Development of integrated cattle genomics knowledge base

Zorc, M., Kunej, T.

Poljoprivreda/Agriculture

DEVELOPMENT OF INTEGRATED CATTLE GENOMICS KNOWLEDGE BASE

Zorc, M., Kunej, T.

Original scientific paper

SUMMARY

Systems biology approaches being applied to animal breeding represent an opportunity to derive greater benefits from animal production systems. The increasingly detailed investigations in systems biology have led to a large amount of data dispersed over various sources; therefore, a centralized knowledge base is in demand. In this study, we have integrated cattle genomics data of heterogeneous sources and types and developed a bioinformatics tool to study genotype-phenotype associations in cattle: http://integromics-time.com/integromics-database/. The tool enables revealing genomic overlaps within trait-associated loci and identification of potential functional candidates. It might be also used as a tool for planning genotype–phenotype research in cattle.

Key-words: breeding, bioinformatics, cattle, data integration, genomics

INTRODUCTION

A recent development in bovine genome databases has potential to provide advances in finding associations between genotype and phenotype and is beginning to revolutionize gene functions study methods in cattle (Pareek et al., 2011). Comparative integromics combined with systems biology approaches are used (Jiang et al., 2012; Cannistraci et al., 2013) to explain genetic factors and molecular pathways underlying complex phenotypes. In the cattle genomics field regularly updated resources containing genotype-phenotype data are available. For instance, the Animal QTL database (Hu et al., 2013) includes several thousands of QTL for six livestock species, detected with large confidence intervals covering several megabases of the genome containing hundreds of genes. The GeneRIF (Gene Reference Into Function) provides an up-to-date functional annotations of genes. However, the percentage of genes associated with at least one GeneRIF remains quite modest (Lu et al., 2006). Online Mendelian Inheritance in Animals (OMIA) (Nicholas, 2003) is an online database providing up-to-date information on inherited disorders and other familial traits in animal species.

The aim of our study was to: 1) review available sources of genotype-phenotype association data in cattle, 2) develop integrated genotype-phenotype knowledge base for the research in the field of cattle genom-ics and 3) develop a bioinformatics web application for querying the integrated knowledge base.

MATERIAL AND METHODS

The Cattle genomics integrated knowledge base containing genotype-phenotype association data from three data sources: GeneRIF, Animal QTLdb and OMIA. 1) GeneRIF (http://www.ncbi.nlm.nih.gov/gene/about-generif), downloaded in April 2015, includes short statements (255 characters in length) describing a function of a gene, supported by at least one publication, 2) Animal QTLdb (http://www.animalgenome.org/QTLdb/), release 26, is a database of QTL and single-nucleotide polymorphism/gene association data in livestock animal species, 3) OMIA (http://omia.angis.org.au/) is the catalogue of inherited disorders, other (single-locus) traits, and genes in animal species (Nicholas, 2003).

The data is stored within a relational database management system, MySQL (http://www.mysql.com). The data from the database is retrieved through the web interface and developed using HTML, CSS, JavaScript and Apache web server and servlet written in programming language Java, run on servlet container Apache

Ph.D. Minja Zorc (minja.zorc@bf.uni-lj.si), Assoc. Prof. Tanja Kunej – University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Groblje 3, 1230 Domžale, Slovenia

Tomcat. Genomic view, a visualization of distribution of phenotype-related loci along chromosomes, is implemented using Flash GViewer, a freely available flash program developed by the GMOD project (http://gmod.org/wiki/Flash_GViewer). It is inserted into a web page to display chromosomes with genomic locations of individual loci (genes and QTL). To keep the knowledge base up to date, regular updates will be performed by source databases updates.

RESULTS AND DISCUSSION

In the present study, we have reviewed available sources of genotype-phenotype association data in cattle (Table 1) and developed the freely available bioinformatics web application: http://integromics-time.com/integromics-database/. It included genotype-phenotype associations from three data sources, from different research approaches and biological layers, including genomics, proteomics, transcriptomics, and epigenomics. The web interface was used for querying the knowledge base and for visualization of the search results. The user can enter the keyword describing the phenotype of interest. If the knowledge base contains any associations with the phenotype of interest it returns a list of candidate loci and the visualization of loci on the genomic view. The visualization of candidate loci on a

genomic view reveals positional overlaps and serves as a tool for analysis of causal genes that underlie complex traits (Figure 1). Numerous regions of the cattle genome have been linked to QTL for body weight and carcass characteristic. For example, there are QTL related with body weight within all cattle chromosomes (Figure 1A). On the bovine chromosome 20 there was an overlap between three QTL associated with body weight and growth hormone receptor (*GHR*) which was found to play a role in body weight determination in three cattle breeds (Qin, Xu, and Gao 2007) (Figure 1B).

Table 1. The number of genotype-phenotype associations, loci, and phenotypes/traits in cattle in the source databases included in the integrated cattle genomics knowledge base

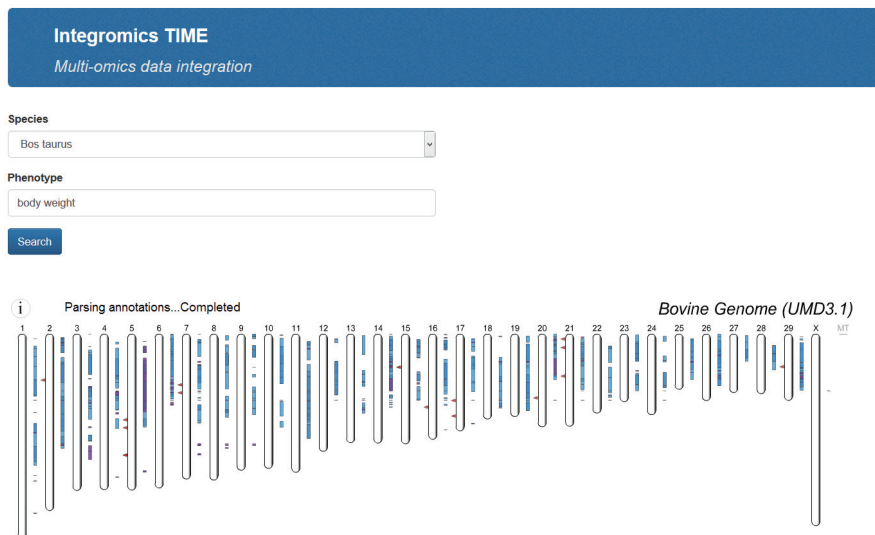| Source	Number of genotype-phenotype associations	Number of loci	Number of phenotypes/traits
GeneRIF	5178	1817	-
Animal QTLdb	17908	17908	514
OMIA	31	31	31

- The number of phenotypes/traits was not provided

Fig. A

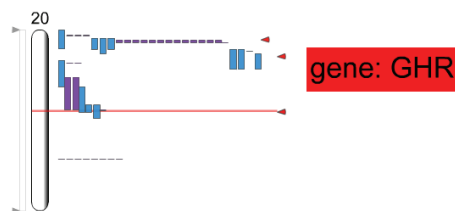gene: GHR(Chr 20: 31890735-32064203)

Fig. B

Figure 1. Web interface for querying the integrated cattle genomics knowledge base. A) Search results (genomic view of body weight related loci), B) Bovine chromosome 20 with body weight related loci

More data sources could be included in the cattle genomics integrated knowledge base. For example, the Bovine Metabolome Database (http://www.cowmetdb.ca/cgi-bin/browse.cgi) is a free web database including cattle metabolites information. Additionally, the databases of bovine candidate genes for milk production and mastitis (Ogorevc et al., 2009) and Obesity Gene Atlas in Mammals (Kunej et al., 2013) are centralized databases of candidate genes for traits in cattle. All efforts related to organization and integration of genomic data are a step toward a faster translation of biomarkers from research into practice.

CONCLUSION

In this study, we have integrated cattle genomics data from heterogeneous sources and types and developed a new bioinformatics tool enabling us to mine the integrated knowledge base. The tool was designed to return genotype associations from the knowledge base for a given phenotype term. The results are presented as a list and as a visualization of genomic locations of loci associated with the phonotype of interest. The presentation of genotype-phenotype associations in this manner enables us to reveal genomic overlaps which might reveal functional candidates at trait-associated loci.

ACKNOWLEDGEMENT

REFERENCES

1. Cannistraci, C.V., Ogorevc, J., Zorc, M., Ravasi, T., Dovc, P., Kunej, T. (2013): Pivotal role of the muscle-contraction pathway in cryptorchidism and evidence for genomic connections with cardiomyopathy pathways in RASopathies. BMC Medical Genomics, 6: 5.
   doi: http://dx.doi.org/10.1186/1755-8794-6-5

2. Hu, Z.L., Park, C.A., Wu, X.L., Reecy, J.M. (2013): Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research, D871-9.
   doi: http://dx.doi.org/10.1093/nar/gks1150

3. Jiang, L., Sørensen, P., Thomsen, B., Edwards, S.M., Skarman, A., Røntved, C.M., Lund, M.S., Workman, C.T. (2012): Gene prioritization for livestock diseases by data integration. Physiological Genomics, 44(5): 305-17.
   doi: http://dx.doi.org/10.1152/physiolgenomics.00047.2011

4. Kunej, T., Jevsinek Skok, D., Zorc, M., Ogrinc, A., Michal, J.J., Kovac, M., Jiang, Z. (2013): Obesity gene atlas in mammals. Journal of Genomics, 1: 45-55.
   doi: http://dx.doi.org/10.7150/jgen.3996

5. Lu, Z., Cohen, K.B., Hunter, L. (2006): Finding GeneRIFs via gene ontology annotations. Pacific Symposium of Biocomputing, 52-63.

6. Nicholas, F.W. (2003): Online Mendelian Inheritance in Animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. Nucleic Acids Research, 31(1): 275-277.

7. Ogorevc, J., Kunej, T., Razpet, A., Dovc, P. (2009): Database of cattle candidate genes and genetic markers for milk production and mastitis. Animal Genetics, 40(6):832-51.
   doi: http://dx.doi.org/10.1111/j.1365-2052.2009.01921.x

8. Pareek, C.S., Smoczynski, R., Pierzchala, M., Czarnik, U., Tretyn, A. (2011): From genotype to phenotype in bovine functional genomics. Briefings in Functional Genomics, 10(3): 165-71.
   doi: http://dx.doi.org/10.1093/bfgp/elr019

9. Qin, Q.M., Xu, S.Z., Gao, X. (2007): Analysis of the correlations between growth hormone receptor gene polymorphism and body size indexes in cattle. Yi Chuan, 29(2): 190-4.