2023 Genetic Prediction Workshop

Opportunities and obstacles to enhancing beef cattle evaluation with sequence data

Thank You

- Program advisors
  - Darrh Bullock
  - Cedric Gondro
  - Larry Kuehn
  - Megan Rolf
  - Troy Rowan
  - Matt Spangler
  - Mark Thallman
  - Bob Weaber
- BIF board & staff
  - Angie Denton
  - Bob Weaber
- Speakers, panelists, moderators and participants

2023 GPW Format

- Three half-day sessions (same as previous GPW)
  - Technical talks followed by reaction from industry

    Please participate – if you have questions, ASK!

- New in 2023
  - Evening poster session

    Get to know grad students, post-docs & young professionals

Sequence in USMARC Genetics Projects

Before low-pass

- Variant discovery – sequence influential GPE bulls to detect variation
  - Develop assays to directly genotype interesting variants

  Easier said than done
  - Custom chips
    - Design and manufacturing is expensive for low volume chips
    - Limited shelf life
  - Targeted sequencing
    - Little success
    - Scalability to 1000s of variants?

Sequence in USMARC Genetics Projects

Before low-pass

- Experimental selection for functional alleles (SFA)
  - Variation in loss-of-function alleles observed in GPE bull sequence
    - ~600 to 900 LOF alleles per bull
  - Could selection against LOF decrease genetic load – increase fitness? (Gary Bennett)
    - Selection initiated in MARC I, II and III composites, Angus
      - Populations split into select & control lines using LOF counts from GGP-F250 chip
        - ~200 to 250 LOF per animal
      - Select replacements picked by GGP-F250 LOF counts
      - Control replacements picked randomly

Sequence in USMARC Genetics Projects

Initial low-pass

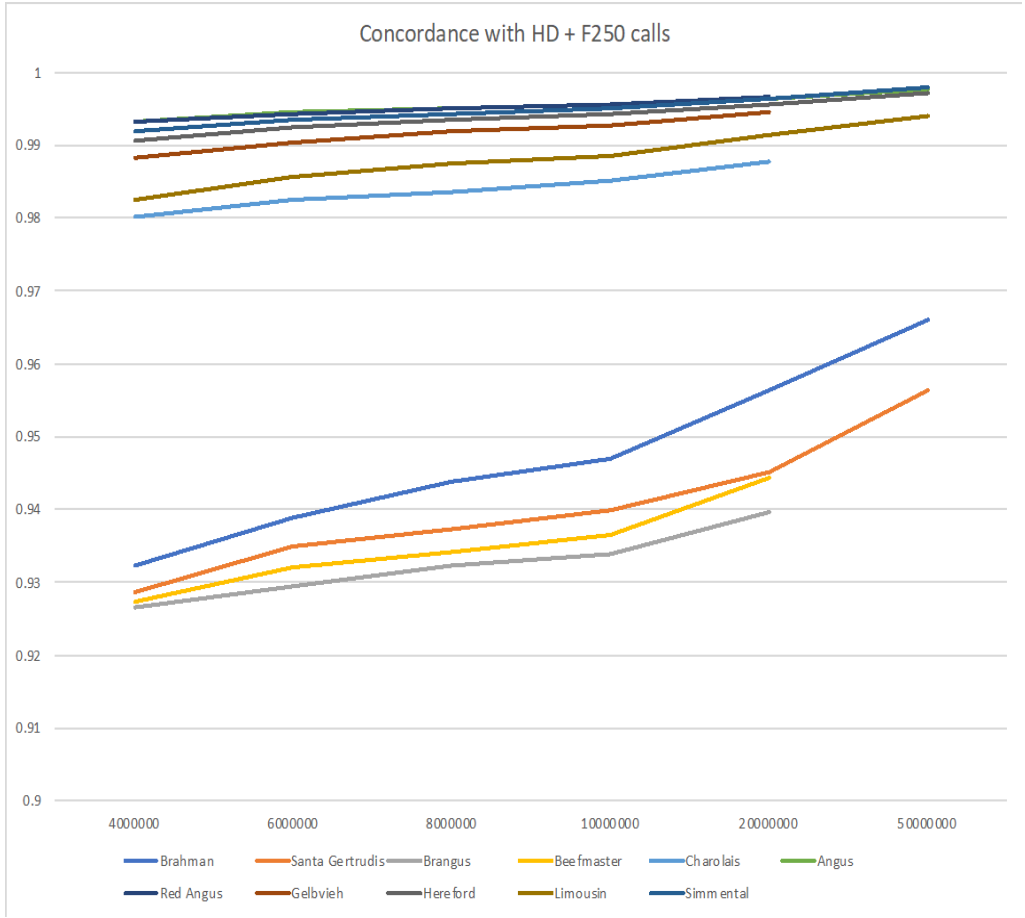
- Samples requested to test multiplex library preparation
  - Low-coverage sequencing with samples of GPE cattle, wheat, corn, humans and dogs in same run
  - Potential cost/sample << chip genotypes
  - Mixed results
    - Gencove imputation with ~250 animal panel
    - Concordance up to 99% with chip genotypes
    - Average concordance ~65% - sample contamination apparent

    - Success with some samples suggests low-pass worth investigating

Sequence in USMARC Genetics Projects

Further investigation (USMARC and Gencove)

- Expanded reference to include additional SRA and GPE sequence
  - 946 animal reference
- GPE downsampled deep sequence to mimic low-pass
- Low-pass is a promising alternative to chip genotypes
- Simple, low-cost approach to genotyping specific variants
  - Avoid frustration and expense associated with custom chips or targeted sequencing



Sequence in USMARC Genetics Projects

Current low-pass

- In-house sequencing and imputation
  - 970 animal reference (added SFA bulls)
  - GLIMPSE imputation
    - 64 million SNP and indels
  - Determine unknown sires by counting exclusions
    - ~5000 high MAF SNP
  - Extract “interesting” variants according to annotation
    - 1.5 million variants in coding sequence, non-coding RNA, UTR
  - Fill low-probability calls with pedigree imputation
    - ~9000 GPE with low pass
    - `27,000 GPE with chip genotypes

Sequence in USMARC Genetics Projects

Selecting functional alleles with low-pass

- SFA calves sequenced and imputed by Gencove (2020-2021)
- In-house sequencing and imputation (2022-)
- Sires determined by counting exclusions
- Low-probability calls filled by pedigree imputation
  - ~900 to 1350 LOF per animal
  - Mean 34 LOF difference between select and control
    - Larger between-line differences in MARC composites (~40) than Angus (13)

Sequence in USMARC Genetics Projects

Future work with low-coverage sequencing and imputation

- Reference panel composition & construction algorithms
- Pan-genome aware imputation
  - Major structural variation
- Low-pass by-products
  - Mitochondrial DNA
  - CNV
  - Sex chromosome coverage – QC metric?
- Low-coverage long reads

Sequence in USMARC Genetics Projects

Low-pass challenges

- Storage
  - Plenty of temporary space for computing, not so much to store results
    - Strategies to better utilize available space
- Genotype accessibility
  - Enable collaborators to access genotypes – low-pass genotypes would overwhelm existing relational database
    - Queryable system to identify animals available for genotype-specific treatments