Development of a Low-Density SNP Panel for Local Ghanaian and Tanzanian Chicken Ecotypes

M. Walugembe[1], E.N. Amuzu-Aweh[2*], K.-S. Lim[1], Y. Wang[3], N. Chouicha[3], T. Kelly[4], A.P. Muhairwa[5], B.B. Kayang[2], P.L. M. Msoffe[5], A. Naazie[2], H.R. Otsyina[6], J.S. Mushi[5], R.A. Gallardo[4], H. Zhou[3], S.J. Lamont[1], and J.C.M. Dekkers[1].

[1]Department of Animal Science, Iowa State University, Ames, IA, United States, [2*]Department of Animal Science, University of Ghana, Legon, Accra, Ghana; [3]Animal Science, University of California, Davis, Davis, CA, United States, [4]School of Veterinary Medicine, University of California, Davis, Davis, CA, United States, [5]Department of Veterinary Medicine and Public Health, Sokoine University, P.O. Box 3000 Chuo Kikuu, Morogoro, Tanzania, [6]School of Veterinary Medicine, University of Ghana, Legon, Accra, Ghana; ***Presenting author** jdekkers@iastate.edu

Abstract

Genetic selection is important in choosing animals for the next generation. Animal breeders now choose selection candidates based on DNA markers, in addition to animal performance. There has been increased use of high-density SNP chips to enable selection of commercial broiler and layer chickens. However, genotyping a large number of selection candidates still carries a high cost. The aim of our study was to develop an economical 5K SNP panel for local Ghanian and Tanzanian chicken ecotypes using targeted genotyping by sequencing (GBS) for imputation to higher density. Low-density panel SNPs were selected from haplotype blocks and other important genomic regions across the chicken genome based on 600K SNP panel genotypes. A total of 188 birds were genotyped by GBS and an in-house shell script pipeline was utilized to obtain SNP calls. Our in-house pipeline was compared to a standard company pipeline and 600K SNP chip genotypes for validation. Selected SNPs were evenly distributed across the genome, with at least one SNP in each megabase region. Comparison of the two pipelines revealed a good genotype match for the 5K SNP panel. The 5K GBS panel and SNP calling pipeline are important tools to aid selective breeding in African chicken ecotypes.

Introduction

In the past decade, there has been an increase in the use of commercial high-density (HD) SNP chips for genotyping in chickens (Kranis et al., 2013), which has accelerated genomic selection in broiler and layer breeding. The availability of genomic and phenotypic information enables genomic prediction of breeding values for selection candidates. Although HD SNP chips are commonly used, the costs to genotype a large number of selection candidates remains high. To reduce costs, the use of low-density (LD) panels followed by imputation to a high density could be a feasible approach (Habier et al., 2009). Array-based genotyping platforms are commonly utilized in genotyping various livestock species. However, they include SNPs that may not be geographically representative and therefore, population diversity and estimates of recombination rates may be biased (Pérez-Enciso and Ferretti, 2010; Zhan et al., 2011). In poultry, such studies have mostly been conducted in commercial layers and broilers, with limited focus on local African chicken ecotypes. New cost-effective next generation sequencing platforms such as genotyping by sequencing (GBS) have recently been developed. GBS was originally developed for plants (He et al., 2014), and follows a simple protocol that provides high SNP coverage (Sonah et al., 2013). The objective of this study was to develop a low-density SNP panel for Ghanaian and Tanzanian local chickens using targeted GBS.

Materials & Methods

Experimental birds. Breeding flocks were established at the University of Ghana, Legon, Accra, Ghana, and Sokoine University of Agriculture, Morogoro, Tanzania. Breeder chickens from three Tanzanian (Kuchi, Morogoro Medium, and Ching'wekwe) ecotypes and three Ghanaian (Coastal Savannah, Forest and Interior Savannah) ecotypes were obtained.

SNP panel genotyping and quality control. Blood samples were collected from 1,440 Ghanaian and 1,399 Tanzanian chickens and stored onWhatman FTA cards (Sigma-Aldrich, St. Louis, MO, United States). All birds were genotyped using the Affymetrix Axiom® 600k Array (Thermo Fisher Scientific Inc., Calsbad, CA, USA) at GeneSeek (Lincoln, NE, USA) and *Gallus gallus* genome version 5 was used for annotation of the genotyping array. Quality control on the combined genotype data was performed using PLINK 1.9 (Chang et al., 2015) and SNPs were screened based on minor allele frequency (MAF) ≥ 0.1 and call rate > 95%. A total of 401,083 SNPs were utilized in the downstream analyses.

Selection of SNPs for the 5K GBS panel. SNPs for the 5K GBS panel were obtained across the chicken genome from various sources including haplotype block estimates based on the 2,839 birds, SNPs in the MHC regions provided by Hy-Line International., genome-wide association studies on the commercial Hy-Line Brown laying line (Rowland et al., 2018; Saelao et al., 2019) and in African local chicken ecotypes (Walugembe et al., 2019; Walugembe et al., 2020), and 379 SNPs selected across the genome to ensure that there was at least one SNP in each Mb region. Imputation for the combined genotype data was performed using Fimpute (Sargolzaei et al., 2014) to fill in missing genotypes. Haplotype blocks were estimated in PLINK 1.9 (Chang et al., 2015) using the following options; --blocks, --blocks-min-maf 0.1, and --blocks-strong-lowci 0.6. A total of 69,622 haplotype blocks were generated across the genome and 4,500 SNPs were selected based on these haplotype blocks. The number of SNPs selected for a chromosome was computed as x*4500/401803, where x is the number of 600K SNPs on the chromosome that passed quality control.

Sequencing and SNP calling. Blood samples were also collected from an additional 92 Tanzanian and 96 Ghanaian chickens, which were sequenced for the 5K low-density GBS panel with 100 bp paired-end on 4 lanes using Illumina Hiseq at GeneSeek. Of these, 25 birds were also genotyped on the Affymetrix Axiom® 600k Array. Sequence data were processed using an in-house pipeline by Interval Bio (Mountain View, CA, USA). With input from Interval Bio staff, we developed our own pipeline to process the raw sequence read data and target variant calling using shell scripts and various publicly available software tools (BWA, SAMtools, PICARDS, and BCFtools). The raw reads were aligned to the *Gallus gallus* 5 reference genome, and genotypes were obtained from the vcf file using in house python script. SNP genotype calls from our pipeline were compared to those of the Interval Bio pipeline for all 188 birds and to the 600K SNP genotypes for the 25 birds.

Results

Selection of SNPs for the 5K GBS panel. Initial selection of SNPs utilized LD based pruning methods, where SNPs in high LD were retained after pruning. Although this resulted in selection of SNPs that were distributed across the genome, most had MAF close to 0.5. This approach method was, therefore, replaced by the haplotype blocks method. The 4,500 SNPs selected based on haplotype blocks were evenly distributed across the genome, with MAF ranging from 0.1 to 0.5, as illustrated in Figure 1. Each Mb included at least one SNP and for regions where this was not achieved, a SNP was selected and added to the panel. Haplotype

blocks were determined to contain SNPs that are in very strong LD with each other to minimize intra-block variance and maximize inter-block variance.

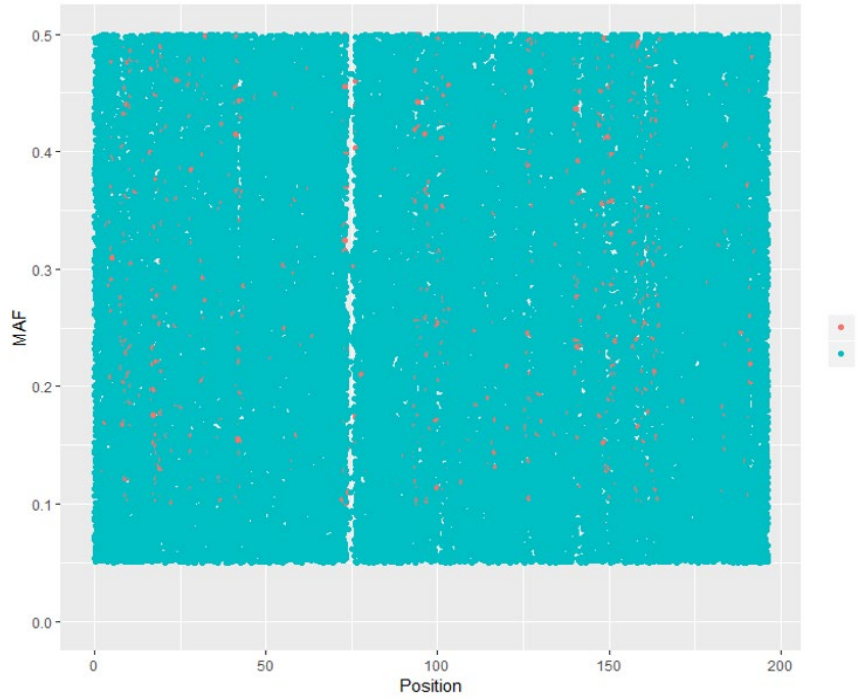

Figure 1. Distribution of the 805 selected SNPs (red) from the 71,725 SNPs that passed quality control on chromosome 1.

Comparison of SNP genotype calls. Figure 2 shows the comparison of genotype calls based on our and the Interval Bio pipeline. SNPs were ranked by the percentage of the 188 birds for which the genotype was called the same by both pipelines (Match%). For most of the 188 birds, the same genotype (Match%) was called by both pipelines, with a few cases where neither pipeline or only one of the pipelines made a genotype call. We removed about 80 SNPs where a different (Non-match%) genotype was called by the two pipelines for more than 20% of the birds.

Figure 2. SNPs ranked by the % of birds for which the same genotype was called by both pipelines (light blue) along with the % for which different genotypes were called (red), the % for which only our pipeline did not (purple) or did (yellow) make a genotype call, and the % for which no call was made by both pipelines (dark blue).

Both pipelines had matching genotype calls with the 650K genotypes for the 25 birds that were genotyped on both the GBS and the 650K panel. These findings confirmed our low-density SNP panel list that would subsequently be used for genotyping on a GBS platform/protocol.

Discussion

To our knowledge, this is the first low-density SNP panel developed for African local chicken ecotypes, which can enable selective breeding of Africa local chickens using a cost-effective GBS. The panel has been used to genotype animals on a low density panel and imputation to a high density panel will be done using available imputation software programs.

The targeted GBS methodology employed here is one of the current methods of next generation sequencing for genotyping of SNPs and the availability of an excellent reference genome provides the basis for the selection of GBS restriction enzymes (de Donato et al., 2013; Gurgul et al., 2019).

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) Feed the Future Innovation Lab for Genomics to Improve Poultry (cooperative agreement number AID-OAA-A-13-00080).

References

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell et al.(2015). *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

de Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. (2013). *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0062137

Gurgul, A., Miksza-Cybulska, A., Szmatoła, T., Jasielczuk, I., Piestrzyńska-Kajtoch, et al. (2019). *Genomics*, *111*(2), 186–195. https://doi.org/10.1016/j.ygeno.2018.02.002

Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2009). Genomic selection using low-density marker panels. Genetics, 182(1), 343–353. https://doi.org/10.1534/genetics.108.100289

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, et al. (2014). In *Frontiers in Plant Science* (Vol. 5, Issue SEP). Frontiers Media S.A. https://doi.org/10.3389/fpls.2014.00484

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, et al. (2013). *BMC Genomics*, *14*(1). https://doi.org/10.1186/1471-2164-14-59

Pérez-Enciso, M., & Ferretti, L. (2010). *Animal Genetics*, *41*(6), 561–569. https://doi.org/10.1111/j.1365-2052.2010.02057.x

Rowland, K., Wolc, A., Gallardo, R. A., Kelly, T., Zhou, et al. (2018). *Frontiers in Genetics*, *9*(August), 1–12. https://doi.org/10.3389/fgene.2018.00326

Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). *BMC Genomics*, *15*(1). https://doi.org/10.1186/1471-2164-15-478

Saelao, P., Kelly, T., Lamont, S., Gallardo, R., Zhou, et al. (2019). Genes, 10(1), 61. https://doi.org/10.3390/genes10010061

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, et al. (2013). *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0054603

Walugembe, M., Amuzu-Aweh, E. N., Botchway, P. K., Naazie, A., Aning, et al. (2020). *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00739

Walugembe, M., Mushi, J. R., Amuzu-Aweh, E. N., Chiwanga, G. H., Msoffe, et al. (2019). *Genes*, *10*(7). https://doi.org/10.3390/genes10070546

Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, et al. (2011). *BMC Genomics*, *12*. https://doi.org/10.1186/1471-2164-12-557