Footprints of Parallel Selection Revealed by Direct Sequencing in Egg Laying Chicken

S.Qanbari[1], M. Seidel[2], TM. Strom[2], R. Preisinger[3], H. Simianer[1]
[1]Georg-August University, Göttingen, [2]Helmholtz Zentrum München, Neuherberg,
[3]Lohmann Tierzucht GmbH, Cuxhaven, Germany

ABSTRACT: We used whole-genome re-sequencing data from two commercial white (WL) and brown (BL) egg laying chicken lines to perform first systematic screening of recent artificial selection in the genome of modern chicken. Evidence of positive selection was investigated in two steps. First, we constructed a map of differentiated loci between WL and BL. Next, we examined evidence of parallel selection for production-related traits in the two populations. Our analyses revealed signatures of parallel selection in separate lines, exemplified with a striking evidence of selection for candidate gene OPG, a gene implicated in bone mineral density, osteoporosis and fracture risk. Our study demonstrates the utility of population based techniques for detecting recent selection in chicken. These results also show that LD decays in commercial chicken at a much faster rate than previously thought.
Keywords: F_{ST}; LD; Selective sweep

Introduction

During the last century, modern selective breeding has made spectacular progress in commercial chicken breeding. During the specialization of egg or meat producing lines, chicken have been subject to intensive selection to increase yield, fertility and other complex traits. Selection affecting these phenotypes has left detectable signatures in the genome of modern chicken. The elucidation of these signatures of selection is of interest to identify production and fitness related genes that help to genetically improve this economically important species. Selection leaves distinct footprints behind such as reduced local variability (Rubin et al. 2010) and increased linkage disequilibrium and extended haplotype structure (Voight et al. 2006) that can be detected in either within or between population comparisons. The methods used to detect selection are based either on the site frequency characteristics (focusing on single loci) or on properties of haplotypes segregating in populations.

Previous genome-wide studies to detect positive selection in commercial chicken have used either SNP arrays (Johansson et al. 2010; Elferink et al. 2012) or pooled sequence data (e.g., Rubin et al. 2010; Qanbari et al. 2011). Analyses based on SNP array data suffer from ascertainment bias and limited resolution. In contrast, pooled sequencing guarantees a high resolution, but produces uncertain allelic frequency profiles and remains blind to the individual genotypes and local haplotype structures. Many selection signals may, therefore, have remained un-detected in previous studies. Using whole genome sequence information can overcome these problems.

In this study, we use whole-genome re-sequencing data from two commercial white (WL) and brown (BL) egg laying chicken lines of the Lohmann Tierzucht GmbH. We employ multiple statistics to scan the genome for regions that underwent recent selection. We identify highly differentiated regions/genes between white and brown layers, suggesting that selection is responsible for the differences in allele frequencies. We also investigate evidences of parallel selection for production-related traits in separate populations. We do this by exploring none (or less)-differentiated regions for overlapping signatures of positive selection. We also use the sequence data to re-examine the pattern of Linkage Disequilibrium (LD).

Material and Methods

Genetic material. For the purpose of this study we used low to medium (~7 fold) coverage sequence of 25 female birds of two commercial white (WL) and brown (BL) egg laying populations of the Lohmann Tierzucht GmbH. After initially editing data for calling depth, clustered SNPs, missing calls and duplicated reads the final sequence panel involved > 9.3 mio. SNPs with an average inter-marker space = 107 bp in both populations. We used Beagle (Browning and Browning, 2011) to reconstruct haplotypes and impute missing genotypes.

Linkage disequilibrium (LD). We quantified LD using r^2 between pairs of SNPs. Evaluations of SNP to SNP pairwise r^2 were performed based on the panels of 15,000 SNPs randomly sampled across identical segments of each 5Mb on two populations. We sampled two subsets of SNPs across macro-chromosomes (chr 1-5) and one subset on chromosomes 6 to 10 with intermediate size. All SNPs on remaining micro-chromosomes were evaluated for estimating LD (classification of chromosomes is according to the ICGSC, 2004).

Detecting positive selection. Evidence of positive selection was investigated in two steps and through multiple statistics. First, we constructed a map of diverged loci between brown and white egg layers using a genetic differentiation statistic (F_{ST}; Reynolds et al. 1983) assuming that recent selection was responsible for extreme local differences in allele frequencies. In the second step, we examined evidence of positive selection in none-differentiated regions. To this purpose we employed the integrated Haplotype Homozygosity Score (iHS) that explores the structure of haplotype (Voight et al. 2006). Evidence of positive selection was also investigated by assessing variation in allele frequency across the genome based on Heterozygosity (Het), which should be reduced in regions affected by selection. To reduce locus-to-locus variation in the inference of selection we averaged single

SNP values for non-overlapping windows of 40 kb across the genome. Window size was adapted based on the extent of LD (Figure 1). The metrics were also estimated using different window size to explore the sensitivity to the choice of window. Empirical p-values were generated by genome wide ranking of F_{ST}, |iHS| and Het values. We annotated candidate genomic regions by aligning the positions to the chicken genome sequence assembly (build 4), to reveal genes located in the respective region.

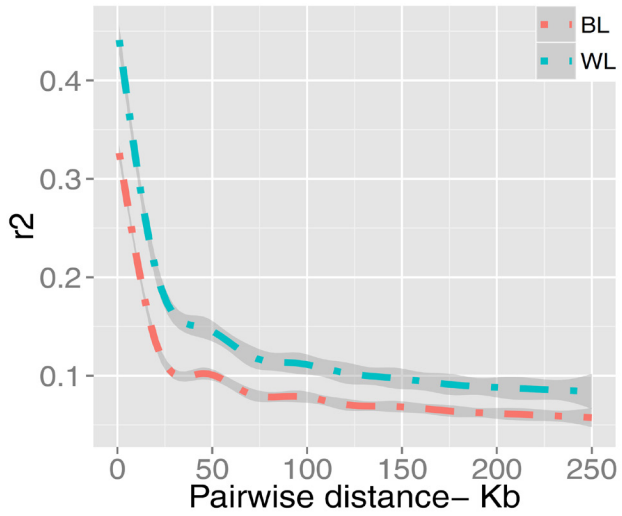

Figure 1. Decay of LD as a function of distance.

Results and Discussion

Extent of LD. A detailed profile of LD over the entire genome is a quantity of interest, especially for the use in breeding programs implementing genomic selection. Previous studies of LD structure in commercial chicken populations have used low resolution panels of ascertained SNPs mainly selected based on their minor allele frequency (MAF) and position on the genome (e.g., Qanbari et al. 2010). The sequence data analyzed in this study includes roughly 200 times more SNPs than the 50K array previously used in chicken for examining LD. The increased SNP density provides a greater coverage of rare and low frequency SNPs than in any previous study based on SNP chip data.

Consistent with the demographic history, WL exhibited a slower decay of LD than BL over physical distance (Figure 1). It is evident that average LD does not extend beyond the inter-marker space of 100Kb across the genome. Previous studies in commercial chicken however, found strong LD extending over several Megabasepairs (e.g., Qanbari et al. 2010). However, LD as measured by r^2 depends on allele frequencies and the difference between this study and previous studies may partially be explained by the biased SNPs selection on the genotyping arrays, where SNPs mainly were ascertained aiming at a uniform allele frequency distribution and coverage of the genome. Additionally, differences in the sample composition may explain the results, as LD is strongly affected by population structure. Characterizing LD using structured populations leads to an inflation of the LD statistics, which might have affected previous studies. Finally, genotyping error reduces apparent LD, and is a major concern for low- and intermediate-depth coverage re-sequencing data. Our analyses also revealed a substantial difference between chromosomes which can be attributed to the much higher recombination rates on short chromosomes.

Exploring putatively selected loci. We calculated F_{ST} values for each SNP and examined its distribution at the level of the genome, the chromosome, and individual genes. The empirical genome-wide distribution of F_{ST} indicates that recent selection has severely operated on the genome (Genome-wide F_{ST} = 0.39±0.33). The average F_{ST} for autosomal and Z-linked SNPs was significantly different (0.32 and 0.40, respectively; t test $p < 2.2e\text{-}16$) (see also Figure 2A). A higher average F_{ST} for Z -chromosome SNPs is expected because of its smaller effective population size compared with that of the autosomes, which makes it more sensitive to demographic events and/or natural selection. Sliding single site values of F_{ST} resulted in a total of 24,797 windows across the genome. Evidence of the positive selection was then assumed for windows in the extreme 1% of the empirical distribution which resulted in 36 significant regions ($F_{ST} > 0.72$). These results based on sequencing entire genome provide a detailed map of differentiated loci, some overlapped with genes previously suggested being under selection (data not shown).

Signatures of selection can be recognized when adjacent SNPs all show high F_{ST}, due to the hitch-hiking effect, implying divergent selection between populations, or where adjacent SNPs all show low F_{ST}, implying balancing selection. A cluster of low-profile- F_{ST} SNPs, however, may also reflect a parallel fixation of beneficial allele in two separate populations which have different phylogenetic history, but have been selected for similar breeding goals (e.g. white-egg layers and brown-egg layers). This scenario suggests that true signals generated by selection would overlap across the populations. To address this hypothesis in more detail, we further explored the pattern of DNA in the region of low-profile- F_{ST} windows (F_{ST} <0.05, see Figure 2A). We examined the local variability (Het), and extent of haplotypes segregating in the region. Strong evidence of a selective sweep reflected by a set of windows extending over 1Mb on chromosome 2 was observed in the region harboring the OPG (TNFRSF11B) gene (Figure 2B,C,D). Osteoprotegerin, the product of this gene, is a key negative regulator of osteoclastogenesis and is secreted by osteoblasts/stromal cells. In humans, polymorphisms within the OPG gene have been widely studied and associated with bone mineral density, osteoporosis and fracture risk (e.g., Richards et al. 2008, among others). Osteoporosis is a progressive loss in structural bone and is a common problem in caged egg-laying strains of hens (Whitehead and Fleming, 2000). Welfare issues associated with osteoporosis have become more urgent due to the increasing use of battery cages. In addition to animal welfare concerns, osteoporosis causes major economic loss in the egg-laying industry (Schreiweis et al. 2005).

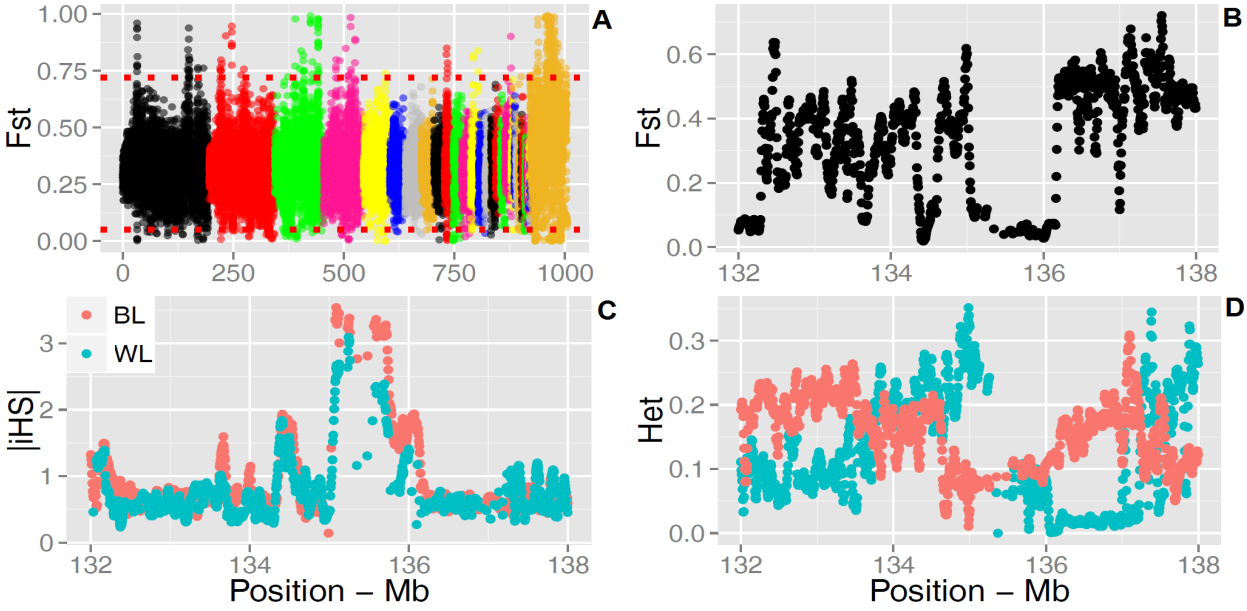

Figure 2. Panel A visualizes genome-wide distribution of F_{ST} measures. F_{ST} for 24,797 non overlapping windows of 40Kb are plotted across chicken genome. The red horizontal lines provide a guide for identifying exceptionally high and low F_{ST} values (corresponding to the upper and lower 1% of the empirical distribution of F_{ST}). Panels B, C and D are high resolution illustration of F_{ST}, haplotype homozygosity and heterozygosity, respectively, on a candidate region of chromosome 5 harboring OPG gene. B, C and D are plotted as overlapping 40Kb windows in steps of 5Kb.

Conclusion

We present the first comprehensive study for localizing signatures of recent selection in commercial egg laying chicken based on full re-sequencing data. We report a map of differentiated loci along with signatures of parallel selection for production related traits in brown versus white egg laying lines, exemplified by a striking evidence of selection at OPG gene. The observation of multiple signals in commercial layer lines is consistent with the hypothesis that egg production is a complex trait controlled by many genes. Our study demonstrates the utility of population based techniques for detecting recent selection in chicken. These results also show that LD decays in commercial chicken at a much faster rate than previously thought.

Acknowledgments

This research was funded by the German Federal Ministry of Education and Research (BMBF) within the AgroClustEr "Synbreed - Synergistic plant and animal breeding" (FKZ 0315528).

Literature Cited

Browning, B.L., and Browning, S.R., 2011. Am J Hum Genet 88:173-182.

Elferink, M.G., Megens, H.-J., et al. 2012. PLoS ONE 7, e32720.

Hillier, L.W., Miller, W., Birney, E., Warren, et al 2004. Nature 432, 695–716.

Johansson, A.M., Pettersson, M.E., Siegel, P.B., Carlborg, Ö., 2010. PLoS Genet 6, e1001188.

Qanbari, S., Hansen, M., Weigend, S., et al. 2010. BMC Genetics 11, 103.

Qanbari, S., Strom, T.M., Haberer, G., et al. 2012. PLoS ONE 7, e49525.

Reynolds, J., Weir, B.S., Cockerham, C.C., 1983. Genetics 105, 767–779.

Richards, J., Rivadeneira, F., Inouye, et al. 2008. Lancet 371, 1505–1512.

Rubin, C.-J., Zody, M.C., Eriksson, J., Meadows, J.R.S., et al. 2010. Nature 464, 587–591.

Schreiweis, M.A., Hester, P.Y., Moody, D.E., 2005. Genet. Sel. Evol. 37, 677–698.

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. PLoS Biol 4, e72.

Whitehead, C.C., Fleming, R.H. 2000. Poult. Sci. 79, 1033–1041.