The Canadian Cattle Genome Project

Interbeef
Dublin - 25 Nov 2014

John J. Crowley
Genome Canada Project
Genome Canada Project

• Involves 7 main Canadian breeds plus a hybrid plus 2 research populations
• Sequence 300+ influential animals
• Genotype ~18,000
• Feed efficiency (FE) and meat quality
So far....

<table>
<thead>
<tr>
<th>Data Type</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td>378</td>
</tr>
<tr>
<td>HD</td>
<td>4825</td>
</tr>
<tr>
<td>50k</td>
<td>5007</td>
</tr>
<tr>
<td>50k with feed efficiency</td>
<td>8027</td>
</tr>
</tbody>
</table>
Sequencing

• ~8X coverage (1 or 2/breed @ 20+)

• Illumina HiSeq platform (and some SOLiD)

• Computing resources from WestGrid to simplify and speed up data analysis
  – 50 TB spinning disk storage (backed up to tape) on Silo storage system (3150 TB).
  – 30 core years and 6 TB on Jasper cluster (2880 cores, 355 TB).
Sequence processing tools

- **FastQC** – sequence data quality control checks
- **BWA** – mapping reads to reference.
- **Picard** – manipulating and summarizing BAM files, duplicate read detection.
- **SAMtools** – variant discovery and genotyping.
- **GATK** – variant discovery and genotyping.
- **Perl** and **Bash** scripts for submitting jobs, comparing sequence genotypes and reports
Sequencing – concordance with HD

Fold Coverage After duplicate Removal

Courtesy of Paul Stothard
Variant Discovery

• Output of run 4.0 of 1000 bulls
  – 35,201,962 SNP
  – 1,714,895 INDELs
  – Other variants to be quantified
    • CNVs, inversions, translocations
    • Causative mutation should be in the data
    • Fine mapping of QTL
Imputation to Sequence

Average accuracy (Concordance rate) of single breed whole-genome sequence imputation of 28.3 million SNP genotypes using FImpute and Beagle software.
# Prediction Accuracy

<table>
<thead>
<tr>
<th>Dataset</th>
<th>n</th>
<th>Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANC</td>
<td>1946</td>
<td>AN, CH</td>
</tr>
<tr>
<td>KIN</td>
<td>923</td>
<td>Beef synt. 1 &amp; 2, Dairy synt</td>
</tr>
<tr>
<td>PG1</td>
<td>4359</td>
<td>HE, AN, RA, CH, M4, TX</td>
</tr>
<tr>
<td>UoG</td>
<td>1972</td>
<td>AN, CH, SM, LM, PD, GV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GBLUP1</th>
<th>GBLUP2</th>
<th>BayesC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADG</td>
<td>0.35 (0.56)</td>
<td>0.29 (0.57)</td>
<td>0.33 (0.52)</td>
</tr>
<tr>
<td>DMI</td>
<td>0.46 (0.70)</td>
<td>0.37 (0.60)</td>
<td>0.54 (0.71)</td>
</tr>
<tr>
<td>RFI</td>
<td>0.35 (0.46)</td>
<td>0.34 (0.51)</td>
<td>0.24 (0.63)</td>
</tr>
</tbody>
</table>

*Highest accuracy in parenthesis
Funders and Sponsors

Alberta Innovates BioSolutions

Canadian Beef Breeds Council

Genome Canada

ALMA Alberta Livestock and Meat Agency Ltd.

Alberta Innovates Technology Futures

Genome Alberta

livestock gentec
Questions?