

---

## Finding a minimal set of SNPs for paternity identification in Nelore cattle

Maurício de Alvarenga Mudado, Embrapa Southeast-Cattle Research Center

Luciana Correia de Almeida Regitano, Embrapa Southeast-Cattle Research Center

Polyana Cristine Tizioto, Federal University of São Carlos

Fabiane Siqueira, Embrapa Beef Cattle

Antônio do Nascimento Rosa, Embrapa Beef Cattle

Roberto Higa, Embrapa Agriculture Informatics

Cattle breeding programs often demand pedigree information to associate the heritability of genetic markers to animal performance, like milk and beef quality or disease predisposition. Paternity mistakes are a common problem in cattle herds and genetic markers have been used in parentage exclusion tests for resolving pedigree disagreements. The most common markers used in these tests are microsatellites and SNPs (Single Nucleotide Polymorphisms). SNPs have also been used in GWAS (Genome Wide Association Studies) based on the novel commercial high density bovine SNP chips. Such platforms possess around 770k SNPs which can be used also as a rich resource for paternity exclusion tests. However genotyping such a large amount of SNPs for individuals of a whole flock to perform paternity exclusion tests is not economically feasible. Furthermore, many of these SNPs are non-informative because of high levels of linkage disequilibrium among them and because of low levels of minimum allele frequency (MAF), call rate and genotyping accuracy (GC score). In this work we devised a method to select a minimal set of SNPs from Nelore cattle samples, which are highly informative and are independent, so they could be used in a small genotyping panel for paternity exclusion tests. Samples from 32 Nelore sires chosen to represent the main genealogies marketed in Brazil and 762 nelore bulls produced by artificial insemination were genotyped. The platform used was Illumina BovineHD Genotyping Beadchip that has ~770k SNPs. The algorithm developed first run a quality control filter that selected, from the starting 770k SNPs, only markers that: i) were in Hardy Weinberg equilibrium, ii) had  $MAF > 0.4$ , iii) had Illumina's accuracy GC Score  $> 0.8$ , iv) had call rate  $\geq 0.95$  and v) had a  $D' < 0.5$  with other marker. 126 pairs of father-sons (151 samples), with the paternity already resolved, were selected to compose the test set. The other 643 animals that were left composed the training set. From the training set, 325 SNPs passed the quality control filter. The selected 325 SNPs were used in the test set to perform paternity tests, following the methodology adapted from Fung et al., 2002. The results show that values of the combined power of exclusion (Q) for all markers, between 0.2 and 0.7, gives the correct paternity answer for all pairs of individuals. We believe that this set of 325 SNPs could be used in a small genotyping chip to perform paternity exclusion tests for Nelore cattle; a low-cost solution to be used by cattle producers to correct paternity disagreements. All procedures were performed using the R language in a local linux server machine.