SNP Discovery and Genomic Architecture of Highly Inbred Leghorn and Fayoumi Chicken Breeds Using Whole Genome Resequencing

A.S. Leaflet R2996

Damarius S. Fleming, Graduate Research Assistant;
James E. Koltes, Post-Doctoral Research Associate;
Eric R. Fritz-Waters, Research Associate;
James M. Reecy, Professor;
Susan J Lamont, Distinguished Professor

Summary and Implications

Advances in the use of next generation sequencing (NGS) and ability to pool individuals into groups that represent distinct livestock populations has made it possible to examine trait differences between breeds of chicken. The breeds examined are very divergent when compared on their history of laying ability and immune response. The long-term objective is to understand the genetic differences between the Leghorn and Fayoumi breeds for use in developing more productive and disease resistant chickens. Statistical testing of the sequence of the two breeds along with Gene set enrichment analysis (GSEA) to make connections between the genetic variation seen in the NGS data and the breed specific traits of egg laying and heightened immune response can be used to characterize these two breeds. Genetic terms having the highest level of differentiation between the lines appear to group into metabolic processes, with terms over-enriched for immune system process, sexual reproduction, and growth for variants examined between lines. Terms for functions within the Fayoumi and Leghorn populations aligned to immune function and reproductive function, respectively. These results are consistent with known breed phenotypes and provide a means to focus on specific DNA variations and the birds' genetic diversity that are potentially of more commercial importance.

Introduction

The onset of new and more powerful next generation sequencing technologies has given researchers the ability to generate massive molecular data sets from individual samples. Discovery and analysis of variants within and between groups of individuals helps define the genetic differences underlying phenotypes or morphology. In this study, we examine genetic variation within and between two highly distinct inbred lines of Leghorn and Fayoumi chickens. Analyzing sequence variants of these very diverse and highly inbred chicken lines allows for examination of the genomic factors underlying trait differences between breeds. The objective of our study was to discover variants to understand how genetic changes affect line-specific differences of highly inbred Leghorn and Fayoumi chicken breeds. The analysis focused on fixed, line-specific genes that were different from the Red Jungle Fowl (RJF) reference sequence (Galgal4).

Materials and Methods

Animals. Chickens from the Iowa State University Fayoumi and Leghorn experimental chicken lines were used for this study. These birds were extremely inbred, with over 70 generations of sib mating. Sixteen birds per line were used for the pooled resequencing of DNA. These Fayoumi and Leghorn lines broadly represent a divergent history of natural selection for disease resistance and artificial selection for reproduction. Variant calling was done using GATK and annotation using SNPeff.

Results and Discussion

The genomes of lines studied were over 99% homozygous reflecting the level of inbreeding within each line. The Fayoumi and Leghorn pooled sequence data were compared against the RJF reference to call all possible variants present in the populations. Analysis of the Fayoumi vs. RJF identified 4,462,467 total variants after quality control filtering. For the Fayoumi, 72.2% of variants found were novel. Similar results were obtained in the analysis of the Leghorn vs. RJF. In Leghorn 4,605,732 variants were discovered of which 71.3% were novel. We uncovered a total of 2,052,537 variants that were unique to Fayoumi and 2,196,553 unique to Leghorn. To examine differences and similarities between the inbred populations of Fayoumi and Leghorn chickens, a fixation analysis (F_{ST}) was conducted. This analysis measures the amount of difference between populations and helped to determine the genes analyzed by GSEA. Analysis of the genes fixed within line indicated that the two lines mainly differed in the variants that each breed used to drive various processes. Annotations for immune system processes, response to stimulus, and metabolic processes (adj p value < 0.05) were overrepresented terms that emerged from the gene set. By comparing data to categories of characteristic traits for each population (Fayoumi= immune traits, Leghorn = reproductive traits), the study highlighted functions indicative of these breed-predominant traits. The Fayoumi had a higher percentage of heterozygous variants than Leghorn within line. Major genetic differences found between breeds by F_{ST} and subsequent GSEA analysis was indicative of functions that showed the highest level of differentiation between the lines. Over-enriched terms included processes such as

immune system, sexual reproduction, and growth. Taken as a whole, the data indicated that there is evidence that the two lines have diverged at both the genetic level and their measureable traits.

Acknowledgements

We thank the NRSP-8 Grant for funding the sequencing of the inbred lines and the USDA National Needs Grant for funding of the research assistant, Damarius S. Fleming.

Table 1. Overrepresented gene ontology (GO) terms for moderate impact, line-specific, fixed variants in inbred Fayoumi and Leghorn lines.

Fayoumi GO Term	Count	P-Value	Benjamini
FN3	22	5.60E-06	1.10E-03
Ribonucleotide binding	154	2.00E-05	1.70E-03
Purine ribonucleotide binding	154	2.00E-05	1.70E-03
Fibronectin, type III	22	8.50E-06	8.10E-03
PTPc	8	1.80E-04	1.70E-02
Nucleotide binding	177	2.70E-04	2.10E-02
Protein kinase activity	61	3.90E-04	2.60E-02

Leghorn GO Term	Count	P-Value	Benjamini
ECM-receptor interaction	25	7.00E-07	9.60E-05
Extracellular matrix	41	5.60E-07	1.90E-04
Metal ion binding	270	9.50E-06	2.30E-03
Proteinaceous extracellular matrix	37	7.60E-06	1.30E-03
Extracellular region	94	2.00E-05	2.30E-03
Extracellular region part	59	2.80E-05	2.50E-03
Extracellular matrix part	15	3.40E-04	2.30E-02
Cell division and chromosome partitioning	26	7.90E-04	2.60E-02
Calcium ion binding	77	5.40E-04	3.90E-02
DNA damage	13	2.80E-04	4.30E-02
Aminophospholipid transporter activity	7	5.40E-04	4.30E-02
Phospholipid-translocating atpase activity	7	5.40E-04	4.30E-02

P-value based on Fisher's exact test and Benjamini is a multiple testing correction of the P-value calculation.