

STATISTICAL METHODS TO EVALUATE REDUNDANCY IN EXPRESSED SEQUENCE TAGS STUDIES

Fernanda Vital de PAULA¹
Fabyano Fonseca e SILVA²
Carlos Souza do NASCIMENTO³
Simone Eliza Facioni GUIMARÃES³
Sebastião MARTINS FILHO²
Moysés NASCIMENTO²
Alencariano FALCÃO¹

- **ABSTRACT:** *Expressed sequence tags (ESTs) surveys are very important to identify genes in sequencing studies of various organisms, but in the presence of high transcript redundancy rate this technique is unviable, since it produces few sequences that were not previously sampled. One of the most relevant redundancy measures is the number of genes, $\Delta(t)$, which may be discovered in a future EST sample t times larger than the original sample. This statistics is useful to direct sequencing protocols of cDNA libraries, since it indicates when this process must be closed, avoiding to sequence stuff already sequenced and the costs related with this one. The present work had as objective to present the theoretical aspects of the statistic $\Delta(t)$ and to propose a classical and bayesian approach for its estimation. Were used data from two cDNA libraries from *Mastigamoeba Balamuthi* organism, and the results showed that the interval estimates obtained for $\Delta(t)$ were significantly more accurate when the Bayesian inference was used.*
- **KEYWORDS:** *EST; cDNA libraries; bayesian inference.*

¹ Universidade Federal do Tocantins – UFT, Escola de Medicina Veterinária e Zootecnia, BR 153, Km 112, CEP 77800-000, Araguaína, TO. E-mail: fernandavmat@yahoo.com.br / ajs.falcao@gmail.com

² Universidade Federal de Viçosa - UFV, Departamento de Estatística, CEP 36571-000 Viçosa, MG. E-mail: fabyanofonseca@ufv.br / martinsfilho@ufv.br / moysesnascim@ufv.br

³ Universidade Federal de Viçosa - UFV, Departamento de Zootecnia, CEP 36571-000 Viçosa, MG. E-mail: carsouzarj@hotmail.com / sfacioniguima@gmail.com