Analyses of the breed integrity of the Goettingen Minipig using pool-sequencing

C. Reimer¹, N.T. Ha¹, A.R. Sharifi¹, S. Weigend², J. Geibel¹, H. Simianer¹

¹University of Goettingen, Department of Animal Sciences, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany
christian.reimer@agr.uni-goettingen.de (Corresponding author)
²Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics, Höltystraße 10, 31535 Neustadt-Mariensee, Germany

Summary

The Goettingen Minipig (GMP), one of the smallest pig breeds, is an established animal model in medical research. The GMP is bred in five isolated stocks and it is of foremost importance to ensure the integrity of the different strains. We sequenced two DNA-pools from every stock and added samples from a diverse set of pig breeds from an earlier study. We estimated the pairwise fixation index for all pools, conducted phylogenetic (UPGMA) and principal component analyses (PCA) and functionally annotated all loci. The PCA revealed, that the GMP is easy to discriminate from all other breeds while there also is remaining differentiation between the five stocks. Annotation of all loci showed that critical functional classes, such as stop codons, were relatively underrepresented and rarely located in genes important in minipig breeding. We conclude that there is a certain level of stratification within the GMP, which might not be compromising breed integrity yet.

Keywords: Goettingen Minipig, pool sequencing, differentiation

Introduction

The Goettingen Minipig (GMP), one of the smallest pig breeds in the world, was established by crossing Minnesota Minipigs, Vietnamese Potbellied Pigs and German Landrace at the former Institute of Animal Breeding and Genetics of the University of Goettingen in the 1960s (Simianer and Köhn 2010). The university owned stock is kept at the research farm Relliehausen (RE). In 1992, a collaboration with Ellegaard Goettingen Minipigs A/S from Dalmose, DK, was started by opening unit DA1. In 2006, animals from DA2 (descendent from DA1) were brought to North Rose, NY, as foundation for a North American population. DA1 was closed down. The next separation happened in 2009 with the opening of a second barrier in Dalmose (DA3). Since the opening of the first Asian facility in Nisshin, Japan (NI) in 2013, branched off from DA3, there are now five active breeding stocks in service worldwide without exchange of animals. Even though all stocks underlie a fully documented and centrally controlled breeding scheme and are bred for the same breeding goal, the genetic isolation might harbor the risk of stratification through genetic drift. As the GMP is today one of the standard non-rodent animal models in medical research, its uniformity and clear characterization are of foremost importance (Bollen and Ellegaard 1997).

This study aims at identifying the traces that separation might have left in the genomes of the different populations, predict their consequences and form the base for a decision-making process as to when interchange of animals is inevitable to maintain the breed integrity.
Material and Methods

30 females representative for the respective stock were chosen for DNA sampling from every unit based on measures of pedigree-based relationship. 20 viable DNA extracts were randomly assigned to two groups of 10 individuals each and equimolarly pooled. Pools were sequenced at a depth of 30X as paired reads on an Illumina X10. All reads were aligned to the reference genome susScr3 (build 10.2; Groenen et al. 2012) with BWA 0.7.2. (Li and Durbin 2009). The subsequent bam file preparation and variant calling followed the “GATK Best Practice” protocol (Broad Institute 2017). Due to unavailability of a high confidence learning SNP set, the 5% SNPs with highest quality contained in dbSNP, were chosen for variant recalibration from the raw callset. Corresponding variants of various pig breeds from an earlier study (Reimer et al. 2014) were added. Individual data was virtually pooled by summation of all reference and alternative reads, respectively. Monomorphic loci were discarded when a subset was used.

Reference allele frequency was calculated for every pool as number of reads supporting the reference allele, divided by the total coverage at the respective locus. Wright’s pairwise fixation index was estimated (eq.1; Eding and Bennewitz 2007). A UPGMA tree, based on 100 subsamples of 50’000 SNPs (‘phangorn 2.2.0’; Schliep 2011) and a principal component analysis were computed with R (R Core Team 2015). All loci were annotated with Ensembl’s variant effect predictor (McLaren et al. 2016).

Results and Discussion

The UPGMA tree (Figure 1) shows that the GMP can still be considered a very distinct breed when compared to other pig breeds. Resampling shows a high robustness of the estimated tree, even when subsets of 50’000 SNPs were used.

![UPGMA tree](image)

Figure 1: UPGMA tree of all analyzed breeds based on $F_{ST}$.

In the PCA (Figure 2), the first principal component (PC) explains 78% and the second 8%. The first PC explains the variation between the GMP and all other breeds, while the second discriminates GMP from European and Asian (including Mini-LEWE) populations. It is remarkable that the first component does not explain the difference between large pigs and minipigs, since the Mini-LEWE is also a minipig, but has a different genetic background than
the GMP. The PCA for the GMP pools only revealed that the DA units and the recently separated NI unit cluster together genetically. RE appears most distant from the other units, which may be explained by the long time since separation. To clarify if this led to critical functional differences, all highly differentiated SNPs were functionally annotated. In Figure 3 it is shown how the relative abundance of the functional SNP classes alters along the level of differentiation.

While, upstream and downstream gene variants show a steady increase towards higher $F_{ST}$ levels, intron variants and intergenic variants remain stable throughout the entire $F_{ST}$ spectrum. Interestingly critical classes were not represented at high differentiation, e.g. ‘stop_lost’ and ‘start_lost’, or were relatively underrepresented e.g. ‘stop_gained’ and ‘missense variants’.

![Figure 2](image_url)  
*Figure 2: PCA based on $F_{ST}$ of all breeds (left) and on the GMP pools only (right).*

![Figure 3](image_url)  
*Figure 3: Relative abundance of selected functional classes in dependence from $F_{ST}$, based on the $F_{ST}$ class 0 – 0.1. Split by functional classes without (left) and with (right) expected consequences on protein syntheses.*

Revisiting all deleterious SNP with $F_{ST} \geq 0.9$ (Table 1), seven loci were found when NR was contrasted against another pool and one comparing RE to DA3. Among the underlying genes are annotation artefacts and novel genes, but also TMEM63A, a membrane protein gene, and PHLDA2, which has been linked to intrauterine growth restriction in humans.
ZNF428, which contains the SNP differentiated between DA3 and RE has no obvious functional link to the GMP breeding goals. Our results support that the GMP is still clearly distinct from all other pig breeds, but inside the GMP, differentiation between RE, NR and a cluster of NI and DA2/3 can be detected. This is sensible, since the split of NI from DA3 was just four years ago and optimal representatives of DA3 were chosen as founders of NI. The functional annotation shows that differentiation happens rather in neutral than in critical genomic regions, and differences found might rather be due to drift than to selection. The few highly differentiated deleterious SNPs are located in genes without obvious functional relation to the typical attributes of the GMP and it seems unlikely, that they might compromise the functional integrity of the GMP. **Conclusions:** Even though genetic drift drives apart the different units genetically, the centralized breeding scheme has ensured breed integrity of the GMP so far and an exchange of animals between units does not yet appear to be necessary.

**Table 1: Missense alleles with deleterious consequence exhibiting F_{ST}=1.**

<table>
<thead>
<tr>
<th>Chr</th>
<th>Pos</th>
<th>Pop1</th>
<th>Pop2</th>
<th>Ens-ID</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>429'370</td>
<td>DA2</td>
<td>NR</td>
<td>ENSSSCG00000021597</td>
<td>PHLDA2</td>
</tr>
<tr>
<td>2</td>
<td>15'249'414</td>
<td>DA2</td>
<td>NR</td>
<td>ENSSSCG00000029368</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>46'206'421</td>
<td>DA3</td>
<td>RE</td>
<td>ENSSSCG0000003059</td>
<td>ZNF428</td>
</tr>
<tr>
<td>10</td>
<td>16'012'840</td>
<td>DA2</td>
<td>NR</td>
<td>ENSSSCG00000010854</td>
<td>TMEM63A</td>
</tr>
<tr>
<td>14</td>
<td>7'880'409</td>
<td>DA3</td>
<td>NR</td>
<td>ENSSSCG00000025094</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7'880'409</td>
<td>NR</td>
<td>NI</td>
<td>ENSSSCG00000025094</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>86'466'849</td>
<td>NR</td>
<td>NI</td>
<td>ENSSSCG00000020913</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>29'494'436</td>
<td>NR</td>
<td>RE</td>
<td>ENSSSCG00000024692</td>
<td></td>
</tr>
</tbody>
</table>

**Acknowledgements**

We acknowledge financial support by Ellegaard Goettingen Minipigs A/S.

**List of References**


