

RNA-Seq Analysis of lncRNAs and cisNATs in the Tomato Genome

S. Hussain Ather, *Boyce Thompson Institute for Plant Research, Cornell University*

ABSTRACT

Since the sequencing of the human genome, it has been revealed that the vast majority of DNA does not code for proteins. Instead, these regions of DNA produce long noncoding RNAs (lncRNAs), which have recently been reported to play important roles such as protein regulation and small RNA processing (Wilusz, Sunwoo, & Spector, 2009). The catalog and functions of lncRNAs in the ripening of tomato species (*Solanum lycopersicum*) are largely unknown. Similarly, the mechanisms of cis-natural antisense transcripts (cisNATs) of proximal complementary RNA strings, which function to inhibit transcription, are also poorly understood (Wang, Gaasterland, & Chua, 2005). Global issues in food production and malnutrition exacerbate the relevance of understanding these biological mechanisms central to the development of fruit. We identified certain functions of lncRNAs and cisNATs in the tomato ripening process using an RNA-Seq pipeline (Wang, Gerstein, & Snyder, 2005). Raw reads from two different stages in the tomato ripening cycle were aligned to a reference genome to test the hypothesis that there would be different expression levels for certain lncRNAs and cis-NATs between the two stages. The two stages were Mature Green, the stage in which the tomato is completely green, and Breaker, the stage in which the tomato shows initial colors of red. Then, the reads were *de novo* assembled, assessed for coding potential, and annotated by transcript and function. Finally, the results were filtered for lncRNAs (length > 200 bp, ORF < 100 bp, noncoding, expression value > 0) and cis-NATs (sense-antisense pairs, overlap length > 50 bp, differential splice patterns, expression value = 0). Differentially-expressed lncRNAs and cis-NATs between the two stages of development were identified, and their functions were analyzed. However, experimental evidence is necessary to confirm our findings and hypothesize models of cis-NAT mechanisms for further classification and identification.

KEYWORDS: bioinformatics, genome, sequencing, evolution, tomato ripening

ACKNOWLEDGMENTS AND CORRESPONDENCE

The author would like to thank Zhangjun Fei for his support and guidance on this research project. This research was funded by a National Science Foundation Research Experiences for Undergraduates Grant (#1061199) and a National Science Foundation Plant Genome Grant (#0923312).

All correspondence concerning this article should be addressed to S. Hussain Ather at sather@umail.iu.edu

Wilusz, J.E., Sunwoo, H., & Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes and Development*, 23(13), 1494-1504. doi: 10.1101/gad.1800909

REFERENCES

- Wang, X.J., Gaasterland, T., & Chua, N.H. (2005). Genome-wide prediction and identification of *cis*-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biology*, 6(4), R30. doi: 10.1186/gb-2005-6-4-r30
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10, 57-63. doi: 10.1038/nrg2484