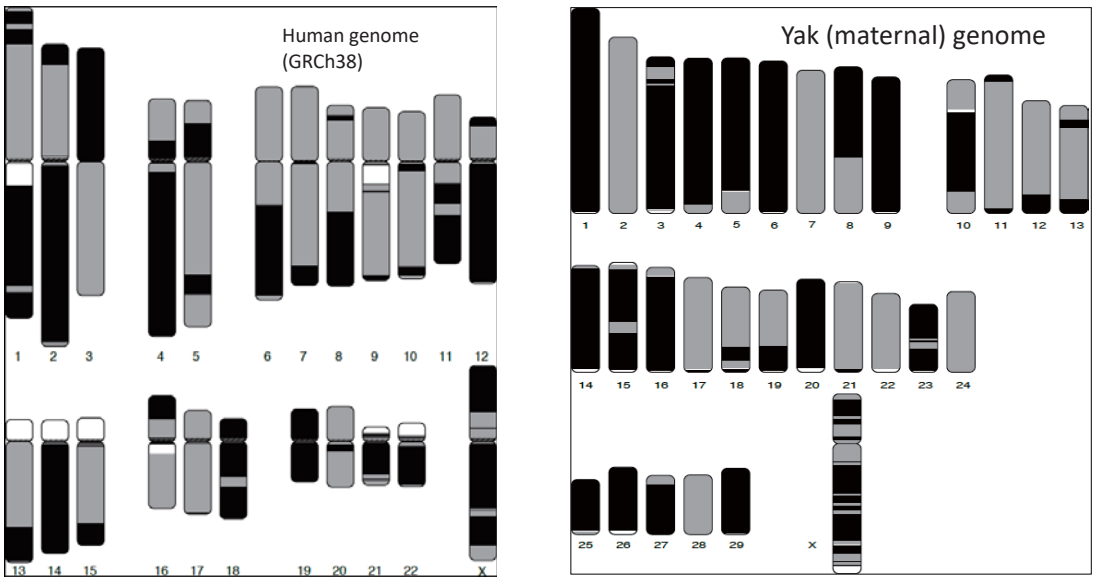
Yaklander interspecies F_1 resulted in the best sequence ever produced



Courtesy : Sergey Koren

The black regions are complete ordered contiguous sequence. Whole chromosomes resulted

Yaklander assemblies are close to (or better than) the current human genome



Using genome sequences, we can identify polymorphisms that are predicted to alter protein function, including loss of function (LOF). Protein function changes are immediate candidates for functional DNA variation.

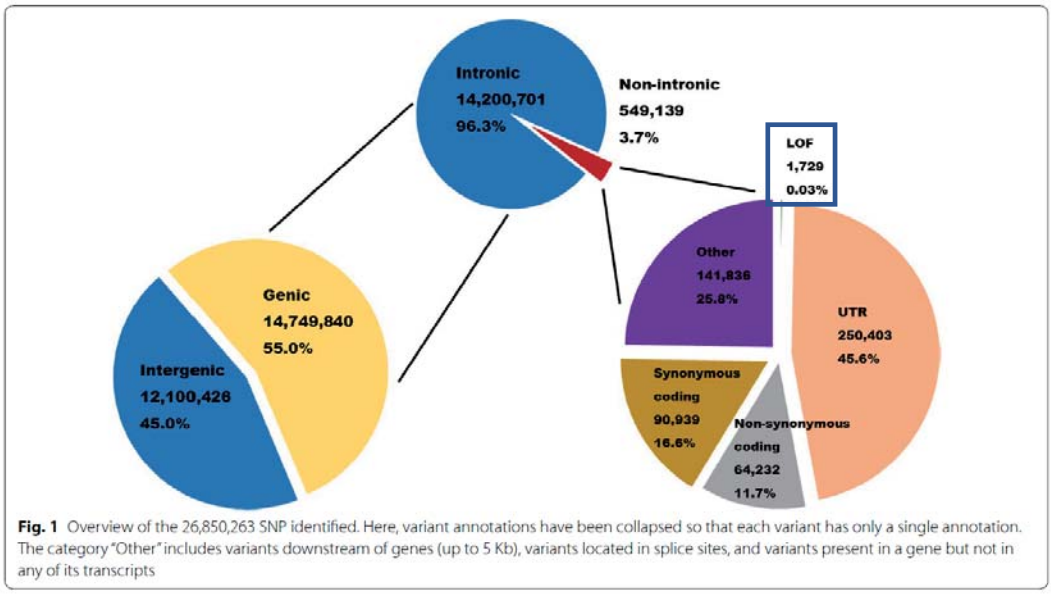
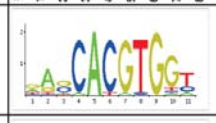
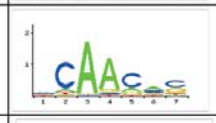
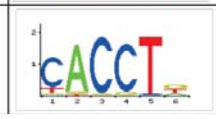


Fig. 1 Overview of the 26,850,263 SNP identified. Here, variant annotations have been collapsed so that each variant has only a single annotation. The category "Other" includes variants downstream of genes (up to 5 Kb), variants located in splice sites, and variants present in a gene but not in any of its transcripts

Impact of butyrate on rumen epithelial cells From Keel et al., 2018

By understanding transcription factor binding sites, we can find polymorphisms that alter gene expression for important traits, or predict their phenotype

Meq Motifs and Gene Expression

Up regulated genes			Down regulated genes		
Motif 1 (AP-1 like)		85.6%	Motif 2 (MERE-II)		84.1%
NFE2L2		5.9%	Myc		5.3%
CREB1		4.2%	BRCA1		3.1%
NR2F1		2.1%	ZEB1		2.2%

Percentages in the figure are the percent of up or down regulated genes that had these motifs in their promoter region.

Combined with SNP information, we can make predictions as to whether a gene will be up- or down-regulated by MDV Meq and ultimately, the response of the bird to Marek's disease infection

Summary:

- Genetic selection using quantitative methods has brought miraculous improvements an animal production
- Genome enhanced selection has already doubled the rate of progress in dairy cattle, and increased selection response 30-50% (depending on the trait) in other livestock species
- Further progress in many livestock species will be obtained by monitoring the actual sequence polymorphisms that alter phenotype
- Dramatic improvements in sequencing technology and functional genomic analysis will deliver that progress