Performance of whole NGS pipelines containing alignment and SNP calling tools

M. Mielczarek1,2, J. Szyda1,2, B. Gulbrandsen3
1Wroclaw University of Environmental and Life Sciences, Department of Genetics, Kozuchowska 7, 51-631 Wroclaw, Poland, 2National Research Institute of Animal Production; Krakowska 1, 32-083 Cracow-Balice, Poland; 3Aarhus University, Department of Molecular Biology and Genetics, Blichers Allé 20, 8830 Tjele, Denmark; magda.a.mielczarek@gmail.com

Objectives

- Evaluation of different NGS data processing pipelines
- In silico investigation of interactions between programs

Conclusions

- SNP calling pipelines provide different results for the same dataset
- Results from NGS and microarrays agree, but not completely
- True SNPs are characterized by high quality scores
- Mappers perform better with GATK, except for SMALT
- SAMtools-based pipelines run faster

Data set

4 Danish Red Dairy bulls:
- Hiseq 2000, whole genome sequences average coverage = 10 X
- Illumina BovineHD BeadArray 770K SNP

Methods

I. Pipeline definition

Alignment to the reference genome (BWA-MEM, BWA-BT, Bowtie2, SMALT)

SNP calling for 770K sites (GATK, SAMtools)

II. Quality assessment

true SNP/genotype = confirmed by microarray

Results

Comparison of 770K SNP sites

<table>
<thead>
<tr>
<th>%</th>
<th>MEM_GATK</th>
<th>BT_GATK</th>
<th>BOW_GATK</th>
<th>SMA_GATK</th>
<th>MEM_SAM</th>
<th>BT_SAM</th>
<th>BOW_SAM</th>
<th>SMA_SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>true SNPs</td>
<td>&gt; 99.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

true genotypes

- **MEM_GATK**
- **BT_GATK**
- **BOW_GATK**
- **SMA_GATK**
- **MEM_SAM**
- **BT_SAM**
- **BOW_SAM**
- **SMA_SAM**

**Run times** (minutes)

<table>
<thead>
<tr>
<th>pipelines</th>
<th>MEM_GATK</th>
<th>BT_GATK</th>
<th>BOW_GATK</th>
<th>SMA_GATK</th>
<th>MEM_SAM</th>
<th>BT_SAM</th>
<th>BOW_SAM</th>
<th>SMA_SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEM_GATK</td>
<td>3617</td>
<td>4080</td>
<td>3455</td>
<td>4852</td>
<td>3815</td>
<td>3815</td>
<td>3274</td>
<td>4646</td>
</tr>
<tr>
<td>BT_GATK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOW_GATK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMA_GATK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEM_SAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BT_SAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOW_SAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMA_SAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Acknowledgements