Next generation sequencing and *de novo* assembly of a Nelore (*Bos indicus*) bull genome

L.C. Cintra[1], A. Zerlotini[1], F.P. Lobo[1], F.R. Da Silva[1], P.F. Giachetto[1], P.K. Falcao[1], L.O.C. Silva[2], A.A. Egito[2], F. Siqueira[2], N.M.A. Silva[3], S.R. Paiva[3], M.E.B. Yamagishi[1] and A.R. Caetano[3]

[1]Embrapa Informática Agropecuária, CP6041, 13083-886, Campinas, SP, Brazil, [2]Embrapa Gado de Corte, Av Rádio Maia, 79106-550 Campo Grande, MS, Brazil, [3]Embrapa Recursos Genéticos e Biotecnologia, CP 02372, 70770-917 Brasília, DF, Brazil; alexandre.caetano@embrapa.br

Bos indicus cattle breeds present several natural adaptations to biotic and abiotic stresses found in the tropics and have been extensively used for dairy and beef production in these regions of the world. A *B. indicus* genome assembly represents an essential tool which will be vital to help identify and understand the underlying genetic variations that distinguish taurine and indicine cattle, which have diverged >250,000 years ago, as well as facilitate the work of breeder associations striding towards incorporating genomic tools into ongoing genetic evaluations and breeding programs to improve productivity and beef and milk quality traits. DNA obtained from semen from a Nelore bull born in 1987, with an estimated cumulative inbreeding coefficient of 29.4%, and that can be traced to animals imported from India, was used to produce 100 bp paired-end sequences from short (300 and 700 bp) and long insert (3, 5 and 10 kbp) libraries, with an Illumina HiSeq platform. A total of 1,201 Gbp were sequenced, corresponding to 45× raw coverage of the genome. The SOAP de novo assembler was used to build contigs and scaffolding. Several parameters sets were evaluated to obtain the best assembly based on the number of scaffolds, number of bases in scaffolds, N50, and total gap length. The best assembly obtained so far contains 2.7 Gbp, 15,103 scaffolds with N50 of 649 Kbp and 756 Mbp of gaps. Current results are being used to target additional sequencing of specific libraries to improve scaffold assembly. In addition, additional data generation using different sequencing technologies is underway to improve sequence assembly quality before comparisons with the reference *B. taurus* sequence are performed.