Computational methods applied to identification of the Dairy Gir breed families

Gisele Silva, Tales Silva, Míria Bobó, Fernanda Almeida, Victor Menezes, Stênio Soares, João Cláudio Panetto, Wagner Arbex

Federal University of Juiz de Fora (UFJF), Brazilian Agricultural Research Corporation (Embrapa)

Abstract

The Embrapa Dairy Cattle and the Brazilian Association of the Dairy Gir Breeders (Associação Brasileira dos Criadores de Gir Leiteiro/ABCGIL) develop for 30 years the Genetic Improvement National Program of the Dairy Gir (Programa Nacional de Melhoramento do Gir Leiteiro/PNMGL). Released in May 2015, the most recent PNMGL's sire summary evaluated and published the results about 300 males, using more than 30,000 information to their daughters or contemporary thereof and more over 32,000 records of dairy production of their daughters and contemporary. Among several sets of data, the PNMGL activities monitoring and maintains two databases (i) for store the certificate (registry) of the individuals (males and females) (ii) with the records of dairy production (females). This work will assists the PNMGL, because it will be important to find traits which can be decisive for increasing milk production. The database contains 92,488 animals registered with the following attributes: the IDs of the animal, father and mother; the animal gender and the animal birth date. The beginning of this study was a descriptive analysis of the database and create a graph where each node represents a animal and each edge represents a relationship between the animals and its parents. This graph was divided into connected components - which can be seen as "animal families" into the database - and a genetic inheritance was assigned according to affiliation of animals, within each of connected components. For the animals whose "genetic load" was supplied by a common ancestor, through father and mother, it is considered one with the biggest amount of transmitted genetic load. The descriptive analysis demonstrated that (a) 94% of animals are female and 6% are males; (b) about 22% of animals does not have a known parents; and (c) nearly 52% of animals do not have descendants. The graph was separated in 3,959 connected components; however, only 12 connected components have more than 10 animals. The largest component covers 94% of animals, containing 15 sires with more than 500 direct descendants. The largest descendants number of a single animal is 4,004, it means that about 4% of registered animals are direct descendant of the same sire. It was observed that 67% animals have some known genetic heritage in his grandparents ascending line. These results were the first steps in a project for explore and reveal new features about important parts of the dataset stored in the last 30 years by PNMGL.

SP 7009