XIII Argentine Congress of Bioinformatics and Computational Biology
XIII International Conference of the Iberoamerican Society of Bioinformatics
III Annual Meeting of the Ibero-American Artificial Intelligence Network for Big BioData

Bioinformatic identification of sequence variants by integrating omic data to design functional DNA markers in tomato

Valentina Goytia Bertero[1], Paolo Cacchiarelli[2], Guillermo Pratta[2] and Débora Arce[1]
[1]Facultad Regional San Nicolás, Universidad Tecnológica Nacional (FRSN-UTN), Buenos Aires, Argentina;
[2]Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR-CONICET-UNR), Sante Fe, Argentina.
valengoytia19@gmail.com

INTRODUCTION

The cultivated tomato (*Solanum lycopersicum*) is one of the most agriculturally important vegetables. Due to its limited genetic variability, related wild species are often used in breeding programs to expand the available diversity. From previous experiments, we were able to obtain the genome and transcriptome of the cv. Caimanta (*S. lycopersicum*, C) and LA0722 (*S. pimpinellifolium*, P) progenitors.
The objective of this study was to develop molecular markers (*MMs*) from the detected induced genes (*IG*), with a particular focus on Heat Shock Proteins (*HSPs*), a biologically significant superfamily that functions as molecular chaperones, folding other proteins and preventing their denaturation.

MATERIALS AND METHODS

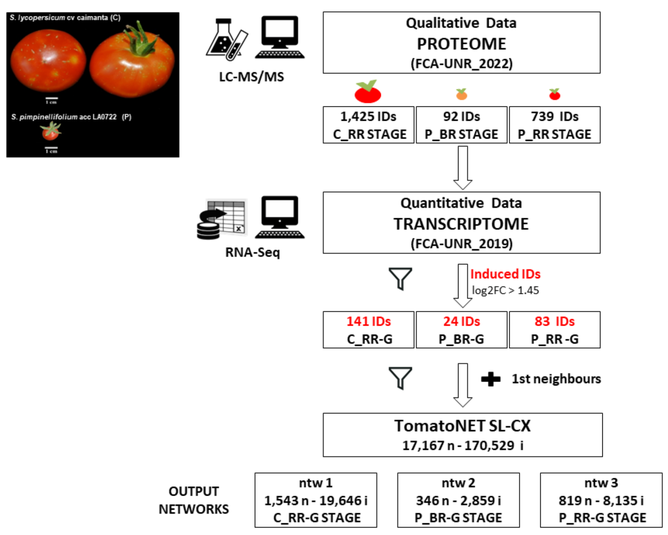

Figure 1. As an original contribution of this experiment, we analyzed proteins obtained through *LC-MS/MS* and subsequently assigned their respective expression levels, obtained from a previous *RNAseq* experiment, allowing us to identify *IG* during the fruit maturation process in both genotypes (C and P).

To accomplish the intended objective, protein sequences of 13 selected chaperones were acquired. Subsequently, a *BLASTn* analysis was conducted on both progenitors to identify the precise start and end positions of the genes. An additional 1000 bases were then added to each start coordinate to extract the *DNA* sequences corresponding to the chaperones from the genomes of C and P. Finally, these sequences were aligned using the *Emboss Needle* program, as depicted in Figure 2.

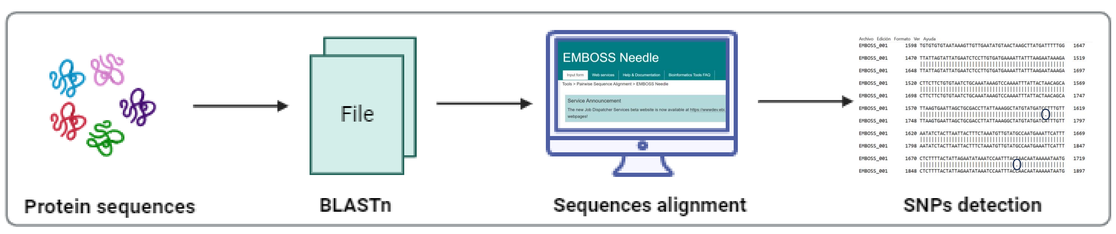

Figure 2. Workflow summary of the work carried out

RESULTS

Using the employed methodology, a substantial number of Single Nucleotide Polymorphisms (*SNPs*) were revealed within 9 of the 13 examined genes, which are itemized in Table 1.

Gene ID	Gene name	SNPs count
SOLYC03G113930	HSP20	11
SOLYC04G082720	HSP20	13
SOLYC05G014280	HSP20	7
SOLYC06G076570	HSP20	28
SOLYC07G006180	RIPENING REGULATED PROT	15
SOLYC07G042250	CHAPERONIN 21	17
SOLYC09G011030	HSP70	15
SOLYC09G075950	HSP70	30
SOLYC12G056780	HSP20	9

Table 1. The nine polymorphic genes detected with their ID, name, and number of SNPs.

Two of the identified small HSPs, marked in blue, have been reported to exhibit the highest differential expression during the fruit ripening process (Krsticevic *et al*., 2016). Among them, **SOLYC05G014280** is also a part of the core set of the four homeostasis genes reported, which appear to offer baseline protection during both fruit ripening and heat shock stress in various tomato tissues. Furthermore, three chaperones, marked in red, have previously been identified as differentially expressed by our research group.

CONCLUSIONS

The development of *MMs* from omics data is anticipated to yield functional markers specifically targeting crucial chaperones involved in the fruit ripening process. These markers are expected to play a significant role in characterizing segregating populations resulting from the crossbreeding of C and P, with a particular emphasis on the key chaperones.

REFERENCES

-Figure 1 belongs to a scientific research work conducted by my research group, under my authorship, which is currently under review.
-Krsticevic, F. J., Arce, D. P., Ezpeleta, J., & Tapia, E. (2016). Tandem duplication events in the expansion of the small heat shock protein gene family in Solanum lycopersicum (cv. Heinz 1706). *G3: Genes, Genomes, Genetics*, *6*(10), 3027-3034..