

ABSTRACT

Genome-wide identification and initial characterization of bovine long non-coding RNAs from EST data

Wen Huang¹, Nanye Long, Hasan Khatib

¹ Department of Dairy Science, University of Wisconsin-Madison, Madison, WI 53706, USA. wen.huang@ncsu.edu

It has become increasingly clear that the mammalian genomes produce many long non-coding RNAs (lncRNAs). Accumulating evidence suggests important functions for lncRNAs in a variety of biological processes. However, little is known about lncRNA identity and characteristics in cattle. Using public bovine-specific expressed sequence tags sequences, we reconstructed transcript assemblies, from which reference sequences were obtained for RNAs. Intergenic regions with evidence of transcription were screened for putative lncRNAs using the combination of a gene-finding program and a support vector machine-based tool for the calculation of protein-coding potential. A total of 449 putative lncRNAs located in 405 intergenic regions were identified. Characterization of these putative bovine lncRNAs suggests that they are generally expressed in a tissue-specific manner, their GC contents are higher than randomly selected intergenic sequences but are lower than protein-coding genes, and they are moderately conserved among mammals. This is the first genome-wide catalogue of putative intergenic lncRNAs in cattle and provides important targets for functional studies.