Finding candidate genes relevant to tomato fruit shape and the corresponding genomic structural variation via a bioinformatic approach

Shamir Harry[1], Nathan Taitano[1], Esther van der Knaap[1], Zachary Lippman[2], Zachary Lemmon[2], Sofia Visa[3]

[1]Center for Applied Gene Technologies, University of Georgia, Athens, GA

[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

[3]The College of Wooster, Wooster, OH

Introduction

- Fruit shape and size varies widely across vegetable crops, affecting their processing ability and fresh market value
- In tomato, the importance of Tonneau1 Recruiting Motif proteins (TRMs) in affecting fruit shape has recently come to light with regard to their interactions with OVATE Family Proteins with emphasis on the collective effect of these interactions on the elongation of tomato fruit [3]
- Homologs of the tomato fruit-shape regulator SUN have also been found regulating fruit shape in other fruits [4]
- Genomic structural variation across tomato populations of South and Central American accessions has been influential in fruit shape diversity

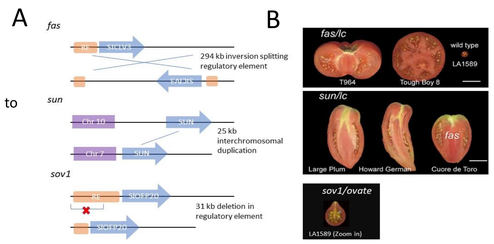

Figure 1. Structural variation underlies *sun, fas,* and *sov1* (A) which in addition to *lc* and *ovate* mutations result in a wide range of fruit shapes (B). [1 2 3 5]

Objectives

- Explore other genes that may act to regulate fruit shape via a coexpression analysis of TRM and SUN genes implicated in fruit shape regulation
- Assess genomic structural variants for overlap with these potential candidate genes in order to examine existing natural variation

Methods

Inputs:

- Gene of Interest (GOI) List: SUN-like genes with fruit shape significance in other organisms and TRM genes including those that were found not to interact with OVATE but either have conserved domain for their interaction or coexpress with it.
- Expression data: RNA-seq data of 4-16 days past initiation flower buds in wild (*S. pimpinellifolium*) LA1589 accession
- Structural Variant (SV) Lists: Two independently generated structural variant lists for mostly semi-domesticated/domesticated (*S. cerasiforme/ S. lycopersicum*) tomato accessions by the labs of Sofia Visa and Zachary Lippman, respectively.
- Annotated Genome: ITAG 2.5 annotated genome for cultivated tomato

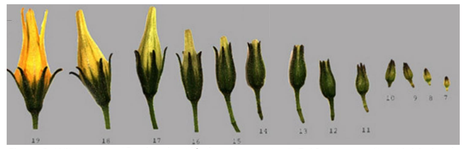

Figure 2. Wild tomato (*S. pimpinellifolium*) flower buds 7-19 days past floral initiation [4]

Final Parameters Ran

- Genes with 0.95 correlation threshold with GOIs
- Only SV deletions that are below 5 Mb
- RE is set as 5'UTR and 5 kb upstream
- Genes defined as relevantly expressed (in a manner likely affecting fruit shape for this case) using conditions that incorporate OVATE and SUN expression patterns and ratios as ideal values for fruit shape regulation associated genes

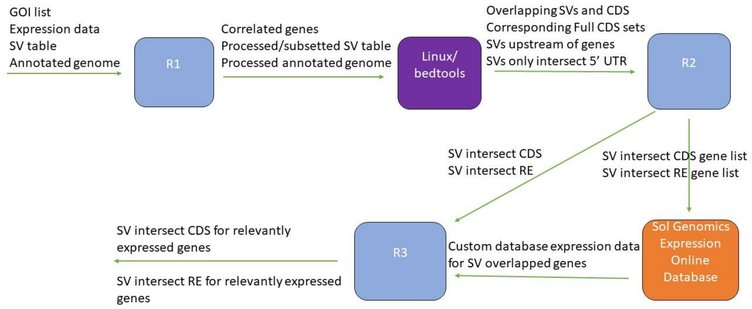

Figure 3. The simplified overall process outlined of the pipeline developed to perform this analysis. In R1, correlated genes from GOI list are found using RNA-seq expression data; additionally, the SV list is subsetted for size and type. In bedtools, SV and correlated gene overlaps are established. In R2, the genes are separated based on if SV overlaps in RE (Regulatory Element) or CDS (Protein Coding Sequence). For those with overlaps in CDS, the full set of CDS for each gene is paired together and the likely mutation induced by each SV is determined. Using the Sol Genomics Expression Viewer online database, publicly available expression data is used to determine if these remaining genes are expressed in preferential fashion via adjustable algorithm.

Results

- Final gene/SV numbers for results ran under Visa SV list is more stringent than those for Lippman's SV list; However, it is yet to be determined if this results in higher quality hits.
- Correlated gene numbers are exceptionally high, indicating tighter parameters or more data necessary for correlations

Figure 4. The number of genes (A) and structural variants (B) through each step of pipeline ran under SV lists generated by the Sofia Visa (green) and Zach Lippman (grey) labs, respectively. Each box shows the number of genes/SVs that meets the given condition and all preceding conditions.

Table 1. A sample of genes taken from pipeline's output.

Gene	Brief Description
Solyc01g091010	YABBY-like transcription factor CRABS CLAW-like protein (Q6SRZ7_ANTMA)
Solyc01g091030	Small auxin up-regulated RNA1 (manually curated)
Solyc01g093960	MADS box transcription factor (Q2NNC0_ELAGV)
Solyc02g081670	anantha (manually curated)
Solyc02g089200	TM29 (manually curated)
Solyc06g071600	Kinetochore protein Spc25 (Q54Q96_DICDI)
Solyc05g014370	Mitotic spindle checkpoint protein

- Prioritize genes from output that directly relate to cell division/growth and those that have been thoroughly studied (manually curated genes).
- Genes that encode for broad or unknown function proteins are largely neglected

Future/Ongoing

- Gather more expression data to increase precision of gene expression correlations
- Run on wide range of parameters to find most statistically significant results
- Find gene expression clusters to outline potential distinguished pathways
- Run again with SNP data and newer SV lists being developed by Lippman lab utilizing nanopore sequencing
- Ensure SVs are real according to sequence alignment data and manual testing
- Research final gene list for those most likely of significance based on literature
- Select the most promising genes for transgenic development/confirmation (generate using CRISPR/Cas9 constructs) and observe any effect on fruit shape
- Hypothesize evolutionary context of observed results using SVs

References

[1]Rodríguez, G. R., Muños, S., Anderson, C., Sim, S. C., Michel, A., Causse, M., ... & van der Knaap, E. (2011). Distribution of SUN, OVATE, LC and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant physiology*, pp-110.

[2] van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., ... & Wu, S. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. Frontiers in plant science, 5, 227.

[3] Wu, S., Zhang, B., Keyhaninejad, N., Rodriguez, G. R., Kim, H. J., Chakrabarti, M., ... & van der Knaap, E. (2018). A novel mechanism underlies morphological diversification in plants. Manuscript submitted for publication.

[4] Xiao, H., Radovich, C., Welty, N., Hsu, J., Li, D., Meulia, T., & van der Knaap, E. (2009). Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape. BMC Plant Biology, 9(1), 49.

[5]Xu, C., Liberatore, K. L., MacAlister, C. A., Huang, Z., Chu, Y. H., Jiang, K., ... & Van Eck, J. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics*, *47*(7), 784.

Acknowledgments

- Thanks to Nathan Taitano and Esther van der Knaap for mentorship, outlining the framework of this project, and conferring insight; Sofia Visa, Zachary Lemmon, Zachary Lippman, for developing SV lists; Yanping Wang for collecting young floral bud RNA-seq expression data.