The Genomes and History of Domestic Animals

D. M. Larkin\textsuperscript{a,}\textsuperscript{b,*} and N. S. Yudin\textsuperscript{a,}\textsuperscript{c,}\textsuperscript{**}

\textsuperscript{a}Federal “Research Center Institute of Cytology and Genetics”, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

\textsuperscript{b}Royal Veterinary College, University of London, NW1 0TU, London, United Kingdom

\textsuperscript{c}Novosibirsk State University, Novosibirsk, 630090 Russia

\textsuperscript{*}e-mail: dmlarkin@gmail.com

\textsuperscript{**}e-mail: yudin@bionet.nsc.ru

Received May 4, 2016

Abstract—This paper reviews how mammalian genomes are utilized in modern genetics for the detection of genes and polymorphisms (mutations) within domesticated animal (mostly livestock) genomes that are related to traits of economic importance to humans. Examples are given of how genetic analysis allows to determine key genes associated with the quality and quantity of milk in cattle and key genes for meat production. Various questions are reviewed, such as how contemporary methods of genome sequencing allow to maximise the effective detection of coding and regulatory DNA polymorphisms within the genomes of major domesticated mammals (cattle, sheep and pigs) and the history of their formation from the standpoint of genetics.

DOI: 10.3103/S0891416816040054

THE STRUCTURE OF THE GENOME

Within mammalian genomes, up to 40% of the DNA is comprised of simple repeats, DNA- and retrotransposons, SINE and LINE elements, most of which do not encode for proteins (Fig. 1). Out of 30% of DNA that belongs to genes, only 2% is coding DNA (exons). Heterochromatin DNA, repetitive and unique noncoding DNA sequences comprise a further 30% of the genome. Variation within DNA sequences in the genome as a result of errors during replication is the basis for natural (during evolution) or artificial (animal breeding) selection.

The main sources of variation of DNA in genomes include: (a) point (nucleotide) mutations within genes (originally resulting from mistakes during DNA replication). Point mutations can lead to changes in the protein if the encoded amino acid undergoes changes or lead to regulatory modification if the regulatory sequences of the gene are disturbed. It is important to note, that the majority of point mutations don’t lead to changes of amino acids. Most of them are neutral (synonymous) and don’t have an immediate effect on the protein. The structure and regulation of the gene can be effected by: (b) insertions of transposons inside or near genes; (c) segmental duplications of DNA and chromosomal rearrangements mediated by them, which likely to provide a “gentler” influence, changing the positions of a gene(s) within the chromosome, thus affecting their regulation. Segmental duplications can also lead to a higher number of copies, if genes are duplicated, thus leading to a higher level of gene expression. Traditional ways of searching for genes that affect economically important traits in domesticated animals use the fact that there is a high probability that the gene that controls a trait in one species (for example, a human), will have a similar function in other species of animals. Approximately 12000 genes within the genome of a placental mammal are also represented by one copy within genomes of other placental mammals; i.e. one gene, for example, in a mouse will have a single ortholog with the similar structure and function in a human, a cow or a sheep. About 4000 genes have multiple copies in different

Fig. 1. The fraction of different types of DNA sequences in the human genome. ([http://commons.wikimedia.org/wiki/File:Components_of_the_Human_Genome.jpg](http://commons.wikimedia.org/wiki/File:Components_of_the_Human_Genome.jpg)).
species of animals. Approximately 500 or more genes are unique to every species [1]. The existence of orthologous genes with similar functions in mammals allows us to use information about the functions of genes from well-studied species (e.g. the human or the mouse) for the search of “candidate genes” that have the same function among those of the gene that is looked for. Then, these genes are cloned in species A and a search is conducted for polymorphisms within these genes, segregating with the desired phenotype.

### GENES AND MUTATIONS AFFECTING ECONOMICALLY-IMPORTANT TRAITS IN DOMESTICATED ANIMALS

The principle of “candidate gene” search is presented in Fig. 2. Through the use of linkage mapping techniques, it can be determined that a specific chromosomal locus controls an economically important trait in livestock. Using the methods of comparative mapping, one can find the homeologous regions of human or mouse chromosomes corresponding to this locus. The genes that are present in human or mouse homeologous intervals are screened for ones performing a role similar to the one which is studied in the livestock animal. Gene(s) that control similar traits in a human or mouse, are then cloned and sequenced in the animal. The mutations (e.g. point mutations) are determined in these genes in livestock animal populations, and an associative analysis is conducted, which reveals the relationship (or lack thereof) between the animal’s phenotype and specific nucleotide substitution(s). If an association of this sort is observed, then the gene that influences a particular trait in the animal’s genome was identified correctly. Obviously, this approach has a number of drawbacks, namely: (a) it “works” only for genes with similar functions in different species, and; (b) generally, mutations in regulatory regions of genes are not detected by this approach because they might be located outside the cloned region. Nevertheless, with the help of the candidate gene search technique, important genes have been identified as such the gene influencing the increase in muscle mass in cattle [2]. On the linkage map of the cattle, this trait is mapped (among other locations) to a locus on chromosome 2. This locus corresponds to region 2q31-22 in the human genome, which, among other genes, contains MSTN gene. The knock out of the Mstn gene in the mouse genome significantly increases muscle mass. As a result, MSTN was chosen as a “candidate gene.” After sequencing of the MSTN gene in cattle, mutations segregating with the muscle mass trait had been found.

In cases where the trait is controlled by multiple genes, the situation becomes more complicated, since often the contribution of each of these genes to the phenotype is quite small and often it is not possible to find one or two major genes/mutations that would significantly change phenotype in the desired direction. As shown in Table 1, most of the economically important traits of livestock species are “quantitative” traits. An example of a successful identification of a mutation that has had considerable effect on economically important quantitative trait through candidate gene approach had been the detection of a mutation affecting the fat percentage in cattle milk. Among other chromosomes that trait had been mapped in the centromeric region of cattle chromosome 14 [3]. The locus influencing this polygenic trait was in a region homologous to a chromosomal interval of the mouse genome containing the gene Dgat1. The knockout of Dgat1 in mice affects lactation. After cloning and sequencing of the homologous gene in cattle, several mutations had been found and their segregation with fat percentage of milk from several breeds had been confirmed.

Further progress in identifying genes and mutations associated with important traits in animals has become possible with the advent of high-throughput, low-cost methods of DNA sequencing. It is now possible to not only add the coding regions of genes to the analysis, but also regulatory regions, and introns. A good example of the effectiveness of full genome analysis may be finding a mutation associated with the appearance of

### Table 1. The link between chromosomes and economically important traits in cattle

<table>
<thead>
<tr>
<th>Trait</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>1, 2, 3, 5, 6, 7, 10, 13, 14, 15, 17, 19, 20, 23, 27, 29</td>
</tr>
<tr>
<td>Carcass</td>
<td>2, 5, 6, 7, 13, 14, 15, 19, 23, 27, 29</td>
</tr>
<tr>
<td>Fertility</td>
<td>5, 6, 7, 19</td>
</tr>
<tr>
<td>Disease resistance</td>
<td>1, 2, 13, 23</td>
</tr>
<tr>
<td>Other</td>
<td>2, 4, 5, 6, 14, 23</td>
</tr>
</tbody>
</table>