Comparative structural genomics in Buffalo using the Cattle reference genome

Presentation by Derek Bickhart

The Scale of Genetic Variation

Variation as a function of size
- Single nucleotide variations – SNP (human millions of variants)
- Indels – Insertions/Deletions (1 bp – 1000 bp)
- Mobile Elements – SINE, LINE Transposition (300bp - 6 kb)
- Genomic structural variation (1 kb – 5 MB)
  - Large-scale Insertions/Deletions (Copy Number Variation: CNV)
  - Segmental Duplications (> 1kb, > 90% sequence similarity)
  - Chromosomal Inversions, Translocations, Fusions.

CNV detection by NGS

Reference Genome

Read Pair (RP)

Concertant

Deletion

Insertion

Inversion

Read Depth (RD)

Averaeng/Baseline

Duplication

Deletion

Split Read (SR)

A candidate MGE

Software used to identify variants

- RAPTR-SV
  - Combined SV detection
  - Read Pair - Split Read

- MGE transposition
  - Custom perl scripts
  - Btau4.0 annotation

- Availability:
  - Manuscript submitted to Bioinformatics
  - https://github.com/njdbickhart/raptr-sv
Methodology

- 14 Italian River Waterbuffalo
  - 5 - 10 X coverage
  - Illumina Hiseq 2000 paired end
- One 20 X coverage River Waterbuffalo

Summary of Results

<table>
<thead>
<tr>
<th>Method</th>
<th># Calls</th>
<th>Megabases</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD CNVs</td>
<td>1,425</td>
<td>59.0</td>
</tr>
<tr>
<td>RAPTR-SV Deletions</td>
<td>21,578</td>
<td>37.8</td>
</tr>
<tr>
<td>RAPTR-SV Duplications</td>
<td>5,054</td>
<td>26.7</td>
</tr>
<tr>
<td>MGE events</td>
<td>2,808</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- Non-redundant coverage: 158 Megabases
- Impact on Genes
  - 3174 annotated Refseq genes
  - 2271 genes affected by deletions

Validation

- 144 BAC FISH experiments
<table>
<thead>
<tr>
<th>Results</th>
<th>BACs</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>84</td>
<td>69.42%</td>
</tr>
<tr>
<td>Not</td>
<td>37</td>
<td>30.58%</td>
</tr>
<tr>
<td>No result</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

- 19 qPCR tested genes
  - 9 out of 19 agreement (47%)
  - Gene structural differences?
  - Primer design issues?

High variability in immune gene copy number

T2R65A is deleted in every individual


Mobile Genetic Elements (MGE)

- MGE within gene promoter regions: 41 genes
- Impact expression profile?

<table>
<thead>
<tr>
<th>GO Term</th>
<th>PValue</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0030496~midbody</td>
<td>0.01993</td>
<td>VPS4A, MAPRE3</td>
</tr>
<tr>
<td>GO:0019904~protein domain specific binding</td>
<td>0.020867</td>
<td>YWHAZ, NCK1, VPS4A</td>
</tr>
<tr>
<td>GO:0007010~cytoskeleton organization</td>
<td>0.049714</td>
<td>SORBS1, TPPP, NCK1</td>
</tr>
<tr>
<td>GO:0008017~microtubule binding</td>
<td>0.061751</td>
<td>TPPP, MAPRE3</td>
</tr>
<tr>
<td>GO:0007015~actin filament organization</td>
<td>0.06299</td>
<td>SORBS1, NCK1</td>
</tr>
</tbody>
</table>
Implications on Biology

• Deleted gene families

• Likely changes in gene expression
  – Copy number of gene families
  – MGE upstream of structural proteins

• Putative wallowing adaptations
  – Trappin family (PI3)
  – CATHL4, LAP, TAP

Acknowledgements

• USDA ARS
  – George Liu
  – Lingyang Xu
  – Yang Zhou
  – Tad Sonstegard
  – Curt Van Tassell

• The International Buffalo Genome Consortium

This work has been supported in part by the National Research Initiative Grant no. 2011-67015-30183 and USDA NIFA grant no. 2013-67015-20951

Questions?