Genotype Imputation in Nelore Cattle

R. Carvalheiro¹, J. Sölkner², H. Neves¹, Y.T. Utsunomiya¹, A.M. Pérez O’Brien², S. Boison², M. da Silva³, C.P. VanTassell⁴, T.S. Sonstegard⁴, J. McEwan⁵, F.S. Schenkel⁶, J.F. Garcia¹
¹Universidade Estadual Paulista, São Paulo - Brasil, ²University of Natural Resources and Life Sciences, Vienna - Austria, ³EMBRAPA, Juiz de Fora - Brasil, ⁴USDA-ARS, Beltsville - USA, ⁵AgResearch, Invermay - New Zealand, ⁶University of Guelph, Guelph - Canada

Objective
Test imputation efficiency in Nelore using different SNP densities and software
• Population vs. Pedigree based

Conclusions
1. FImpute outperformed BEAGLE in imputation accuracies
2. Pedigree information increased accuracy just for 6k
3. Relatives in reference panel increased accuracy (BEAGLE)
4. LD between markers was important for imputation accuracy

Table 1. Mean percentage ± SD of correctly called genotypes (PERC) and correlation between imputed and observed genotype (CORR) for different scenarios of customised SNP panels

<table>
<thead>
<tr>
<th>SNP panel</th>
<th>BEAGLE</th>
<th>FImpute – Pedigree</th>
<th>FImpute – Pedigree free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%PERC</td>
<td>CORR</td>
<td>%PERC</td>
</tr>
<tr>
<td>6k</td>
<td>87.7 ± 4.4</td>
<td>89.8 ± 3.9</td>
<td>90.6 ± 4.1</td>
</tr>
<tr>
<td>11k</td>
<td>96.1 ± 2.3</td>
<td>97.0 ± 1.9</td>
<td>97.6 ± 1.5</td>
</tr>
<tr>
<td>15k_Ev</td>
<td>95.5 ± 2.5</td>
<td>96.4 ± 2.1</td>
<td>97.2 ± 1.8</td>
</tr>
<tr>
<td>15k_LD</td>
<td>96.4 ± 2.2</td>
<td>97.1 ± 1.8</td>
<td>97.9 ± 1.4</td>
</tr>
<tr>
<td>46k</td>
<td>98.2 ± 1.2</td>
<td>98.6 ± 1.1</td>
<td>99.1 ± 0.7</td>
</tr>
</tbody>
</table>

Figure 1. Effect of genomic relationships between reference and test set on imputation accuracy

Figure 2. Impact of Linkage Disequilibrium on accuracy

Background
• Genotype imputation reduce costs of breeding programs implementation
• Available low density panels developed using Bos taurus breeds
• Need to develop alternative low SNP panels for Bos indicus (Nelore breed)

Materials and Methods
• 793 sires used as reference and 202 sires used as testing set

Scenarios
Illumina 6k, 11k (Illumina 6k + customised*), evenly spaced 15k (Ev), 15k customised*(LD), 46k (Illumina 6k + customised*)

Software
FImpute, BEAGLE

*SNPs with the highest MAF and LD within 20 SNP-windows across the genome