In Silico Prediction of Cell Wall Remodeling Genes in Tomato, Banana, Melon and Grape

Nurul Hidayah Samsulrizal¹*, Nik Yusnoraini Yusof ²

ABSTRACT
Ripening is a complex developmental process and involves many events such as textural and constitutional changes. The texture of fleshy fruits is one of the major criteria for consumer choice. However, the molecular determinants of ripening-associated changes in texture or “softening” are relatively poorly understood and seem to involve a large number of cell wall remodelling factors. The recent completion of the tomato genome sequence has revealed more than 50 cell wall structure-related genes that are expressed during fruit development and ripening and may impact texture changes in this fruit. The aim of the project is to compare, on a genome-wide scale, ripening-related gene expression in a range of fleshy fruits and especially those linked with cell wall remodelling via computer simulation. Then by identifying orthologous genes in different fruit species to make predictions about those genes likely to be important for the softening process in all fleshy fruits. Comparative genomics analysis of tomato (Solanum lycopersicum), banana (Musa acuminate), melon (Cucumis melo) and grape (Vitis vinifera), has been undertaken using Inparanoid, Multiparanoid and BLAST2GO software. This analysis showed that a total of 8,982 (25.86%) gene models could be identified in common between all four genomes based on comparison of amino acid sequences. Of these genes, 262 in tomato, 252 in grape, 261 in melon, and 198 in banana were identified as encoding cell wall structure-related proteins. However, comparison of the expression patterns of these genes revealed that most were expressed in tissues other than ripening fruits, and of the fruit expressed genes only a small number were common between different fruit species. This in silico analysis should provide additional clues as a target for manipulation of fruit softening in a range of fleshy fruit species. These also provide new opportunities to develop varieties of tomatoes that can survive the trip from the farm to the grocery store whilst maintaining excellent flavour and shelf-life.

Introduction
The revolution in DNA sequencing technology in the last 10 years has enabled the sequencing and assembly of hundreds of genomes from organisms across the tree of life. These genomes include those from a wide range of fruit species. The objective of this

1 Department of Plant Science, Kulliyyah of Science, International Islamic University Malaysia, Bandar Indera Mahkota, Kuantan, Pahang, Malaysia
2 Institute for Research in Molecular Medicine (INFORMM), Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia
*Corresponding Autor: Nurul Hidayah Samsulrizal, e-mail: hidayahsamsulrizal@iium.edu.my;
project was to undertake a comparative genomics study to reveal similarities and
differences in the types of cell wall-structure genes expressed in a range of fleshy fruit
species. The fruit species chosen were tomato (*Solanum lycopersicum*), melon (*Cucumis
melo*), grape (*Vitis vinifera*) and banana (*Musa acuminata*). Tomato is the model for
investigations on the mechanistic basis of ripening and was chosen as our experimental
platform and the genome assemblies for melon, grape and banana were of sufficient
quality, annotated and with expression data to allow the bioinformatics analysis. The
approach was made to identify related gene families and identify orthologous genes across
these species.

Gene families can be defined as a set of genes inherited from a common ancestor that
maintain their sequence and functional similarities [1]. This concept includes the gene
paralogs (pairs of genes with similarities in sequence in the same species) and orthologs
(genes that share similar sequences and have the same function in different species). Genes
in the same family are expected to maintain their molecular structure and biochemical
functions in different organisms [2] and sequence clustering helps to identify the gene
family.

Studying relationships between genes within a family can provide important structure-
function information and provides evidence for ancient genome duplications and
*neofunctionalisation* as is apparent in tomato [3]. The changes in number of genes that are
involved in certain biological processes could occur in several scenarios, for example, gene
duplication that leads to gene families with multiple copies of genes which encode the same
or related functions. The classic model in molecular biology assumes that duplication of a
gene will generate several new genes that are free to be mutated, as long as one of the
original genes retains its function [4]. The most likely outcome for these paralogues is that
they degenerate into pseudogenes that are not transcribed (*nonfunctionalization*) [5, 6].
Alternatively, paralogs may still be functional and through mutation diverge from the
original function (*neofunctionalization*) [4].

Here, the cell wall genes in melon, grape and banana were compared with those in tomato.
The hypothesis to be tested was that cell wall structure-related genes that are common to all
fleshy fruits and expressed during softening are likely to be key targets to allow the
manipulation of fruit firmness. These genes will be the targets of our next experiments in generating transgenic tomato. By identifying and manipulating these genes, it is hoped to improve tomato production in the future.

Materials and Methods

Similarity Search using BLAST

BLAST [7] programs used in this study were blastp (Protein-protein BLAST), blastn (Nucleotide-nucleotide BLAST) and tblastx (Nucleotide 6-frame translation-nucleotide 6-frame translation). The parameter used is the default parameter which is BLOSUM (BLOckS SUbstitution Matrix). BLOSUM62 with opening value space is 11 and the extension value space is 1. The BLAST program that is used in this study was downloaded from NCBI database and installed into a LINUX system on an external server. The similarity search analyses were then undertaken by batch collection (batch blast).

BLAST database format

In this study, formatDB was used for making a custom BLAST database. Putative cell wall structure-related genes were selected in each of the genome. This file was then used to construct the index for the BLAST database using a program from NCBI using a “formatdb” command.

Comparative analysis using in parand and multiparanoid

The genome sequences from four species of fleshy fruit, banana (Musa acuminata), melon (Cucumis melo) and grape (Vitis vinifera) were compared with that of tomato (S. lycopersicum). Only genomes that were assembled, annotated and had associated fruit-related expression data were used. All the pairwise proteome data from the Inparanoid [8] approach were then brought together by using Multiparanoid script. MultiParanoid is a powerful approach for searching for gene clusters among multiple strains so the pairwise orthologous clusters that had been generated from Inparanoid were directly transferred to it.

Gene classification by gene ontology (GO)

GO will provide a comparison of the classification of genes among genomes being studied [9]. It also overcomes issues in linking genes that have been annotated by different researchers when comparison of genes between species is required [10]. The results from
GO classification were used to identify all genes that had a putative cell wall base on cellular component distribution.

**Manual curation with transcriptome database**

Candidate genes will be manually selected and annotated based on GO classification and annotation from tomato genome (our reference genome). Then, the cell wall remodelling genes were selected out of all the proteins that were identified as orthologs in all four genomes. The curations also have been done using all transcriptome data from NCBI database, which focused on those genes expressed during ripening.

**Results**

**Ortholog analysis**

In this research, comparative analysis was done with the detection of orthologous groups using the Inparanoid program which applies *all-versus-all* sequence comparisons of two genomes with the special rules of cluster analysis. Although phylogenetic tree is a well-established method to distinguish orthologs and has been used to study the evolution of organisms, it is time-consuming and prone to errors [11]. Thus, the Inparanoid program was used as an alternative to the phylogenetic method. The *S. lycopersicum* gene models were compared to the list of genes from *C. melo, V. vinifera* and *M. acuminata*. The orthologs for each genome that were generated from Inparanoid program were sorted and viewed using Microsoft Excel program. The numbers of genes shared among all four species were obtained and calculated. Then, the cell wall remodelling genes in this group were identified. *V. vinifera* and *C. melo* were compared with *S. lycopersicum* because all of them were dicot genomes and the genomes have been completely sequenced [3, 12, 13, 14]. *M. acuminata* was used in a comparative analysis because it is a representative for monocot fleshy fruit bearing species [15]. From this analysis (Figure 1.1), a total of 3,013 of the predicted *S. lycopersicum* genes have orthologs in *C. melo*, whereas 2,763 of the gene models are shared in *V. vinifera* sequences and 1,607 of the gene models are represented by orthologous sequences in *M. acuminata*. The different trends of orthologous relationships between *S. lycopersicum* with the two dicots, *C. melo* and *V. vinifera*, and with the monocot *M. acuminata* likely reflect evolutionary processes that occurred in the ancestral genomes.
of each group. Figure 1.1 also shows that all the dicot species (*S. lycopersicum*, *C. melo*, and *V. vinifera*) share a total of 1,340 orthologous sequences while 8,982 (26%) orthologous genes were found to be common to all four species.

**Fig 1-1** Venn diagram of the orthologous genes between *S. lycopersicum, M. acuminata, C. melo* and *V. vinifera*. Numbers in the area of overlap indicate the number of orthologs predicted by reciprocal Inparanoid v 2.0 analysis (threshold E-value = 1x10^{-5})

**Gene classification from GO**

The Gene Ontology (GO) [http://www.geneontology.org] approach is probably the most widespread and the most extensive annotation scheme for the functional description of gene products [16]. A straightforward mapping of gene sequences was made using homology searches (blast hit) [17] to retrieve the GO terms associated with the hit obtained from the Blast results. The 8,982 sequences that had significant blast hits were loaded for GO mapping in Blast2GO suite (Table 1-1). GO terms of a total of 7,791 sequences were successfully assigned based on Gene Ontology Consortium in term of biological process, molecular function and cellular components which then were loaded through mapping to Gene Ontology database. From these, all sequences were assigned to biological processes, molecular functions and cellular components which GO terms could be assigned for more than one term for one sequence and each category was divided into other subcategories.
Table 1-1 Gene Ontology analysis generated using Blast2GO suite

<table>
<thead>
<tr>
<th>Description</th>
<th>Quantity</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence has significant similarity (e-value ≤10^{-5})</td>
<td>8,982</td>
<td>100</td>
</tr>
<tr>
<td>Sequence has Gene Ontology assignment</td>
<td>7,791</td>
<td>86.74</td>
</tr>
</tbody>
</table>

Data mining for cell wall remodeling genes

Fruit softening is a complex process characterized by sequential disassembly and degradation of cell wall components mediated cooperatively by cell wall modifying enzymes [18]. There is also evidence for renewed cell wall biosynthesis during this process [19]. Five GO IDs were used in the identification of cell wall genes common between all orthologous groups, these were GO: 0005623, GO: 0005618, GO: 0044464, GO: 0030312, and GO: 0071944. Then, by using tomato as a model, comparison between the GO classification and tomato genes annotated as linked to cell wall structure, and remodelling [3] was undertaken manually using Microsoft Excel. The Tomato Genome Consortium (2012) reported there were 718 genes that were identified as cell wall structure-related genes in tomato genome. This information was then mapped to the orthologous cell wall sequences. The mapping resulted revealed that there is 262, 261, 252 and 198 cell wall structure-related genes where sequence were highly related in tomato, melon, grape and banana, respectively.

Discussion

Although there are numerous cell wall-related genes, not all are involved in the fruit softening process. For example, in tomato, out of more than 700 genes linked to cell wall metabolisms, only just over 50 were expressed in developing and ripening fruits [3]. Thus, peach and nectarine cultivars also reported only 14 cell wall related genes changed in expression in all cultivars tested [20]. Cell wall classification results using GO terms were mapped to transcriptome data from each of the fruits to identify the genes expressed during fruit development and ripening.
In tomato, 52 cell wall-related genes have been identified as being expressed during fruit development and ripening. However, only 12 genes showed large changes in expression during the ripening process and these included pectin methyl-esterases (PME), pectate lyases (PL), polygalacturonases (PG), and xyloglucan endotransglicosidases (XET) [3]. In melon and grape, approximately 100-200 cell wall genes were identified as expressed during fruit development and ripening which included those that encoded polygalacturonase, pectate lyase, cellulose, and xyloglucan endotransglucosylase. In banana fruits, around 90 genes were expressed during the developing and ripening stage [15], during ripening in bananas some of the most highly expressed cell wall genes were those encoding pectate lyase (2 genes), polygalacturonases (6 genes), pectinacetyl esterases (6 genes), xyloglucan endotransglucosylase/hydrolases (5 genes) and expansins (3 genes) [21].

Although the range of fleshy fruit species studied showed the expression of many of the same families of cell wall-related genes, the most surprising observation was that only a small number of truly orthologous genes were apparent that were expressed in all species (Supplementary Table 1). These include β-glucosidase, cellulose synthase, expansins, polygalacturonase and pectate lyases which were expressed in a wide range of fruits including others not studied in detail here such as apple [22] and strawberry [23]. In tomato, there are three PL genes that are expressed during fruit development and ripening (Solyc03g111690, Solyc05g014000 and Solyc06g083580) (Figure 1-2). The only one of these genes that has a close ortholog in melon, grape or banana was Solyc05g014000. However, this gene is expressed in tomato principally, during fruit development [3, 15]. Solyc06g083580, is expressed only in developing tomato fruits, but expression during ripening is very low [3] which fruit expressed orthologues in melon and grape [12, 13].
Fig 1-2 Molecular phylogenetic tree of pectate lyase (PL) amino acid sequences and the similar amino acid sequences in studied fruits. The dendrogram was generated by Mega 7.0 software using MUSCLE for the alignment and the maximum likelihood method for the construction of the phylogeny. Bootstrap tests were performed using 1,000 replicates and the percentage of the bootstrap value are shown in each branch where the value exceeds 50% is considered significant. The branch lengths are proportional to the phylogenetic distances.

Interestingly, Solyc03g111690 is highly expressed in tomato during fruit ripening, but it does not have a close ortholog in melon, grape or banana. PL was shown to be important in fruit softening in banana [24], strawberry [25], apple [14] and very recently work in the Seymour lab has shown it is very important in tomato [26].

Tomato PG is perhaps the best known pectin degrading enzyme in tomato and is encoded by the gene Solyc10g080210. Orthologues of this gene are present in melon, grape and banana and fruit related expression of these orthologs occurs in tomato, melon and grape. In apple the PG most highly expressed during ripening [22] was more closely related to the tomato gene Solyc05g049980 (Figure 1-3) which was lowly expressed in ripening tomato fruits. In addition, Solyc06g009200, which is not expressed in developing or ripening tomato fruits, is orthologous to a gene that modulates softening in strawberry [27]. These data help highlight that species utilize a range of gene family members during cell wall disassembly.
Fig 1-3 Molecular phylogenetic tree of polygalacturonase (PG) amino acid sequences and the similar amino acid sequences in studied fruits. The dendrogram was generated by Mega 7.0 software using MUSCLE for the alignment and the maximum likelihood method for the construction of the phylogeny. Bootstrap tests were performed using 1,000 replicates and the percentage of the bootstrap value are shown in each branch where the value exceeds 50% is considered significant. The branch lengths are proportional to the phylogenetic distances.

One of the gene families where orthologues showed expression in all fruits were those encoding a CesA-like gene (Solyc08g061100) and also a glucan endo-beta glucosidase-like protein (Solyc03g115200) (Supplementary Table 1). During fruit development cellulose synthases are highly expressed in tomato and then their levels decrease at the breaker stage [3]. They are likely to be involved in the biosynthesis of cellulose [28], but the function of the CesA-like gene product from Solyc08g061100 has not been investigated in fruits. The role of the putative glucan endo-beta-glucosidase-like protein is even more obscure where in tomato its expression declines during fruit development and then increases during
breaker stage (Figure 1-4). Glucan endo-beta-glucosidase has a role in callose decomposition and others have reported that they are expressed during ripening [29].

![Gene expression patterns based on Reads Per Kilobase of transcript per Million mapped reads (RPKM) value of cesA-like gene and β-glucosidase in tomato during fruit development and ripening (Tomato Genome Consortium, 2012)](image)

**Fig 1-4** Gene expression patterns based on Reads Per Kilobase of transcript per Million mapped reads (RPKM) value of cesA-like gene and β-glucosidase in tomato during fruit development and ripening (Tomato Genome Consortium, 2012)

**Conclusion**

In this project, the aim to select genes that were common to all the fleshy fruit species examined with respect to expression during fruit ripening. A comparison of the relationship between cell wall related genes in ripening tomato, melon, grape and banana revealed that there were only a small number of cell wall genes that were likely orthologues and expressed in all fruits that were surveyed. A limited number of these would then be targeted for further functional analysis.

**Acknowledgments**

I wish to thank International Islamic University Malaysia (IIUM) Research Initiative Grant Scheme (RIGS17-022-0597) for sponsored this work and The University of Nottingham, UK for providing the facilities.
References

**Supplementary Table 1** Cell wall structure related genes expressed in tomato and likely orthologues in melon, grape and banana. These genes are expressed during fruit development and ripening

<table>
<thead>
<tr>
<th>Gene Description</th>
<th>Short name</th>
<th>Tomato Gene ID</th>
<th>Melon Gene ID</th>
<th>Grape Gene ID</th>
<th>Banana Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xyloglucan endotransglucosylase/hydrolase 9</td>
<td>SIXTH3d</td>
<td>Solyc03g093080</td>
<td>MELO3C017478P1</td>
<td>GSVIVT01029162001</td>
<td>GSMUA_Achr10P30870_001</td>
</tr>
<tr>
<td>Xyloglucan endotransglucosylase/hydrolase 14</td>
<td>SIXTH5</td>
<td>Solyc01g081060</td>
<td>MELO3C003441P1</td>
<td>GSVIVT01013055001</td>
<td>GSMUA_Achr1P23980_001</td>
</tr>
<tr>
<td>Xyloglucan endotransglucosylase/hydrolase 5</td>
<td>SIXTH16</td>
<td>Solyc07g052980</td>
<td>MELO3C018785P2</td>
<td>GSVIVT01000416001</td>
<td>GSMUA_Achr9P29980_001</td>
</tr>
<tr>
<td>Xyloglucan endotransglucosylase/hydrolase 2</td>
<td>SIXTH26</td>
<td>Solyc05g005680</td>
<td>MELO3C012004</td>
<td>GSVIVT01020228001</td>
<td>GSMUA_Achr3P10480_001</td>
</tr>
<tr>
<td>Xyloglucan endotransglucosylase/hydrolase</td>
<td></td>
<td>Soly03g031800</td>
<td>MELO3C002480P1</td>
<td>GSVIVT01012635001</td>
<td>GSMUA_Achr1P08500_001</td>
</tr>
<tr>
<td>Beta-galactosidase</td>
<td>TGB6</td>
<td>Solyc02g084720</td>
<td>MELO3C007872P1</td>
<td>GSVIVT01018853001</td>
<td>GSMUA_Achr9P21610_001</td>
</tr>
<tr>
<td>Polygalacturonase A</td>
<td>PG-2a</td>
<td>Solyc10g080210</td>
<td>MELO3C023556</td>
<td>GSVIVT01033303001,</td>
<td>GSMUA_Achr10P08930_001</td>
</tr>
<tr>
<td>Polygalacturonase</td>
<td></td>
<td>Soly05g049980</td>
<td>MELO3C006092P1</td>
<td>GSVIVT01019405001</td>
<td>GSMUA_Achr1P02870_001</td>
</tr>
<tr>
<td>Pectinesterase inhibitor</td>
<td></td>
<td>Soly06g009190</td>
<td>MELO3C015963P1</td>
<td>GSVIVT01028041001</td>
<td>GSMUA_Achr1P05430_001</td>
</tr>
<tr>
<td>Pectinesterase inhibitor</td>
<td></td>
<td>Soly07g017600</td>
<td>MELO3C013699P1</td>
<td>GSVIVT01023135001</td>
<td>GSMUA_Achr7P15620_001</td>
</tr>
<tr>
<td>Pectate lyase 1-27</td>
<td></td>
<td>Soly06g083580</td>
<td>MELO3C012791</td>
<td>GSVIVT01028548001</td>
<td>GSMUA_Achr7P04580_001</td>
</tr>
<tr>
<td>Pectate lyase</td>
<td></td>
<td>Soly05g014000</td>
<td>MELO3C002319P1</td>
<td>GSVIVT01000592001,</td>
<td>GSMUA_Achr6P28260_001</td>
</tr>
<tr>
<td>Mannan endo-1 4-beta-mannosidase</td>
<td></td>
<td>Soly02g084990</td>
<td>MELO3C007842</td>
<td>GSVIVT01009746001,</td>
<td>GSMUA_Achr4P02940_001</td>
</tr>
<tr>
<td>Glucan endo-1 3-beta-glucosidase 1</td>
<td></td>
<td>Soly03g115200</td>
<td>MELO3C017220</td>
<td>GSVIVT01007873001</td>
<td>GSMUA_Achr4P32790_001</td>
</tr>
<tr>
<td>Fasciclin-like arabinogalactan protein 19</td>
<td></td>
<td>Soly07g045440</td>
<td>MELO3C024192P1</td>
<td>GSVIVT01014684001</td>
<td>GSMUA_AchrUn_randomP25500_001</td>
</tr>
<tr>
<td>Protein</td>
<td>Accession</td>
<td>Genus</td>
<td>Species</td>
<td>Chromosome</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------</td>
<td>-------</td>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>Fasciclin-like arabinogalactan protein 10</td>
<td>Solyc10g005960</td>
<td>MELO3C024938P1</td>
<td>GSVIVT01030085001</td>
<td>GSMUA_AchrUn_random P25500_001</td>
<td></td>
</tr>
<tr>
<td>Expansin</td>
<td>LeEXP1</td>
<td>Solyc06g051800</td>
<td>MELO3C025907P1, MELO3C003134P1</td>
<td>GSVIVT01024946001</td>
<td>GSMUA_Achr2P16370_001</td>
</tr>
<tr>
<td>Expansin (EXPA3)</td>
<td>Solyc03g031840</td>
<td>MELO3C015695P2</td>
<td>GSVIVT01023857001</td>
<td>GSMUA_Achr1P19730_001</td>
<td></td>
</tr>
<tr>
<td>Endoglucanase 1</td>
<td>Cel8</td>
<td>Solyc08g082250</td>
<td>MELO3C016287P1</td>
<td>GSVIVT01019523001</td>
<td>GSMUA_Achr4P08520_001</td>
</tr>
<tr>
<td>Endoglucanase 1</td>
<td>Solyc04g081300</td>
<td>MELO3C003760P1</td>
<td>GSVIVT01009881001</td>
<td>GSMUA_Achr4P19910_001</td>
<td></td>
</tr>
<tr>
<td>Cellulose synthase-like</td>
<td>Solyc11g066820</td>
<td>MELO3C017935P1</td>
<td>GSVIVT01028071001</td>
<td>GSMUA_Achr3P24160_001</td>
<td></td>
</tr>
<tr>
<td>Cellulose synthase</td>
<td>Solyc01g087210</td>
<td>MELO3C023114</td>
<td>GSVIVT01033297001</td>
<td>GSMUA_AchrUn_random P09460_001</td>
<td></td>
</tr>
<tr>
<td>Cellulose synthase</td>
<td>Solyc08g061100</td>
<td>MELO3C003689</td>
<td>GSVIVT01035830001</td>
<td>GSMUA_Achr5P06050_001</td>
<td></td>
</tr>
</tbody>
</table>