Analysis of global genetic variation in pig: Discovery of novel SNPs by whole genome re-sequencing

Bujie Zhan, Jakob Hedegaard, Christian Bendixen and Frank Panitz
Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, 8830 Tjele, Denmark

Objectives
Our goal is to investigate genetic variation in pig among different individuals at the whole genome level and survey functional effects for some essential variations.

Background
- A close to finished Sus scrofa reference genome is available.
- Single nucleotide polymorphism (SNP), the most frequent type of variation in the genome. SNPs can affect how individual develop diseases and respond to pathogens, chemicals, drugs, vaccines, thus making it a valuable resource of information for livestock breeding programs.
- The SNP information available for Sus scrofa, however, is still limited when compared to other sequenced mammalian species.

Materials and Methods
- **Reference genome** - Ensembl Sus scrofa 9 [www.ensembl.org]
- **Data produced** - About 260 Gb Illumina solexa paired-end sequence generated for 4 Duroc boars by whole genome shotgun sequencing.
- **Calling SNPs** - Two algorithms (tools) were applied, bwa[1] and gigaBayes[2], and common calls shared by both algorithms were collected.
- **SNPs annotation** - Customized perl script by using Ensembl Sscrofa9 features information.
- **Prediction of functional effects on non-synonymous SNPs** - SIFT[3] and PolyPhen[4].

Results

**Table 1.** SNPs called in all sequenced four boars.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Coverage</th>
<th>SNPs Called by gigaBayes &amp; bwa</th>
<th>Homozygote sites on 60K chip</th>
<th>Non synonymous SNPs (Homo + Hetero:nsSNPs)</th>
<th>Deleterious nsSNPs (Homo + Hetero)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear 1</td>
<td>78.7 %</td>
<td>Total: 3,388,663 Homo: 1,944,896 Hetero: 1,543,767</td>
<td>6471 (0.2%)</td>
<td>4456+1025 = 5481</td>
<td>4563+134 = 4697</td>
</tr>
<tr>
<td>Bear 2</td>
<td>78.5 %</td>
<td>Total: 2,697,010 Homo: 1,325,607 Hetero: 1,371,403</td>
<td>4959 (0.2%)</td>
<td>3395+2016 = 5411</td>
<td>306+134 = 440</td>
</tr>
<tr>
<td>Bear 3</td>
<td>78.1 %</td>
<td>Total: 2,494,136 Homo: 1,094,649 Hetero: 1,399,487</td>
<td>6315 (0.2%)</td>
<td>4215+2286 = 6497</td>
<td>468+116 = 584</td>
</tr>
<tr>
<td>Bear 4</td>
<td>71.1 %</td>
<td>Total: 2,471,558 Homo: 1,035,056 Hetero: 1,436,502</td>
<td>6599 (0.2%)</td>
<td>5868+3922 = 9790</td>
<td>701+165 = 866</td>
</tr>
</tbody>
</table>

Figure 1. Number of SNPs called as a function of amount of sequences.

Figure 2. Uniquely mapped sequence coverage on all chromosomes.

Figure 3. SNPs called for all sequenced 4 boars on chromosome 8. Some regions show common SNP density patterns appeared in all 4 individuals maybe indicating evolutionary conserved variants or assembly errors in the reference.

Figure 4. A higher percentage of homozygous SNP compared to heterozygous variants is shared in all sequenced boars. False positive SNP calls caused by errors in reference may introduce common homozygote calls in all individuals.

Summary

With the current sequencing depth, millions of SNPs, mostly novel polymorphisms can be identified for each boar we sequenced. Comparisons with genotyping data generated with Illumina porcine 60k chip indicate a detection rate about 80%(Table 1). Combined with Ensembl pig genome features, thousands of polymorphisms in coding regions were annotated as non-synonymous SNPs and hundreds of them were predicted to be deleterious. All the variations discovered in this study provide a valuable resource to be used as high density genetic markers for genotyping and for investigating phenotypic variations.

Meanwhile in order to increase the SNP detection rate and accuracy, additional shotgun sequencing and an updated reference assembly may needed.

Acknowledgements
This work was supported by grants of the Danish National Advanced Technology Foundation, Pig and Health project and Danish Strategic Research Council (Nabiit; grant no. 2106-07-0021).

Reference

Figure 5. A higher percentage of homozygous SNP compared to heterozygous variants is shared in all sequenced boars. False positive SNP calls caused by errors in reference may introduce common homozygote calls in all individuals.