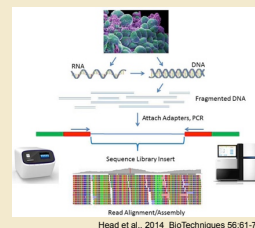


Low-pass sequencing to genotype cattle: Promises & Problems

Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer.

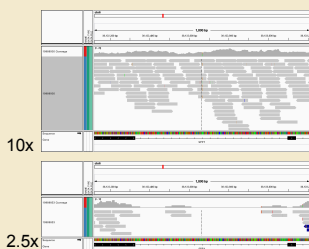
Genome sequencing

- cannot read chromosome sequence from end to end
- random process
 - "library" of randomly fragmented DNA
 - read ends of random fragments
 - align reads to reference assembly

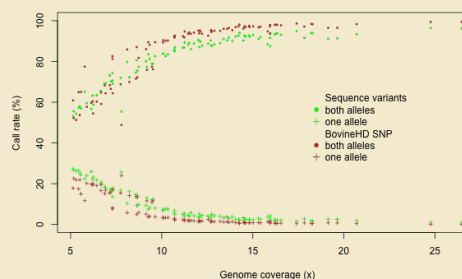


Genome coverage

- bases read / genome length
- substantial variation around average coverage
- portion of genome read increases with coverage



Genotyping calls from sequence

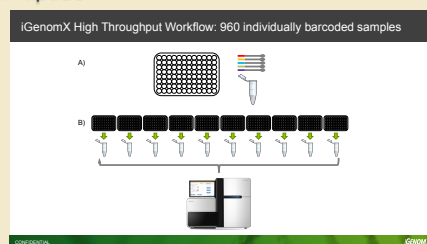


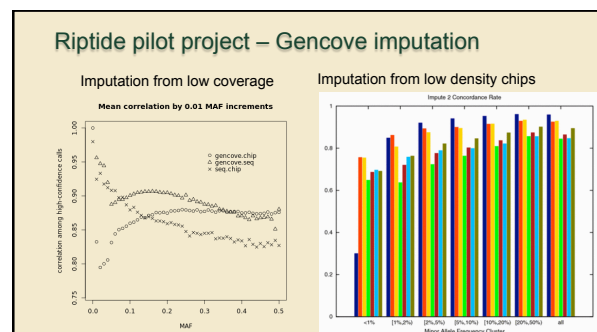
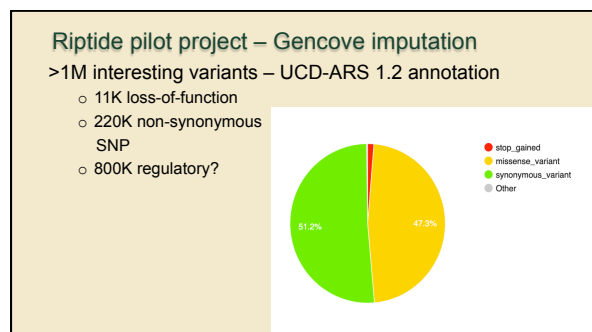
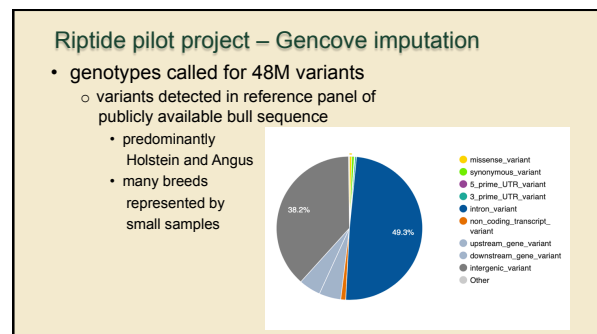
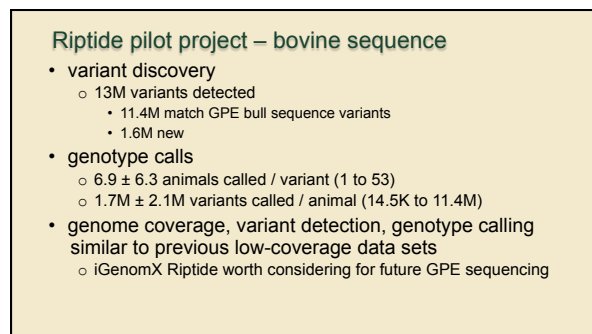
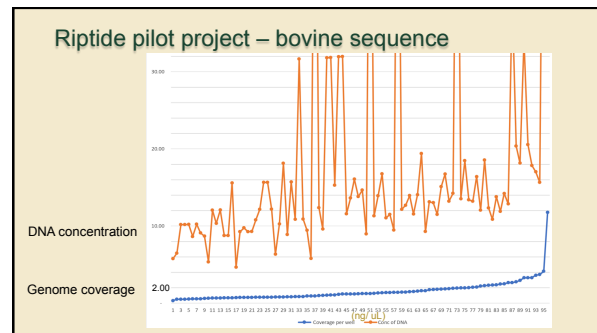
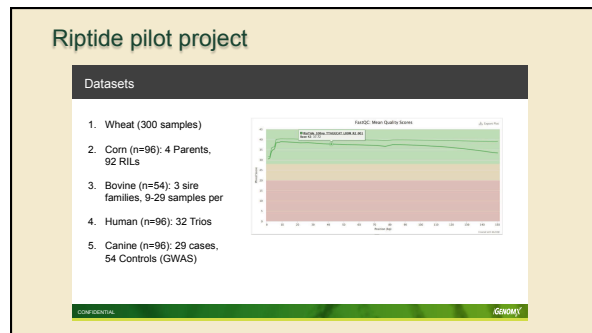
using low-pass sequence

- variant discovery
 - same cost and effort to sequence many individuals at low coverage as few individuals at high coverage
 - broader sampling to detect sequence variation in population
- genotyping?
 - low direct call rate
 - imputation – match low-coverage reads to reference haplotypes
 - higher power for genome-wide association studies
 - Li et al., 2011; Pasanuic et al., 2012; Gilly et al., 2018
 - iGenomX Riptide pilot project
 - requested bovine samples for sequencing

iGenomX Riptide

low cost per sample library preparation





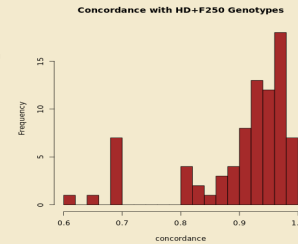
Riptide pilot project – Gencove imputation

- 21 animals in first set imputed by Gencove
 - 2 sire-progeny pairs
 - zero parentage SNP exclusions between Charolais bull and progeny (6 replicates)
 - 15 or 16 parentage SNP exclusions between Angus bull and progeny (6 replicates)
 - 6 to 17 exclusions between Angus-sired progeny and other animals
 - zero exclusions between pair using chip genotypes
 - 64% concordance between progeny Gencove & chip calls
 - sample ID mixup?
 - 92% to 99% concordance between replicated sire Gencove & chip

Riptide pilot project – Gencove imputation

Agreement between Gencove and HD+F250 genotypes

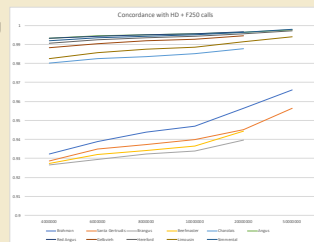
- good agreement possible
- why not all samples?
 - sample ID
 - contamination
 - imputation reference



GPE sequence – Gencove imputation

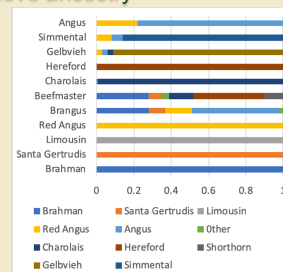
GPE sequence downsampling

- one bull from each Cycle VII breed, indicus-influenced composites, Brahman
- downsampled to 0.4x, 0.6x, 0.8x, 1x, 2x



GPE sequence – Gencove ancestry

Breed composition of downsampled bulls



GPE sequence – Gencove

- Strong agreement between between GPE HD+F250 genotypes and Gencove calls from downsampled British-breed bulls shows genotype accuracy possible for imputing from low-pass sequence
- Weaker agreement for Continental and indicus-influenced breeds suggests need for broader representation of those breeds in the reference panel
- Unexpected ancestry suggests need for broader reference of all breeds?
 - GPE sequence available
 - hybrid taurus – indicus genome?

GPE vs Riptide – Gencove imputation

- Lower agreement between for Gencove calls from Riptide sequence suggests sample contamination
 - physical contamination – low level sample mixing?
 - "index hopping" - sequence barcodes mis-assigned, reads for one barcode may represent more than one sample
 - exacerbated by variation in DNA concentration
 - mitigated by up-front QC of input DNA, steps added to library prep

low-pass sequencing & imputation

- Promises
 - genotype calls for comprehensive set of known sequence variants
 - 50K, HD, functional variant panels can be extracted
 - replace 50K with variants more likely to affect phenotypic variation
 - reduce dependence on LD between 50K & QTL
 - enable more accurate genomic predictions across breeds and crosses
 - lower cost than current chips
 - encourage complete genotyping of all seedstock calves
 - reduce bias in genetic evaluations due to selective genotyping
 - justify genotyping commercial calves
 - genomic predictions to support calf management and marketing decisions
 - heifer retention; genetic potential for growth, meat quality

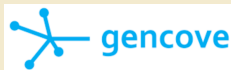
low-pass sequencing & imputation

- Problem
 - genotype call accuracy too low
 - addressable
 - imputation reference – broader sampling of all cattle
 - DNA QC and library preparation
 - tests with human samples underway
 - planning further bovine tests

Acknowledgments



Paul Doran
Keith Brown



Joe Pickrell
Jeremy Li
Tomaz Berisa



NEOGEN

Stewart Bauck



Entire crew involved
with GPE, tissue sampling
& repository, sequencing,
...
(too many to name)

Questions?

