RNA-Seq analysis of lncRNAs and cis-natural antisense transcripts (cis-NATs) involved in tomato ripening

Syed Hussain Ather, Yi Zheng, Zhangjun Fei
Boyce Thompson Institute for Plant Research, 533 Tower Road
Ithaca, NY 14853

Background

- Some RNA doesn't code for proteins, but may have other purposes.
- Issues in food supply and nutrition cause the need to study the biological factors that control fruit development.
- Tomato ripening is controlled by coding and noncoding RNA.
- **Long noncoding RNAs (lncRNAs)** -noncoding transcripts longer than 200 bp.
- **Cis-natural Antisense transcripts (cis-NATs)** – might inhibit the transcription of opposite (sense) DNA strand with complementary antisense pairing.
- **Next-generation sequencing** allows high throughput, low-cost alignment and assembly of genomes for analysis.

RNA-seq data from tomatos at **mature green** (left) and **breaker** (right) stages were used to identify lncRNAs and cis-NATs and to investigate their expression in ripening.

Method

Clean raw reads to remove contamination

↓

Align RNA-Seq reads to the reference tomato genome (Tophat)

↓

Reference-guided de novo assembly (Cufflinks & METSS)[1]
Multi-Exonic Transcribed Loci by Splice-Site
Filter out false positive assemblies

↓

Merge all samples from both assemblies with ITAG 2.3 annotations

↓

Function annotation for all transcripts (AHRD)

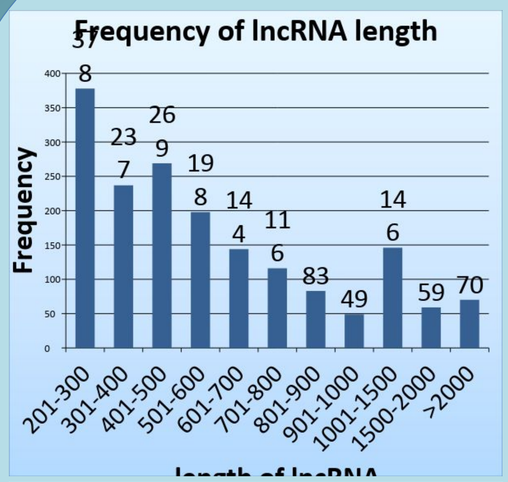
↓

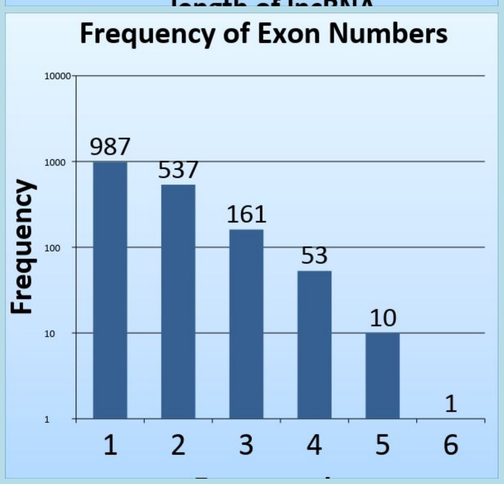
Filter out lncRNA's:
1. Length > 200 bp
2. ORF < 100 bp
3. Noncoding (evaluated by CPC[2])
4. Expression value > 0.

Filter cis-NATs:
1. Sense-antisense pairs
2. Overlapping strands (>50 bp)
3. Diff. splice patterns
4. Expression value = 0

↓

Downstream analysis





- GC Content (% of DNA that is Guanine or Cytosine):
  -lncRNA 35.31%
  -ITAG cDNA 40.56%

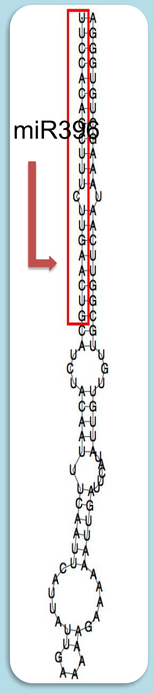
lncRNA

- 1749 lncRNAs were identified.
  -987 had 1 exon
  -762 had multiple exons.
  -66 intronic, 401 intergenic (lincRNA), and 144 cisNAT lncRNAs

(top-left) The frequency of each lncRNA transcript length shows that the median lncRNA length lies between 401 and 500 bp. This might be optimal for functionality since lncRNAs can be negatively affected by small- and large-scale mutations[3].

(bottom-left) With more exons per transcript, more genetic variation occurs, which could lead to improved survival through, for example, disease-resistance.

lncRNAs may fold into the hairpin structure, the precursor of microRNA (miRNA). We detected 18 lncRNAs that could fold into hairpin structures for 24 different miRNAs. (right) miR386 was identified as a potential miRNA from one of these structures.

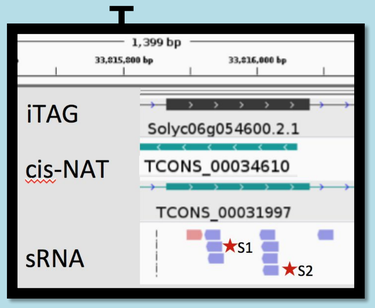


- 102 lncRNAs differentially-expressed:
-64 upregulated in Breaker
-38 upregulated in Mature Green
(39 of the lncRNAs were exclusive to Breaker and 16 to Mature Green.)

cisNAT

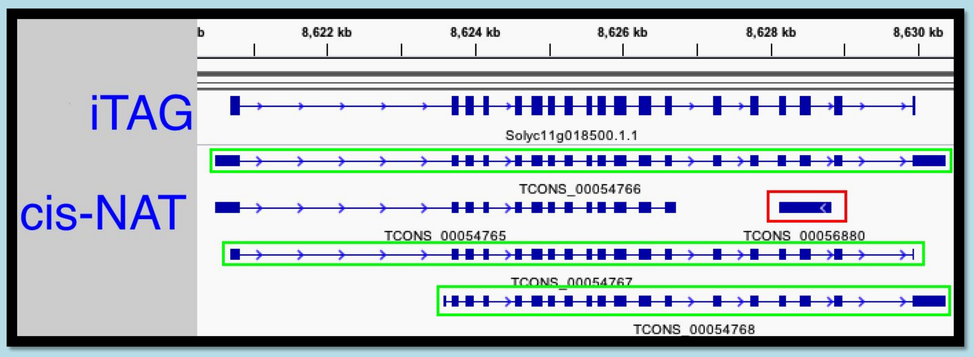
2411 cisNAT pairs
107 noncoding-noncoding pairs
907 noncoding-coding pairs
1397 coding-coding pairs.

(right) TCONS_00031997 (sense) with TCONS_00034610 (antisense) code for the Zinc Finger CCCH domain-containing protein.
8 potential sRNAs (S1 and S2 are highly-expressed).



144 lncRNA cis-NATs
209 lncRNA cis-NAT pairs
190 noncoding-coding pairs
19 noncoding-noncoding pairs

(below) TCONS_00056880 (antisense) with different sense transcripts (in green) code for β-galactosidase. Expressed in the Mature Green stage, and its role is unknown[4].



Conclusion

We used a combination of Cufflinks and METSS to identify lncRNAs and cisNATs in the tomato genome and analyzed how they were involved in tomato development.

What's next?

- lncRNA &Cis-NATs
  -Mechanisms behind cis-NAT transcription not fully understood.
  -Experimental evidence and theoretical models necessary .
  -Further classification and identification
- Improve our statistical techniques of RNA-Seq Analysis to identify lncRNA roles with understanding the nature of lncRNA.

References

[1]Necsulea, Anamaria. S. Magali, M. Warnefors, L. Angélica, D. Tasman, Z. Ulrich, B. Julie, G. Frank, K. Henrik. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505 635–640 (2013).
[2]Lei Kong, Z. Yong, Y. Zhi-Qiang, L. Xiao-Qiao, Z. Shu-Qi, W. Liping, G. Ge. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research.* 35 W345:W349 (2007).
[3]Wapinski and Cheng. Long noncoding RNAs and Human Disease. *Cell Biology* 21 10:561 (2011).
[4]Edgar Moctezuma, S. David, G. Kenneth. Antisense suppression of a β-galactosidase gene (TB G6) in tomato increases fruit cracking. *J. Exp. Bot.* 54 (390):2025-2033. (2003).

Acknowledgements

Special thanks to Tiffany Fleming, Dr. Zhangjun Fei, Yi Zheng, NSF, and the Boyce Thompson Institute for making this project possible.