Revisiting genome-wide suggestive regions for daily milk production in Egyptian buffalo using a multi-locus approach

H. Abdel-Shafy¹, N.K. El-Halawany², A.A. Shawky² & A.F.M. Al-Tohamy²

¹ Department of Animal Production, Faculty of Agriculture, Cairo University, 12613, Giza, Egypt
² Department of Cell Biology, National Research Centre, Dokki 12622, Giza, Egypt
nk.el-halawany@nrc.sci.eg (Corresponding Author)

Summary

Our initial GWAS for daily milk production in Egyptian buffalo using SNPs has detected several loci with suggestive signals on chromosomes 1, 5, 6 and 27. In the current study, we investigate the use of haplotypes to confirm associations of suggestive genomic regions. A total number of 250 animals were used to estimate average milk yield deviations for the first five lactations. AffyPipe and PLINK were used to check the quality of genotypes. Haplotypes were constructed and partitioned into haplotype blocks using fastPHASE and Haplovview, respectively. Association analysis for each suggestive region was performed using PLINK. Our results confirmed the previous findings of genome-wide suggestive signals on the four chromosomes after Bonferroni correction. The most significant haplotype (P=3.9 x 10^-6) was identified on chromosome 5 between 72.54 and 72.57 Mb, while the most significant SNP from our initial investigation was located on chromosome 27 (P=1.87 x 10^-6) at 30.14 Mb. The identified regions were subjected to enrichment analysis using the GeneCodis databases. The analysis ranked immune response at the top of all biological processes (P<0.001). In conclusion, multi-markers based approach can be used beside single markers to identify genomic associations for small population especially for phenotypes with moderate variation. Further investigations of the validated loci in different populations are necessary to identify the potential candidate genes and probably causative mutations.

Keywords: GWAS, haplotypes, candidate genes, buffalo

Introduction

The availability of affordable high-throughput genotyping technologies make it feasible to identify genomic loci associated with the trait variation with high precision through genome-wide association studies (GWAS). Recently, we have performed GWAS using Axiom Buffalo Genotyping Array to detect genomic loci associated with daily milk yield deviation (YD) in Egyptian buffalo (El-Halawany et al., 2017). Although several genomic regions on chromosomes 1, 5, 6 and 27 have been suggested, the power to detect significant association was limited after correction for multiple testing. The same issue have been raised in the GWAS for total milk yield in Italian and Brazilian buffalo (Iamartino et al., 2013), which may be due to the weak or moderate effect size of identified SNPs.

Haplotype based approach may have higher power to detect associations driven from cis-interactions among nearby loci in the suggestive genomic regions. Because haplotype comprise several SNPs with high linkage disequilibrium (LD) among each other, it can have larger aggregated effect than SNPs. Thus, haplotype based approach may increase the power to identify genomic loci, even if they have small effects (Bickel et al., 2011, Abdel-Shafy et
al., 2014). Therefore, the objective of this study is to investigate the use of haplotypes to test significant effect of suggestive genomic regions identified form our initial GWAS in Egyptian buffalo.

Material and methods

Animals and phenotypes

In the current study, we used the same phenotypes as in our initial GWAS based on SNP markers (El-Halawany et al., 2017). Briefly, YDs for the first five lactations of 250 buffalo were estimated based on mixed model procedure (SAS 2014, SAS Institute Inc., Cary, NC, USA) using the original yield records on a daily basis, 5 to 290 days-in-milk (DIM). A yield deviation (YD) is a weighted average of the animal's milk yields adjusted for non-genetic factors (VanRaden and Wiggans, 1991).

Genotypes and quality check

Only 96 buffalo were genotyped with Axiom Buffalo Genotyping Array (90,000 SNPs). Stringent quality control was applied to remove markers and individuals with insufficient genotyping rate using AffyPipe (Nicolazzi et al., 2014) and PLINK (Purcell et al., 2007). Only SNPs with unknown position, low minor allele frequency (MAF<0.01), or missing genotype rate > 10% were excluded. In addition, markers significantly (P<0.001) deviated from Hardy-Weinberg proportion were eliminated. Individuals with call rate <90% were removed. The remaining data set included 70,182 SNPs and 95 animals with 99.8% genotyping rate. Haplotypes for each chromosome were constructed using fastPHASE (Scheet and Stephens, 2006). Phased genotypes were partitioned into haplotype blocks using the solid spine algorithm implemented in Haploview (Barrett et al., 2005). The exclusion criteria applied for SNPs were also applied for haplotypes. In total, 18,290 haplotype blocks containing 59,394 SNPs were constructed for all chromosomes. The average and standard error of D’ (a common measures of LD, (Lewontin, 1964)) among SNPs in the haplotype blocks were 0.89±0.07, respectively. The average size of haplotype blocks was 80.4 kb. The frequencies were ranged from 0.01 to 0.99.

Association analyses

Haplotypes within a window of 1 Mb interval around previously suggestive regions were tested for association with YDs, since LD remains at distances up to 1 Mb in buffalo (El-Halawany et al., 2017) and cattle (Gibbs et al., 2009). The regions were located on chromosomes 1 (from 19.8 to 22.5), 5 (from 71.5 to 109.5), 6 (from 78.0 to 92.7) and 27 (from 27.4 to 41.6 Mb). The association analyses for each region were performed using the linear regression model in PLINK (Purcell et al., 2007) with population stratification as covariates. Haplotypes were considered significant at the threshold of α ≤ 0.05 after Bonferroni correction if the nominal P-value×N ≤ 0.05, where N is the number of tested haplotypes in each region separately. The number of haplotypes were 96, 1134, 403 and 495 on the regions located on chromosomes 1, 5, 6 and 27, respectively.

The gene search was performed using Ensembl data base (UMD3.1, build 85). Haplotypes were assigned to genes using Ensembl Perl API tools (http://www.ensembl.org) through a Perl script (http://www.perl.org). Integrative information of annotations for a given
gene list was performed by GeneCodis (http://genecodis.cnb.csic.es) using information from different bioinformatics tools and functional enrichment analyses (Carmona-Saez et al., 2007, Nogales-Cadenas et al., 2009, Tabas-Madrid et al., 2012).

Results and Discussion

Our results confirmed the previous findings of genome-wide suggestive signals on the four chromosomes after Bonferroni correction. The most significant haplotype ($P=3.9 \times 10^{-65}$) was located on chromosome 5 between 72.54 and 72.57 Mb (Table 1), while the most significant SNP from our initial investigation was located on chromosome 27 ($P=1.87 \times 10^{-6}$) at 30.14 Mb (El-Halawany et al., 2017). The frequencies of significant haplotypes were ranged from 0.27 to 0.85, while SNPs used to form the corresponding haplotypes were ranged from 0.13 to 0.45. The haplotype analysis did not identify new loci than those previously suggested by single marker analysis, although haplotypes within a window of 1 Mb around previous suggestive SNPs were tested (Table 1). If several markers within a small region are in strong LD, haplotypes can capture the LD information better than SNPs and therefore the power of association is increased to reach the significance level (Bickel et al., 2011, Abdel-Shafy et al., 2014), which poses an advantage for using haplotypes beside SNP markers.

Since the reference buffalo genome is not ready yet, markers on the genotyping array used in the current study were aligned to the latest reference assembly of the bovine genome (UMD 3.1). Therefore, SNPs are located on 30 chromosome pairs including the X-chromosome instead of the 25 pairs as known for water buffalo (Iamartino et al., 2013). Genomic loci reported here are overlapped with previously reported QTL mapped for milk yield in different cattle populations. Identified loci on chromosome 5 coincide with QTL previously reported in American Holstein cattle (Cole et al., 2011), Ireland Holstein cattle (Meredith et al., 2012), and Canadian Holstein cattle (Nayeri et al., 2016). The genomic region on chromosome 1 overlaps with previously recognized QTL in Blonde d’Aquitaine and Limousin cattle (Michenet et al., 2016). Moreover, loci on chromosome 6 are supported by known QTL in American Holstein cattle (Cole et al., 2011, Cochran et al., 2013), and Ireland Holstein cattle (Meredith et al., 2012). While, genomic loci on chromosome 27 are located within QTL previously reported in American Holstein cattle (Cole et al., 2011), Ireland Holstein cattle (Meredith et al., 2012), and Blonde d’Aquitaine cattle (Michenet et al., 2016). Identifying the same loci for different species and populations with different experimental designs and analytical methodologies increases the confidence that the linked neighbourhood genomic regions contain true causative mutations and potential candidate genes.

The significant haplotypes on the four chromosomes were located within or close to several candidate genes. SNPs used to form the identified haplotype on chromosome 5 are located within Glycosyltransferase-like protein (LARGE1) and ELKS/Rab6-interacting/CAST family member 1 (ERC1) genes. The other genomic loci on chromosomes 1, 6 and 27 are closed to some other genes with a distance ranging from 0.19 to 0.31 Mb. The candidate genes for identified regions on chromosomes 1, 5, 6 and 27 were subjected to enrichment analysis using the GeneCodis databases. In this regards, immune response was ranked at the top of all biological processes (8 genes with P-value ≤ 0.001 adjusted by FDR; GO:0006955). Interestingly, several immune related genes were identified as potential candidates for milk production traits in dairy cattle (Beecher et al., 2010, Verbeke et al., 2014, Yang et al., 2016).
Conclusion

Multi-markers based approach can be used beside single markers to identify genomic associations for small population especially for phenotypes with moderate variation. Further investigations of the validated loci in different populations are necessary to identify the potential candidate genes and probably causative mutations.

Table 1. Haplotypes associated with YDs for milk production over the first five lactations in the Egyptian buffalo.

<table>
<thead>
<tr>
<th>HTB ID</th>
<th>Chr</th>
<th>Start [Mb]</th>
<th>End [Mb]</th>
<th>Length [Mb]</th>
<th>NHAP</th>
<th>Alleles</th>
<th>HF</th>
<th>β</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTB 1</td>
<td>1</td>
<td>21.05</td>
<td>21.08</td>
<td>0.03</td>
<td>4</td>
<td>AT</td>
<td>0.36</td>
<td>0.85</td>
<td>0.000108</td>
</tr>
<tr>
<td>HTB 2</td>
<td>5</td>
<td>72.54</td>
<td>72.57</td>
<td>0.03</td>
<td>3</td>
<td>CG</td>
<td>0.30</td>
<td>0.91</td>
<td>0.000039</td>
</tr>
<tr>
<td>HTB 3</td>
<td>5</td>
<td>108.49</td>
<td>108.52</td>
<td>0.03</td>
<td>3</td>
<td>TG</td>
<td>0.85</td>
<td>-1.05</td>
<td>0.000044</td>
</tr>
<tr>
<td>HTB 4</td>
<td>6</td>
<td>80.87</td>
<td>81.01</td>
<td>0.15</td>
<td>4</td>
<td>TTCTT</td>
<td>0.27</td>
<td>-0.86</td>
<td>0.000086</td>
</tr>
<tr>
<td>HTB 6</td>
<td>27</td>
<td>37.14</td>
<td>37.20</td>
<td>0.07</td>
<td>5</td>
<td>CGG</td>
<td>0.41</td>
<td>0.76</td>
<td>0.000073</td>
</tr>
</tbody>
</table>

HTB: haplotype block, Chr: chromosome, Mb: mega base, NHAP: number of haplotypes in each haplotype block, HF: haplotype frequency, β: change per significant haplotype. Positions are according to University of Maryland bovine genome assembly Build 3.1 (UMD3.1)

References


analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC genomics 12:408.


Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: a tool set for whole-genome
association and population-based linkage analyses. American journal of human genetics 81(3):559-575.


