Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (*Meleagris gallopavo*) population

G. Marras¹, B.J. Wood¹,³, B. Makanjuola¹, F. Malchiodi¹, K. Peeters², P. van As², C.F. Baes³* & F. Biscarini⁴,5*

¹Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada
gmarras@uoguelph.ca (Corresponding Author)
²Hendrix Genetics Ltd., Spoorstraat 69, 5831 CK Boxmeer, Netherlands
³Hybrid Turkeys, 650 Riverbend Drive, Suite C, N2K 3S2, Kitchener, Canada.
⁴CNR-IBBA, Milan, Italy
⁵Cardiff University, Cardiff, UK
* Equal contributions

Summary

Runs of homozygosity have been used for genomic characterization and inbreeding estimation, or to detect selective signatures in the genome of mammal and avian species. Here we report runs of homozygosity and heterozygosity in the turkey genome (*Meleagris gallopavo*). A total of 5,297 turkeys from one commercial line collected between 2010 and 2016 were genotyped using a proprietary 60k array, and pedigree records of 773,414 individuals were available. Runs of both homozygosity and heterozygosity were detected. Relatively long and abundant runs of homozygosity were detected, with comparatively few runs of heterozygosity. Inbreeding coefficients calculated using ROH were higher than those calculated using pedigree. These results provide a preliminary characterization of the turkey genome in terms of runs of homozygosity and heterozygosity-rich genomic regions.

Keywords: runs of homozygosity, heterozygosity, turkey, autozygosity, Froh, R package

Introduction

The availability of dense SNP arrays has made investigating the distribution of both homozygosity and heterozygosity within livestock genomes possible. Segments of continuous homozygous genome, known as runs of homozygosity (ROH), can be useful for characterizing livestock genomes and understanding implications of strong selection. Although livestock genomes are mostly homozygous (polymorphic sites account for only 1-2% of the genome), one may alternatively look at where heterozygosity clusters into “runs of heterozygosity” (ROHet; Williams *et al.*, 2016): these are not actual “runs”, but rather heterozygosity-rich regions. ROHet can be used to study balancing or negative selection, introgression, admixture or hypervariable regions.

In chickens, runs of homozygosity have been used for genomic characterization and inbreeding estimation (Szwaczkowski 2017), or to detect selective signatures in the genome (Fleming *et al.*, 2017). The turkey genome has yet to be explored for such characteristics. Here we describe runs of homozygosity and heterozygosity in the turkey genome (*Meleagris gallopavo*).
Material and Methods

A total of 5,297 turkeys from one commercial line collected between 2010 and 2016 were analyzed. Blood samples were genotyped using a proprietary SNP array. Individuals and markers with a call-rate < 90% were removed, resulting in 56,450 SNP after editing. Pedigree records of 773,414 individuals with a maximum depth of 29 generations were available.

Runs of homozygosity (ROH) and “runs of heterozygosity” (ROHet) were detected using a consecutive method as described in Marras et al. (2014, 2017). Only ROH with a minimum of 50 SNPs and longer than 1 Mbps were considered. No missing or heterozygous genotypes were allowed within a ROH, and the maximum gap between consecutive SNPs was set at 10^6 bps. For ROHet, the minimum length was set at 20 SNPs and no homozygous or missing genotypes were allowed. To identify ROH and ROHet the R package “detectRUNS” was used (currently hosted in a private GitHub repository -https://github.com/bioinformatics-ptp/RoHet; released at publication).

Inbreeding was estimated from the pedigree genealogies (Fped) and from ROH (Froh). Froh was estimated as:

\[(1)\]

where \(L_{roh}\) is the cumulative sum of all ROH length in an individual, and \(L_{auto}\) is the length of the turkey autosomal genome (~900 Mbps) (McQuillan et al., 2008).

Results and Discussion

Table 1 shows the distribution of ROH and ROHet per length class obtained. An average total of 126.21 ROH and 57.79 ROHet, were identified per turkey. The number of ROH found was comparatively high compared to that in other livestock species, which may be a consequence of the strong directional selection and high level of autozygosity present in commercial turkey populations. The average ROH length (~1.7 Mbps) observed in turkey is longer than that reported in chickens (~< 0.5 Mbps, Fleming et al., 2016).

Table 1: Average number of ROHom and ROHet per bird, standard deviations in brackets.

<table>
<thead>
<tr>
<th>Class (Mbps)</th>
<th>ROHom</th>
<th>ROHet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average n Runs / bird</td>
<td>Mean length of ROH (Mbps) / Bird</td>
</tr>
<tr>
<td>1-2</td>
<td>37.52 (7.13)</td>
<td>1.49 (0.05)</td>
</tr>
<tr>
<td>2-4</td>
<td>27.39 (6.35)</td>
<td>2.75 (0.11)</td>
</tr>
<tr>
<td>4-8</td>
<td>8.30 (3.62)</td>
<td>5.28 (0.40)</td>
</tr>
<tr>
<td>8-16</td>
<td>1.88 (1.19)</td>
<td>10.08 (1.49)</td>
</tr>
<tr>
<td>&gt;16</td>
<td>1.10 (0.34)</td>
<td>19.06 (3.67)</td>
</tr>
<tr>
<td>Total</td>
<td>126.21 (17.71)</td>
<td>1.73 (0.20)</td>
</tr>
</tbody>
</table>
Inbreeding estimated from pedigree genealogies range from 0.06 to 0.33. The range of Froh is between 0.014 and 0.44. The correlation between Fped and Froh ranged from 0.21 to 0.37, depending on the length-class of ROH that was used. Such low correlations are related to the different distributions of inbreeding coefficients estimated from pedigree or molecular data (Figure 1). Using pedigree, the expected inbreeding is estimated, which in a highly structured livestock population tends to have similar values for a large proportion of the animals. The Fped distribution is narrow, with only few animals showing extreme values. In contrast, molecular data allows the estimation of realized inbreeding coefficients. As a consequence, the Froh distribution is wider, with a larger interquartile range and a finer-grain differentiation between animals. The average per-chromosome inbreeding estimated from ROH (Froh) is shown in Figure 2. The highest inbreeding was observed on chromosome 22, the lowest on chromosome 21. No ROH were found on chromosome 18, likely due to the small size of the chromosome, which is only 330,095 bps long.

![Figure 1: Distribution of Froh and Fped]

![Figure 2: Average Froh per chromosome along the turkey genome.]

Both in ROH and ROHet, the proportion of times any single SNP locus falls inside a run in the population can be calculated. Such SNP-inside-runs proportions can be plotted against the position of SNP loci along the chromosome, similar to the familiar GWAS Manhattan plots. Figure 3 shows such Manhattan plots for ROH (3a) and ROHet (3b). Chromosomes 10 and 22 had the highest proportion of SNP within ROH regions, while chromosomes 4, 8, 10 and 11 also contained regions close to 100%. Although the proportion of SNP within ROHet regions were not greater than 50% on any chromosome, heterozygous clusters could still be noted within the genome.
Figure 3: Manhattan plot for Runs of Homozygosity (a) and Heterozygosity (b)

Conclusions

This paper is the first to report runs of homozygosity and heterozygosity in the turkey genome. Relatively long and abundant ROH were detected. Heterozygosity islands do not seem to cluster on specific chromosomes. Inbreeding calculated from ROHs was higher than from the pedigree. These results provide a preliminary characterization of the turkey genome in terms of runs of homozygosity and heterozygosity-rich genomic regions.

List of References