

Topic: **Databases and Bioinformatics Tools**

PI: **DBT03**

## ORGANIZING OF DATA WAREHOUSE USING PERL AND DBMS FOR GENOTYPING DATA

Andrade L G D<sup>1</sup>, Muniz M N M<sup>1,2</sup>, Tagliatti R F<sup>1</sup>, Guedes E<sup>1</sup>, Silva M V G B D<sup>1</sup>, Arbex W A<sup>1</sup>

<sup>1</sup>Laboratório de Bioinformática e Genômica Animal - Embrapa Gado de Leite - MG - Brazil

<sup>2</sup>Núcleo de Bioinformática - Universidade Federal de Juiz de Fora/Embrapa Gado de Leite - MG - Brazil

The process of genotyping SNPs in domestic animals for using in animal breeding programs generates large sets of data. All these large data sets need to be collected, and stored for further analysis. Extract-transform-load (ETL) process might be very useful to select the desired data from a raw database. After then, it is possible to add relations between data sets into a data warehouse and used it to assist in a decision through analysis of minor allele frequency or any other parameter that needs to be implemented as, for instance, in a stored procedure. For this study, three cattle breeds were used: (a) Gyr (with 319 animals); (b) Nellore (with 959 animals); and (c) Girolando (with 54 animals), as well a database management system (DBMS). A total of 77,256,000 genotypic information from 58,000 SNPs markers for each animal (Illumina Bovine SNP50 Bead Chip) were generated. MySQL was chosen as DBMS for its high performance in query operations. DBMS showed higher performance when compared to the text files, besides the reduction of corrupted data. This study showed that the integration of DBMS and ETL applications to create a data warehouse for genotyping SNPs data in domestic animals is more efficient compared to the traditional way.

Financial support: FAPEMIG, Embrapa, CNPq

SP  
4940  
P.1584940  
P.158

P.A. 3