Detection of deleterious genomic variation in domesticated animals

Martijn Derks[1], Hendrik-Jan Megens [1], Mirte Bosse[1], Christian Gross[2,3], Marcel Reijnders[3], Dick de Ridder[2], Martien Groenen [1]

[1]Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands
[2]Bioinformatics Group, Wageningen University, The Netherlands
[3]The Delft Bioinformatics Lab, Delft University of Technology, The Netherlands

Background

Small effective population sizes of livestock can lead to deleterious recessive alleles drifting more rapidly to higher frequencies, impacting production and animal welfare. In corporation with breeding companies we have access to large genotype and phenotype data sets. **These data sets can be used to test for statistical depletion, even absence, of certain haplotypes in homozygous state.** Significant depletion of haplotypes is an indication of decreased viability. The identification of deleterious alleles using a haplotype approach is a powerful tool originally developed by vanRaden. et al 2011 in cattle [1].

We also have whole genome sequence data available for many individuals of the same populations. These sequences can be used to identify potential phenotype-altering mutations in coding regions and predict their effect [2, 3]. **We expect to be able to identify some of the underlying causative mutations**, aided by the availability of phenotype data and validation by designing specific crosses.

Table 1: Heterozygous carrier matings. Both parents carry a lethal allele.

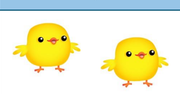
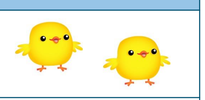
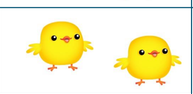
Allele	A 	B
A 		
B		

Table 2: Typical example output, number of observed (O-Hom) homozygotes is significantly lower than the expected (E-Hom) homozygotes.

Locus	Haplotype	Fq %	O-Hom	E-Hom	P-val
X	0000010010	4.1	0	20	0.0001
Y	0011101000	2.1	0	16	0.0002
Z	1110001000	3.2	0	12	0.0003

Data & Methods

11,896 chickens on 60K SNP chip

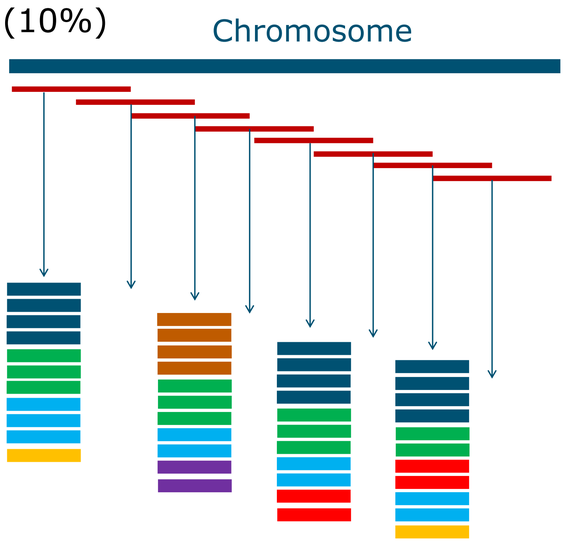
Pre-processing (PLINK [4])

- Missing genotypes per individual(10%)
- Minor allele frequency (MAF) < 2%
- Missing genotypes per SNP (10%)
- Mendelian errors

Phasing

- Beagle [5]
- Shapeit2 [6]

Sliding window approach:



Expected homozygotes is calculated based on the pedigree:

$$E(k) = \sum_{i=1}^{ns} p_{ik} \sum_{j=1}^{nd} r_{jk} n_{ij}$$

Significance: Exact binomial test (P < 10-4)

Whole genome resequensing data: 270 individuals (sires), mapping: BWA-MEM v0.7.5, SNPcalling: GATK

Score deleteriousness: SIFT [2], PROVEAN [3]

Results

Significant haplotypes

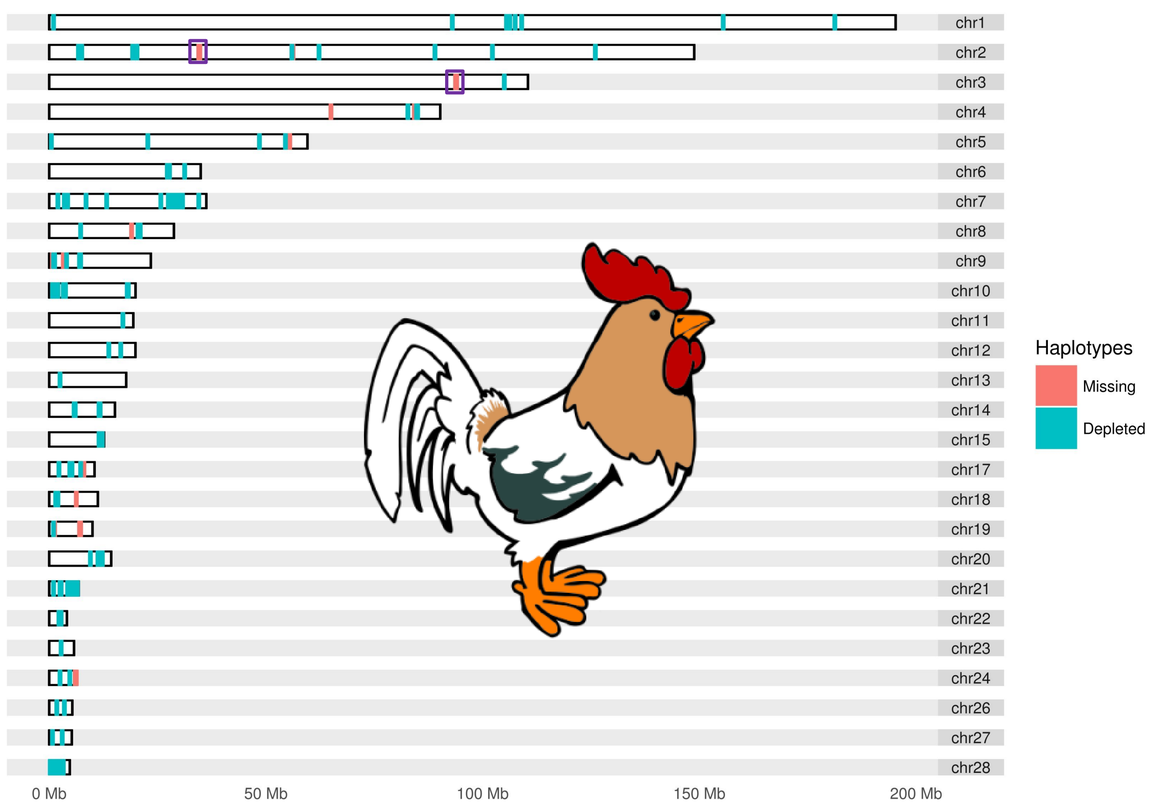
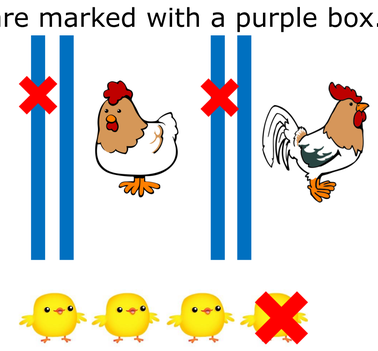


Figure 2: Genomic locations of significant haplotypes identified. Missing homozygotes are marked in pink, significantly depleted haplotypes in blue. The two most highly significant missing haplotypes (E-Hom > 20) are marked with a purple box.

Underlying variants

- We constructed a catalogue of variants in the identified regions (±500Kb on either side) and listed all loss-of-function (LoF) and predicted deleterious variants to identify possible causative variants.



Conclusions

- Scanning for depletion of haplotypes provides a powerful tool to identify deleterious recessive alleles
- Based on our pipeline we **identified a number of haplotypes carrying a potential deleterious or lethal allele** in several chicken populations.
- These results can help to avoid **specific matings** producing affected or non-viable progeny in breeding programs.
- Further development of the method will focus on the **identification of causative variants** using the whole genome re-sequencing data.

References

1. VanRaden, P. M., et al., *Harmful recessive effects on fertility detected by absence of homozygous haplotypes.* J Dairy Sci, 2011. **94**(12): p. 6153-61.
2. Kumar, P., S. Henikoff, and P. C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.
3. Choi, Y., et al., *Predicting the functional effect of amino acid substitutions and indels.* PLoS One, 2012. **7**(10): p. e46688.
4. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
5. Browning, S. R. and B. L. Browning, *Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.* Am J Hum Genet, 2007. **81**(5): p. 1084-97.
6. Delaneau, O., J. Marchini, and J. F. Zagury, *A linear complexity phasing method for thousands of genomes.* Nat Methods, 2012. **9**(2): p. 179-81.

Animal Breeding and Genomics Centre
P.O. Box 123, 6708 PB Wageningen
Contact: martijn.derks@wur.nl
T + 31 (0)317 48 30 19, M +31 (0)6 20 26 09 33