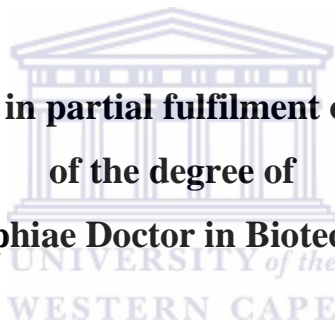# A transcriptome analysis of apple (*Malus x domestica* Borkh.) cv 'Golden Delicious' fruit during fruit growth and development

By

ZEDIAS CHIKWAMBI

**A thesis submitted in partial fulfilment of the requirements
of the degree of
Philosophiae Doctor in Biotechnology**

Department of Biotechnology

Faculty of Science

University of the Western Cape

Supervisor: Dr. D.J.G. Rees

Co-supervisor: Prof. Alan Christoffels

2013

Abstract

A transcriptome analysis of apple (*Malus x domestica* Borkh.) cv 'Golden Delicious' fruit during fruit growth and development

Zedias Chikwambi

Department of Biotechnology, Faculty of Science, University of the Western Cape,
South Africa.

The growth and development of apple (*Malus x domestica* Borkh.) fruit occurs over a period of about 150 days after anthesis to full ripeness. During this period morphological and physiological changes occur defining fruit quality. These changes are a result of spatial and temporal patterns of gene expression during fruit development as regulated by environmental, genetic and environmental-by-genetic factors. A number of previous studies partially characterised the transcriptomes of apple leaf, fruit pulp, whole fruit, and peel plus pulp tissues, using cDNA micro arrays and other PCR based technologies. These studies, however, remain limited in throughput and specificity for transcripts of low abundance. Hence, the aim of this project was to apply a high throughput technique to characterise the full mRNA transcriptome of the 'Golden Delicious' fruit peels and pulp tissues in order to understand the molecular mechanisms underlying the morpho-physiological changes that occur during fruit development.

An RNA-seq analysis of the 'Golden Delicious' fruit peel and pulp tissues transcriptomes was used to identify genes, to analyse gene expression trends and to characterise pathway enrichment in developing apple fruits, which had been harvested at 25, 35, 60, 87, 135, and 150 days after anthesis (daa) from the orchard. Genome-guided (Cufflinks) and *de novo* (Trinity) algorithms were compared to evaluate their ability to reconstruct full-length transcripts. The accuracy of the reconstructed transcripts by the two approaches was evaluated against *M. x domestica* genome v1.0 gene models and ESTs genes, obtained from the Genome Database of *Rosaceae* (GDR) and Refseq respectively. The evaluation revealed that genome-guided assembly was better. A set of non-redundant

transcripts was constructed from the combination of the genome-guided and *de novo* assembly methods, which provided a comprehensive description of the fruit transcriptome.

A two-way analysis of variance (ANOVA) was used to test if tissue types (peel and pulp) and/or fruit developmental period had significant effect on transcript expression patterns. Approximately 10% of the transcripts showed significant tissue and developmental expression patterns, and were clustered into eleven well-correlated groups of expression trends.

Transcripts that showed significant expression in pulp or peel tissue were annotated for molecular function using the main functional categories from the Munich Centre for Protein Sequences (MIPS) database. In the pulp tissue, transcripts encoding proteins with binding function, interaction with the environment, and cell death were mostly enriched. In the peel tissue, transcripts encoding proteins involved in cell fate, interaction with the environment, protein with binding function, development and metabolism functional categories were over represented across time points. In total, approximately 30% of the reconstructed transcripts remained without annotation.

Transcripts that showed significant up-regulation of expression in pulp, encoded enzymes involved in photosynthesis, Calvin cycle, and starch biosynthesis. Peak expression of genes encoding carbon assimilation suggested carbon-fixation and starch accumulation to be highest between daa35 and daa87. The temporal increase in expression of genes encoding sucrose biosynthesis as the fruit matured suggested increased energy demand and degradation of polysaccharides at ripeness, correlating with the conversion of starch to sugars at ripeness.

The plant hormones ethylene and abscisic acid (ABA) are known to regulate a plethora of physiological processes in developing fruit. The genes encoding proteins involved in ethylene and ABA biosynthetic pathways showed increased expression with fruit maturity and were associated with fruit ripening.

The expression of genes encoding cuticle biosynthesis in the peel tissue progressively increased with fruit development, suggesting thickening of the wax layer during fruit development to protect the fruit against uncontrolled non-stomatal water loss.

Fruit colour associated genes encoding for enzymes of the anthocyanin biosynthetic pathway were highly expressed during early fruit development and were associated with the reddish colouration in immature 'Golden Delicious' fruits. The increase in expression of genes encoding carotenoid biosynthesis pathway enzymes with fruit development was associated with the increase in golden colouration with fruit development.

Cell wall integrity associated genes encoding cell wall loosening enzymes were highly expressed in immature peel tissue and were associated with cell division, cell enlargement and protection from turgor pressure and pathogen attack. The genes encoding cell wall degrading enzymes were reduced to very low levels of expression in the later stages of fruit development, which is ideal for maintaining the firmness of the fruit for long shelf life.

The comparative analysis of genes with more than two-fold up-regulated expression in peel or pulp tissue revealed spatial and temporal enrichment of pathways defining tissue properties and functions. In peel, the up-regulation of genes encoding proteins involved in photosynthesis and carbon fixation pathways confirms the peel as an important tissue for carbon assimilation. Up-regulation of genes encoding proteins involved in terpenoid, cuticle, carotenoid and flavonoid biosynthesis pathways in the peel, affirms the contribution of peel tissue to fruit aroma, juiciness, glossiness, texture and colour, which are all important in fruit marketing. On the other hand, genes encoding oxidative phosphorylation, plant hormone signal transduction and C_4-carbon fixation pathway enzymes were enriched in the pulp tissue. Enrichment of both the C_3 and C_4-carbon fixation pathway genes suggests dual carbon fixing mechanisms in 'Golden Delicious' fruit, possibly to cope with the high levels of CO_2 produced by high levels of oxidative phosphorylation in the pulp. This suggests most physiological and morphological changes

in developing fruit to be driven by the pulp tissue. Consequently, the pulp is an important yield component determining fruit size, weight, taste and texture.

Validation of the existing apple genome gene models proves the accuracy of the gene models reconstructed in this project. Spatial and temporal expression of genes in peel and pulp tissue, possibly defining tissue properties and function, has been confirmed. This project has thus identified genes and characterised pathways enriched in peel and pulp tissue of 'Golden Delicious' fruit, which define tissue properties and function.

2013

DECLARATION

I declare that **"A transcriptome analysis of apple (*Malus x domestica* Borkh.) cv 'Golden Delicious' fruit during fruit growth and development**" is my own work that has not been submitted for any degree or examination in any other University and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Zedias Chikwambi   November 2013

Signed:……………………………………..

Key words

Apple fruit

Fruit pulp

Fruit peel

Transcriptome

Differential gene expression

Full-length transcript reconstruction

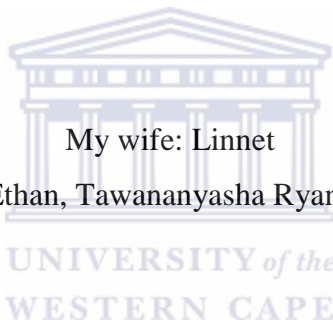Fruit development

Genomics

RNA-seq

Bioinformatics

Acknowledgements

DEDICATION

My wife: Linnet

My sons: Tanatswa Ethan, Tawananyasha Ryan, Taonanyasha Bryan

UNIVERSITY *of the*
WESTERN CAPE

"You are what are important to me"

'Just One Step at a Time'

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE S

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Ab	*Ab initio*
ANOVA	Analysis of Variance
BAM	Binary version of the Sequence Alignment/Map format
CHISAM	Chloroform: Isoamyl alcohol
CTAB	Hexadecyltrimethylammoniumbromide
cv	Cultivar
Daa	Days after anthesis
DEPC	Diethylpyrocarbonate
FA	Formaldehyde Agarose
FAO	Food and Agriculture Organisation
FPKM	Fragments per kilo base of exon per million-mapped sequence reads
Gm	Genome guided transcript assembly
GMAP	Genomic map alignment program
KEGG	Kyoto encyclopaedia of genes and genomes
KO	KEGG Orthology
KOBAS	KEGG Orthology Based Annotation System
MeV	Multi experiment viewer
RT-qPCR	Real time quantitative polymerase chain reaction
SAM	Sequence Alignment/Map format
SSTE	Sodium dodecyl sulphate-Tris-HCl-EDTA

CHAPTER 1

A transcriptome analysis of apple (*Malus x domestica* Borkh.)cv 'Golden Delicious' fruit during fruit growth and development

Abstract

Morphological and physiological changes, which define fruit quality traits, occur during apple fruit growth and development. These changes are a result of spatial and temporal gene expressions as regulated by environmental and genetic factors. Analyses of gene expression in fruit, separated into peel and pulp tissues, sheds light on a molecular understanding of fruit growth and development in apple. Comparisons of expression levels of genes from apple pulp with those from peel tissues can be used to identify genes that define fruit tissue specific characteristics and hence account for most of fruit development and quality traits. This chapter reviews knowledge of what is known to date about the apple tree, fruit botany, physiology, important structural genes in fruit quality determination and their regulation. The characteristics of the peel and pulp apple fruit tissues are reviewed with regards to morpho-physiological properties. However, less is known of their transcriptome and the contribution to the whole fruit transcriptome and quality of these apple fruit tissues. As this review shows, there are gaps in the knowledge of the morpho-physiological changes in apple during fruit growth and development at the molecular level, hence the need for a more in-depth study of the apple fruit transcriptome during fruit development using 'next generation sequencing' approaches.

1.0 INTRODUCTION

The domesticated apple, *Malus x domestica* Borkh (Korban and Skrivin, 1984), also known as *Malus pumila* (Mill.), belongs to the family *Rosaceae*, subfamily *Maloideae*(Janick, 1986; Juniper and Mabberley, 2006; Juniper*, et al.*, 1996; Moore*, et al.*, 1998). This species is believed to be the result of inter-specific hybridisation between *Malus* species (Kellerhals, 2009). A number of theories have been put forward regarding

the origin of *M. x domestica*, however, *M. sieversii* of the central Asia is thought to be the progenitor of this domesticated apple (Coart*, et al.*, 2006; Forte*, et al.*, 2002; Harris*, et al.*, 2002; Juniper and Mabberley, 2006; Juniper*, et al.*, 1996; Vavilov, 1930; Velasco*, et al.*, 2010; Wagner and Weeden, 1999). Merchants plying the Old Silk route between Europe and Asia, spread apples and apple cultivation throughout Asia (Harris*, et al.*, 2002).

The apple is one of the major economically important deciduous temperate fruits and is ranked fourth on the global fruit market. Consumers appreciate the fruit for its flavour, nutritional and health attributes (Boyer and Liu, 2004; Fattouch*, et al.*, 2008; Kahle*, et al.*, 2005; Newcomb*, et al.*, 2006; Soglio*, et al.*, 2009; Vorstermans and Creemers, 2007; Wolfe*, et al.*, 2008). In 2010-2011 production season, world apple fruit production was approximately 65 million metric tons, valued at approximately US$16 billion (FAO, 2010, 2011). This represents 13% of the global fruit production, with almost half of this total being produced by China. Within the South African fruit industry, apple production is ranked 12th. Fruit yields of 0.8 million tons valued at approximately US$200 million were reported to have been produced from 22 500 ha in South Africa (FAO, 2010, 2011). Thus apple fruit production is economically important and any fruit quality and yield improvements are critical for the fruit industry.

1.1 Apple fruit growth and development

The pomaceous simple fleshy fruits of the *Maloideae* members have an expanded inferior ovary surrounded by the cortex or edible fleshy portion of the fruit, the hypanthium. This

structure develops from the fused base of stamens, petals and sepals (Esau, 1977; Janick, 1972; Lee, *et al.*, 2007; MacDaniels, 1940; White, 2002).

1.1.1 Fruit growth

Growth is perceived as an irreversible increase in protoplasm through a series of events in which water, inorganic salts and CO_2 are transformed into living material (Moore, *et al.*, 1998; Raven, *et al.*, 1999). In plants, this process involves the production of carbohydrates in photosynthesising tissue (leaf, green stems and fruit), uptake of water and nutrients, and the formation of complex proteins and fats (Janick, 1972, 1986).

Apple fruit growth follows a smooth sigmoidal response to an increase in fruit fresh weight (Esau, 1977; Hulme, *et al.*, 1968). This response can be divided into three phases based on growth rate. The initial slow increase in fruitlet weight occurs 6 to 12 days from anthesis. This is followed by a rapid exponential increase in fruitlet weight starting from three weeks post anthesis and finally, at a declining rate of growth up to harvesting (McGlasson and Hulme, 1970). Cell division in the pit and cortex of developing apple fruits ceases between 3-4 weeks after anthesis, although fruit thinning can prolong this period. In the epidermis, cell division continues for a considerably longer time than in the cortex. Fleshy cells expand exponentially for 4-6 weeks after blossoming and thereafter, at a diminishing rate, until the fruit is mature (McGlasson and Hulme, 1970).

A reduction in the rate of fruit growth occurs when evaporation and presumably transpiration rates are high. High evaporation and transpiration markedly increases

internal water stress and causes water efflux from fruits resulting in fruit shrinkage especially during the day (McGlasson and Hulme, 1970).

1.1.2 Fruit development

The orderly cycle of development that the whole plant undergoes involves complex patterns of changes in cells, tissues and organs (Janick, 1972, 1986). This cycle in developing apple fruit includes pre-flowering, flowering, pollination, fruit set, fruit ripening, maturation and senescence. The increase in size in the pre-pollination phase is primarily the result of cell division, in response to stimuli and nutrients supplied by the main body of the plant in a source-sink relationship. After pollination cell enlargement is responsible for most of the increase in fruit size (Janick, 1972; White, 2002). In plants with perfect flowers, the stamen primordia are often differentiated before the ovary primordia, and have been shown to be a source of growth stimulus. Early surgical removal of the immature stamens in the flower can adversely affect the growth of the ovary (Moore*, et al.*, 1998).

1.1.2.1 Pollination

Apple fruit trees are self-incompatible; as such they exchange male gametes between compatible cultivars. This inter-cultivar pollen transfer to the stigma (cross-pollination) is aided by insects and sometimes by breeders. The pollen grains that lands onto the sticky stigma, germinate producing a pollen tube, which grows down the style to the ovary. This process functions to initiate physiological processes that culminate in fruit set by inhibiting fruit or flower abscission, and to provide male gametes for fertilization. As pollen tubes grow down the style an auxin hormone is synthesized, which may diffuse

into the ovary inducing growth (Nitsch, 1951). Two sperms from the pollen grain travel through the pollen tube and enter the female gametophyte. One sperm fertilizes the egg, forming a diploid zygote (2n) and the other fertilizes two polar nuclei, forming a triploid cell (3n) (Moore, *et al.*, 1998).

1.1.2.2 Fruit set

Fruit set is the stage when the fertilized ovary is initiated to grow into fruit. The zygote, which is the gateway to the next sporophyte generation, develops into an embryo, while the triploid cell develops into the nutritive endosperm of the seed (Moore, *et al.*, 1998). Even though pollination takes place and fruit set is obtained, fertilization is not assured. Sometimes pollen does not germinate, or if it does, the pollen tube may burst in the style (Moore, *et al.*, 1998). In culture, the germination of pollen is dependant upon the presence of a medium of the proper osmotic concentration and is stimulated by the presence of certain inorganic substances, such as manganese sulphate, calcium, and boron (Moore, *et al.*, 1998).

If the growth of the pollen is very slow, the style or even the entire flower may abscise. This may be artificially modified by the application of auxins, gibberellins, ethylene or cytokines (Gonkiewicz, 2011). When fruit set and growth are obtained without fertilization, the fruit is said to be parthenocarpic. The effect of pollination on fruit set and the influence of the resultant seed on fruit growth make pollination a crucial phase in the production of the apple fruit crop (Moore, *et al.*, 1998). Normally, growth begins in the fertilized ovules and extends to the placentae and the wall of the ovary. Growth of the

nucellus is probably stimulated by auxin associated with pollination. In order to develop, some fruits need only the initial stimulus resulting from pollination, but others such as the apple need an additional supply of hormones to prevent fruit abscission (Nitsch, 1951). Apparently, only the young seeds, possibly in the developing endosperm and embryo, produce auxins (Moore, *et al.*, 1998).

1.1.2.3 Fruit maturation

In this phase, the developing fruit no longer depends primarily upon the parent plant for growth stimuli but on the developing seed within fruit. Deformed fruits have been reported to result from uneven distribution of auxins in developing seeds (Choi, *et al.*, 2011; Janick, *et al.*, 1981; Kühn, *et al.*, 2011). In many fruits, a direct correlation exists between weight and length with seed number. The effect of the seed on fruit development is mediated through chemical substances such as auxins present in extracts of immature seeds, which can stimulate growth of unpollinated tomatoes (Janick, 1986). Furthermore, it is possible to correlate cell division, cell enlargement, differentiation, change in rate of respiration, senescence and other various physiological events in the developing fruit to the presence of auxins, gibberellins, cytokinins, abscisic acid (ABA) and ethylene hormones (Choi, *et al.*, 2011; Janick, *et al.*, 1981; Kühn, *et al.*, 2011). The embryos and endosperms of the developing seeds are sites of many substances that promote fruit growth including several auxins, gibberellin-like compounds, and cytokinins (Janick, *et al.*, 1981). Although the control centre of fruit growth is located in the seed, the plant supplies the raw materials for fruit development at a ratio of 40-50 leaves per fruit for adequate nourishment of the fruit. The availability of nutrients and moisture to the plant

directly affects the growth and development of the fruit (Choi*, et al.*, 2011; Janick*, et al.*, 1981; Kühn*, et al.*, 2011).

1.1.2.4 Fruit ripening

Fruits have two different ripening mechanisms, climacteric and non-climacteric. In climacteric fruits, a peak in respiration accompanies ripening and a concomitant burst in ethylene production. In non-climacteric fruits, respiration rates do not rise as high as in climacteric fruits and ethylene production remains low. In apple fruit a sudden burst in respiration and ethylene production rates is the final physiological process that marks the end of maturation and the beginning of apple fruit senescence (Bangerth*, et al.*, 2012; Kondo*, et al.*, 2012). Increases in CO_2, pigmentation, reduction in texture, and susceptibility of the fruit to pathogen infection are some of the observed changes during fruit ripening (Bangerth*, et al.*, 2012; Kondo*, et al.*, 2012).
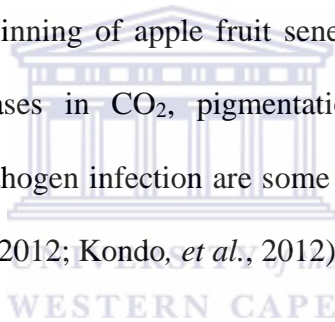
Fruit ripening is transcriptionally controlled, as indicated by an increase in the mRNA of ripening related genes in climacteric fruits followed by an increase in enzymes such as synthetases, hydrolases, and oxidases. Ethylene and perhaps other volatiles appear to have key functions in the ripening process (Janick*, et al.*, 1981; Kondo*, et al.*, 2012; Kühn*, et al.*, 2011)

1.1.3 Fruit anatomical changes during growth and development

The ovary develops rapidly early in the season, reaching maturity well before the hypanthium(White, 2002). This physiological adaptation allows for species perpetuation

even when the conditions are unfavourable for full development of the hypanthium. The developing seeds have been reported to act as reservoirs of auxins for exocarp and mesocarp cell proliferation and enlargement (Cova, *et al.*, 2012).

The hypanthium consists of fleshy parenchyma cells, which show spatial and temporal differentiation defining fruit peel and pulp characteristics throughout fruit development (Esau, 1977; White, 2002). Substantial fruit size increases have been attributed to the actively dividing and enlarging parenchyma cells of the mesocarp, which constitutes the bulk of the pulp. These cells are considered responsible for most of the physiological changes associated with ripening including climacteric phase (Cutter, 1971; Esau, 1977; White, 2002). The pulp is thus the major determinant of fruit total soluble solids (TSS) and total acids (TA). Besides these taste quality factors, the composition of the pulp parenchyma cells are in constant modification throughout fruit development to allow for cell division, expansion and structural roles. These cell-wall modifications determine the fruit texture; biotic and abiotic factor susceptibility, shelf life and fruit crispiness, which all contribute to the consumer experience.

The outer layers of the hypanthium, consisting of epidermis and sub-epidermis (Figure 1.1) have relatively thicker walls and together form the peel. In the early stages of growth, the tender fruit has numerous hairs (trichomes, Figure 1.1), used for protection against desiccation as well as insects and pathogens attack. Trichomes are gradually shed-off and are replaced by a thickening waxy layer, the cuticle(Dong, *et al.*, 2012; Mintz-Oron, *et al.*, 2008).
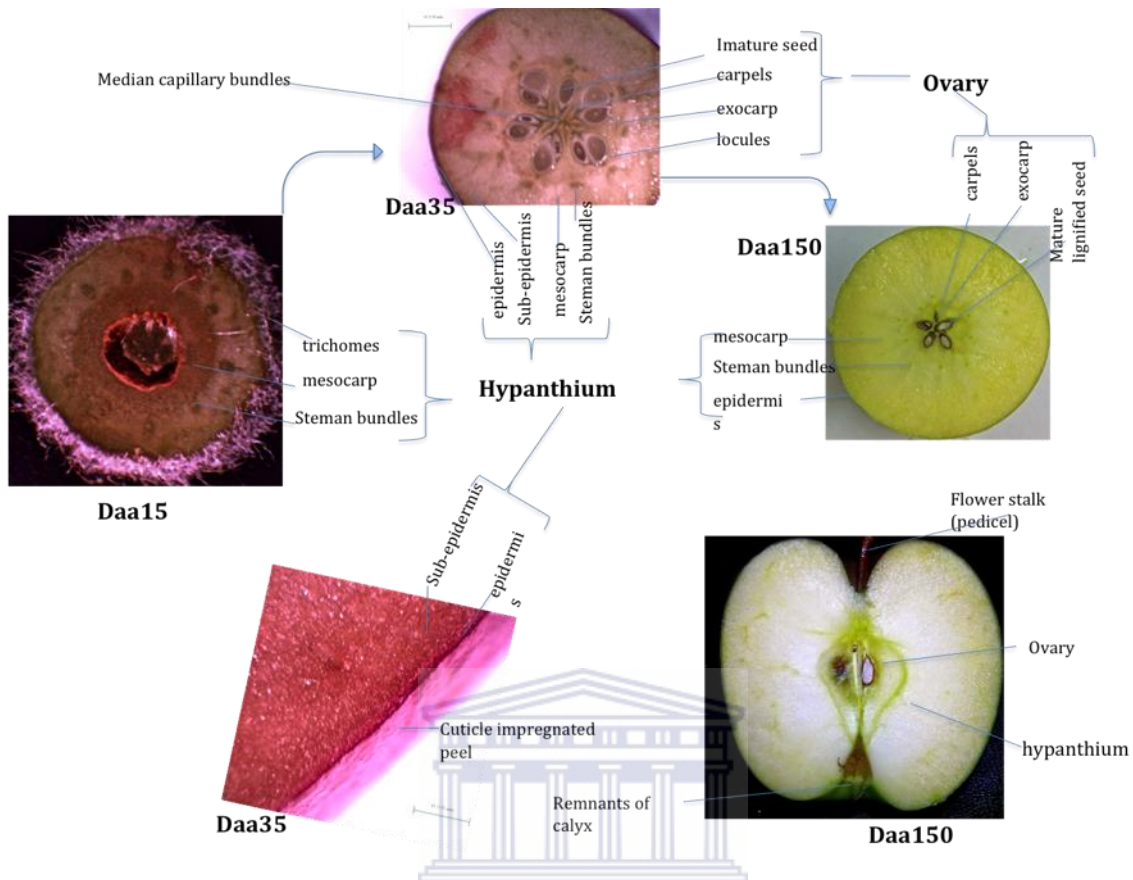
Figure 1. 1 Anatomical and morphological changes in developing apple fruit, cv 'Golden Delicious'. The Figure shows fruit transverse sections at 15,35 and 150 days after anthesis (Daa), as well as longitudinal section of the apple fruit at Daa150. The Figure provides a summary of major changes in internal and external anatomy and morphology of 'Golden Delicious' fruit during development from immature fruits at Daa15 to physiologically mature fruit at Daa150. Each anatomical and morphological attribute manifesting at any particular stage of development has defined role towards fruit development.

The cuticle aids the survival of plants, serving as the interface between plants and their biotic and abiotic environments. The primary physiological function of the cuticle is to seal the tissue against a relatively dry atmosphere, preventing desiccation by minimizing non-stomatal water loss (Mintz-Oron*, et al.*, 2008; Riederer and Burghardt, 2006). Plant surfaces are of importance in ecological interactions, pest resistance and as the first line of defence from mechanical damage. Outside damage can be invoked by herbivores, biting insects or growing fungal hyphae. Inside damage may be due to growing turgor

pressure during fruit growth (Mintz-Oron*, et al.*, 2008). Mechanical failure of fruit surface leads to substantial losses of production through fruit splitting or cuticle cracking.

The young epidermal cells of the developing 'Golden Delicious' fruit have chromoplasts in addition to chloroplasts. These chromoplasts may contain anthocyanins and carotenoids, which give the young fruits a reddish colour (Figure 1.1). Anthocyanins have been reported to act as free radical scavengers, as such may function in protecting the young fruits from oxidative damage. As the fruit matures the green colouration predominates, but is gradually replaced by a golden yellow colour as chloroplasts change into chromoplasts containing carotenoids. The build-up to the golden yellow colour of 'Golden Delicious' is as important as the final uniform colouration of the fruit. Knowledge, therefore, of the physiological contribution to the final fruit colour of each colour change during fruit development is invaluable if a uniformly golden yellow coloured 'Golden Delicious' is to be attained.

The fruit surface influences the outward appearance of the fruit (colour, glossiness, texture, and uniformity), efficacy of post-harvest treatments, storage, transport, and shelf life. The knowledge regarding fruit peel properties is therefore, fundamental for the improvement of fruit quality traits.

1.1.4 Physiological and molecular changes during fruit growth and development

Cells within an organism have the same genome and therefore, genetic potential to give rise to any cell type or embryo, which is regarded as totipotent. Most cells do not divide

after they mature but may dedifferentiate and revert to the meristematic state when induced due to totipotency (den Boer and Murray, 2000).

For cells to enlarge, acids are secreted into the cell walls, activating pH-dependent enzymes that break bonds between cellulose molecules in the wall. Cell-wall acidification associated with cellular elongation is strongly influenced by auxins (Park and Cosgrove, 2012). Positive turgor pressure within the cell resulting from the osmotically driven uptake of water stretches the loosened cell walls, expanding the cells (den Boer and Murray, 2000; Peaucelle*, et al.*, 2012; Wolf*, et al.*, 2012). Plant fruit growth and development are intimately linked with the water and nutrient status of the fruit (den Boer and Murray, 2000). Cell division and enlargement are important for growth but are not the only determinants of development.
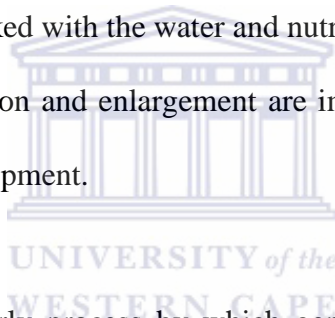
Cell differentiation is an orderly process by which genetically identical cells develop differently in structure and function. The process is always preceded by complex biochemical changes and is governed by a number of factors. Unidirectional signals, such as gravity and light, define cell polarity and influence growth and development. Polarity is controlled by the environment and regulation of the plant genome. Genes are the final arbiters of cellular division, enlargement and differentiation that create polarity (Moore *et al*., 1998). However, these genes are activated by environmental signals such as light, gravity, touch and temperature. Asymmetric divisions are sometimes necessary for differentiation but may not be sufficient for complete differentiation. Additional developmental signals are therefore, necessary to direct the process.

1.1.5 Regulation of growth and development

In addition to totipotency, cellular differentiation results from differential expression of the cell genetic potential as determined by internal and external signals. These signals involve electrical, hormonal, positional, biophysical, and genetic controls (Kohorn and Kohorn, 2012; Moore, *et al.*, 2002). Electrical currents influence division, elongation and polarity of cells, hence differentiation (Moore, *et al.*, 2002). The release of organic molecules, hormones by plant parts elicit physiological responses of immediate or distant tissues of the plant affecting growth and development (Kohorn and Kohorn, 2012; Moore, *et al.*, 2002). There are five major groups of plant hormones, namely auxins, gibberellins, cytokines, abscisic acid, and ethylene affect fruit physiological processes and other hormones or substances such as Ca^{2+} modify their effect. How the hormones affects growth and development of cells programmed to differentiate in a certain way is usually a function of the cell's position and biophysical constraints (Kohorn and Kohorn, 2012; Moore, *et al.*, 2002).

1.1.5.1 Developmental regulation

The overall increase in size of plant organs is driven by cell proliferation with the concomitant generation of cytoplasmic mass, and cell expansion due to uptake of water into the central vacuole (Berckmans, *et al.*, 2011; Tank and Thaker, 2011). The increase in cell number starts well before pollination and is governed by cell cycle genes (Inze, 2003). These repress or promote organ size and are regulated at the transcriptional level early in fruit development (Gasser and Dean, 2009; Janssen, *et al.*, 2008). Upon pollination, the flower releases substantially higher amounts of auxins and gibberellins, stimulating cell division and cell expansion genes, thereby promoting rapid mitotic

12

division of the cells. During this period the genetic material is replicated, new cell organelles are formed and partitioned. The fruitlet relies mostly on carbohydrates photo-assimilated by leaves, and mineral elements translocated by plant roots from the soil. The available photo-assimilates, water, and mineral elements for off-loading in developing fruit determine the rate and extent of fruit growth and fruit size. Translocation and off-loading of assimilates are energy demanding processes, whose energy requirements is met by localized respiration of photo-assimilates (Moore, *et al.*, 1998).

The temporal and spatial gene expression patterns of the genes define the morpho-physiological characteristics of developing apple fruit (Lee, *et al.*, 2007; Sugita and Gruissem, 1987). During apple fruit development, genes are spatially and temporally expressed and can be grouped into functional categories such as, cell proliferation and differentiation, protein synthesis, cell enlargement, metabolism, and stress response (Lee, *et al.*, 2007; Sugita and Gruissem, 1987).

1.1.5.2 Ripening

The ripening process is of considerable importance to the apple industry, as the desire to present fruit to the consumer in the perfect state of ripening, depends on controlling or suspending the normal course of ripening (Bai, *et al.*, 2012; Tan, *et al.*, 2013; Wiersma, *et al.*, 2007). After attaining maximum size, metabolic and physiological modifications occur that confer the characteristics of a mature fruit. These modifications have been attributed to the accumulation of ethylene during the ripening (Bai, *et al.*, 2012; Tan, *et al.*, 2013; Wiersma, *et al.*, 2007). There are two mechanisms of fruit ripening; that is,

climacteric, in which ripening is accompanied by a peak in respiration and a concomitant burst of ethylene, and non- climacteric, in which respiration shows no dramatic change and ethylene production remains at a very low level. The ethylene biosynthesis pathway and its effects in tomato and apple, which are climacteric fruits, have been extensively studied (Alexander and Grierson, 2002; Fleancu, 2007; Soglio*, et al.*, 2009).

Ethylene is formed from methionine via the conversion of methionine (Met) to S-adenosyl-L-Met (SAM) by S-adenosyl-L-Met synthetase (Kondo*, et al.*, 2012). The rate-limiting step involves the transformation of S-adenosyl-L-methionine (AdoMet) into 1-aminocyclopropane-1-carboxylic acid (ACC), which is catalysed by ACC synthase (ACS) (Bai*, et al.*, 2012; Tan*, et al.*, 2013; Wiersma*, et al.*, 2007). ACC is then converted to ethylene by ACC oxidase (ACO) (Kondo*, et al.*, 2012). In addition to ACC, ACS produces 5'-methylthioadenosine, which is utilised for the synthesis of new methionine via a modified pathway that preserves the methyl thio group per cycle (Burstenbinder and Sauter, 2012; Passam*, et al.*, 2011). The high rates of ethylene production can thus, be maintained even when the pool of free methionine is small (Alexander and Grierson, 2002). Positive feedback regulation of ethylene biosynthesis by exposure to exogenous ethylene results in increased ethylene production due to the induction of ACS and ACO, both encoded by small multi-gene families and their expression is differentially regulated by various developmental, environmental and hormonal signals (Alexander and Grierson, 2002).

In higher plants two systems of ethylene production have been proposed. System 1, the ethylene auto-inhibitory system, is considered to function during normal vegetative growth and to be responsible for the basal level of ethylene production. System 2 is instrumental in the upsurge of ethylene production during the ripening of climacteric fruit when ethylene is auto-stimulatory. The transition of ethylene production from system 1 to system 2 is considered to be an important step during fruit ripening, and is developmentally regulated (Bai, *et al.*, 2012; Tan, *et al.*, 2013; Wiersma, *et al.*, 2007). In tomato (Solanum lycopersicum), which is the primary model for climacteric fruit ripening, system 1 ethylene production is regulated by the expression of *LeACS1A* and *LeACS6*. During the transition period from system 1 to system 2, *LeACS1A* expression is increased, and this is followed by induction of *LeACS4* and *LeACS2* expression in system 2 when a competence to ripen has been reached (Davies, 2010; Tan, *et al.*, 2013).
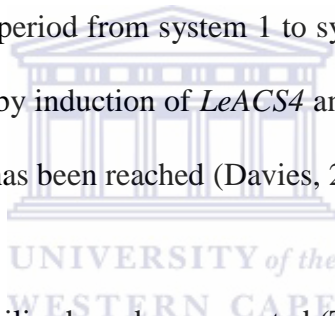
In apple, four *MdACS* gene families have been reported (Tan, *et al.*, 2013), which include *MdACS*1 (accession no U89156), *MdACS*3a (accession no U73816), *MdACS*4 and *MdACS*5 (accession no AB034992). Among these, only *MdACS*3a and *MdACS*1 are expressed specifically in apple fruits, while *MdACS*4 and *MdACS*5 are not expressed in fruits. Both*MdACS*1 and *MdACS*3a play an important role in apple fruit ripening.*MdACS*1-2, an allele of *MdACS*1, is related to long shelf life of apple fruits, while*MdCAS*3a-G289V, a null allele of *MdACS*3a, is responsible for low ethylene production in apple fruits (Bai, *et al.*, 2012). *MdACS*1, which is predominately expressed in the apple fruit, is considered to be involved in system 2 of ethylene biosynthesis because its expression is enhanced by ethylene. *MdACS*3 consists of three family genes

(a, b, c) located on separate loci, but two of them (b, c) possess a transposon-like insertion in their 5'-flanking region, which causes failure of their transcriptions (Tan*, et al.*, 2013). Therefore, *MdACS*3a is the only functional *ACS*3 gene in apple. *MdACS3a* is expressed transiently just before the expression of *MdACS*1, *MdACO*1 and other ripening-related genes such as *ß-polygalacturonase* and *expansin* genes, indicating that the function of *MdACS3*a is pivotal in regulating the transition from system 1 to system 2 of ethylene biosynthesis. (Bai*, et al.*, 2012; Davies, 2004)

Plants respond to ethylene in their environment through a signal transduction pathway involving receptors, a kinase cascade and transcription factors, bringing about changes in the mRNA/protein profile of their tissues (Bai*, et al.*, 2012; Giovannoni, 2001). An increase in the expression levels of ethylene responsive genes leads to ethylene autocatalytic production, which causes changes to the cell metabolism (Soglio*, et al.*, 2009). These changes include synthesis of volatile compounds and increase in sugars and acids. Increase in ethylene production has been reported to stimulate the synthesis of carotenoids in tomato and flavonoids, and anthocyanins in apple (Giovannoni, 2001). These changes affect fruit quality as determined by the consumer.

1.1.5.3 Flavour

Fruit flavour is a complex quality trait determined by the presence of many chemical compounds produced by interacting and interconnected biosynthetic pathways. The principal compounds determining flavour in fruits are alcohols, aldehydes, ketones, sesquiterpenes, polypropanoids, and volatile esters (Schaffer*, et al.*, 2007; Vogt*, et al*,

2013; Zhu, *et al*, 2013), which are involved in the interaction between the plant and the environment. These aromatic compounds are produced from primary metabolites via mevalonate (MVA), methyl-D-erythritol 4-phosphate (MEP), ester biosynthesis and fatty acid biosynthesis pathways among others (Schaffer*, et al.*, 2007).

1.1.5.4 Sugar accumulation

The apple fruit has developed an elaborate system for sugar metabolism and accumulation in the sink cells, called the sucrose-sucrose cycle (Li*, et al.*, 2012, Zhu, *et al*, 2013). This system operates in such a way that it not only allows carbon to be allocated into different pathways to satisfy sink growth and development, but also coordinates sugar metabolism and accumulation and maintains the balance in osmotic potential and turgor between cytosol and other subcellular compartments (Li*, et al.*, 2012). The accumulation of soluble sugars including sucrose, glucose and fructose not only determine the sweetness of the fruit at harvest, but also act as signalling molecules to regulate the expression of key genes involved in plant metabolic processes and defence responses, consequently regulating plant growth and development (Li*, et al.*, 2012; Zhu, *et al*, 2013). However, knowledge of the homeostasis and accumulation of sugars in fleshy fruits still remains limited.
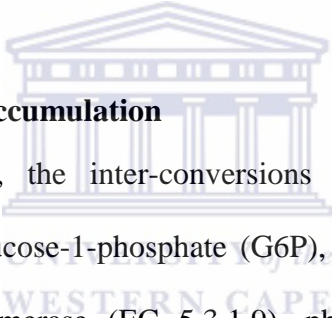
Sucrose in most fruits is the major transported photosynthate, however, in *Rosaceae*, sorbitol accounts for 60-70% of fixed carbon exported from leaves (Li*, et al.*, 2012; Schaffer*, et al.*, 2007; Zhu, *et al*, 2013). Most of the enzymatic steps required for the synthesis of sorbitol in source tissues and its metabolism in sink tissues are known, and

genes encoding these enzymes are well represented in tomato (Moore*, et al.*, 2002) and apple (Li*, et al.*, 2012; Suzuki and Dandekar, 2014) respectively. In leaves, sorbitol is derived from the same hexose phosphate pool as sucrose. The enzyme aldose 6-P reductase (A6PR) is the rate-limiting step, synthesising sorbitol-6-P from glucose-6-P (Zhou*, et al.*, 2003). According to Zhou*, et al.*, antisense suppression of the aldose reductase transcripts in apple plants resulted in a switch from sorbitol production to starch synthesis with no overall effect on the amount of CO_2 fixed (Zhou*, et al.*, 2003). Sorbitol is transported to sink tissues by a family of specialised sorbitol transporters (SOTs) for unloading (Zhou*, et al.*, 2003). Once in the sink tissues such as the fruit, NAD^+-sorbitol dehydrogenase (SDH) converts sorbitol to fructose, thereby re-entering the pool of available sugars (Li*, et al.*, 2012; Moore*, et al.*, 1998; Moore*, et al.*, 2002). The apple fruit is unique in that 80% of the carbon flux passes through fructose because almost all the sorbitol is converted to fructose while only half of the sucrose is converted to fructose (Li*, et al.*, 2012). Apple fruit surfaces contribute up to 15% of the fruit's photoassimilate requirements (Zhou*, et al.*, 2003), but it is not clear how much is directly contributed by the fruit surface towards fruit sorbitol accumulation. According to Suzuki and Dandekar, (2014), sorbitol-to-sucrose ratio in leaves determines fruit quality. High levels of sucrose in the source tissues (leaves) activates transcription of *S6PDH* triggering the synthesis of sorbitol and hence translocation to sink tissues (Suzuki and Dandekar, 2014).

1.1.5.5 Ester accumulation

Straight chain esters in apple are derived from fatty acids such as linoleic and linolenic acid, which in turn are synthesised from acetyl-CoA (Schaffer*, et al.*, 2007; Vogt, *et al*,

2013). In apple, branched chain esters for example, 2-methylbutyl acetate, are derived from isoleucine, which is synthesised from aspartate via threonine (Schaffer, *et al.*, 2007). A considerable number of gene expression profiles have been described for the production of esters in apple (Schaffer, *et al.*, 2007; Vogt, *et al*, 2013). The expression levels of *acetolactate synthase* and *pyruvate decarboxylase*, which are involved in ester biosynthesis, increases with ripening (Soglio, *et al.*, 2009) and so does the aroma of the fruit. Different fruit sections will contribute to the formation of aroma compounds. It is still not clear in which of the peel or pulp fruit sections the aroma- associated genes are mostly expressed.

1.1.5.6 Starch accumulation

In the sucrose-sucrose cycle, the inter-conversions between fructose-6-phosphate, glucose-6-phosphate (F6P), glucose-1-phosphate (G6P), and UDP-glucose (UDPG) are catalysed by phosphoglucoisomerase (EC 5.3.1.9), phosphoglucomutase (PGM, EC 5.4.2.2) and UDP-glucose pyrophosphorylase (EC 2.7.7.9) in reversible reactions. F6P produced in sugar metabolism enters glycolysis and the TCA cycle to generate energy and intermediates for other downstream processes. The G1P is used for starch synthesis, while both F6P and UDPG can be combined to re-synthesise sucrose via sucrose phosphate phosphatase (SPS, EC 3.1.3.24) (Li, *et al.*, 2012). *PGM* is a primary metabolism gene, whose expression increases with ripening and reaches peak in the over-ripening stage. The fate of cellular carbon molecules is determined by this step, catalysed by PGM, resulting in the formation of either malate or complex carbohydrates such as starch. With fruit development, starch is converted to sucrose as respiration increases.

Malate is the main organic acid in ripe apple fruit. It is decarboxylated to pyruvate by a NADP–dependent malic-enzyme, which reaches peak activity when respiration activity is at its maximum. A reduction in the level of malic acid is accompanied by a decrease in NAD-dependent malate dehydrogenase, thus affecting fruit acidity (Soglio, *et al.*, 2009; Soglio, *et al.*, 2007).

1.1.5.7 Colour development

Chlorophyll is distributed throughout the fruit, but during fruit growth and development a negative gradient from epidermis to sub-endodermis is established (Figure 1.1). The onset of ethylene autocatalysis increases chlorophyll degradation throughout the fruit (Chagné, *et al.*, 2012). Chlorophyll is metabolised faster in the epidermis while carotenoids accumulate, which gives the characteristic golden colour of full ripe 'Golden Delicious' fruit (Chagné, *et al.*, 2012). The decrease in chlorophyll pigment content decreases the rate of photosynthesis as the fruit matures. Fruit colour development is also influenced by environmental and cultural factors such as temperature, light and agronomic factors (Dussi, *et al.*, 1995).

The regulation of carotenoid biosynthesis during ripening is due, in part, to the ripening-related and ethylene inducible gene expression in both tomato and melon. Through biochemical and genetic approaches, the enzyme lycopene- €-cyclase, responsible for the relative levels of ß-carotene and lycopene in tomato-fruit was isolated (Foolad, 2009). This led to the development of the tomato ß and/or (crimson) mutants, which results in fruit that accumulates either ß-carotene or lycopene depending on whether the expression

of the *cyclase* gene is enhanced or reduced, respectively (Ashrafi*, et al.*, 2012; Namitha*, et al.*, 2011).

A recessive *high pigment-1 (hp-1)* mutation, results in increased accumulation of both lycopene and ß-carotene during fruit development (Ashrafi*, et al.*, 2012; Namitha*, et al.*, 2011). HP-1 mutation has also been implicated in heightened levels of chlorophyll in leaves and green fruit at all stages of development in lines homozygous for the mutant allele. Homozygous, *hp*-1/*hp*-1 seedlings show inhibition of hypocotyl elongation and intense anthocyanin pigmentation, relative to seedlings of the normal *NIL*, which show maximal phenotypic expression in response to red light (Ashrafi*, et al.*, 2012; Namitha*, et al.*, 2011). The de-etiolation process is a phytochrome (red light) response, which can be enhanced by blue light receptor action and /or signalling. Homologues to the tomato negative regulators of light signalling genes have been identified in apple, suggesting general light signalling in fruit pigmentation (Ashrafi*, et al.*, 2012; Namitha*, et al.*, 2011). A greater understanding of the processes may lead to successful effects in fruit quality and nutrient modification.

Apple is a functional food, containing compounds such as flavonoids, proposed to have antioxidant and anti-carcinogenic properties (Wolfe*, et al.*, 2003). Such traits are important in targeting new breeding strategies. Details of the biosynthetic pathways of two classes of flavonoids, flavonols and anthocyanins, are well known (Mehrtens*, et al.*, 2005). The enzymes of the flavonoid biosynthetic pathway, chalcone synthase 1 (CHS 1) and chalcone-flavone isomerase (CHI), show a decrease in expression with fruit

development except for a putative flavanone 3-ß-hydroxylase enzyme located at the bifurcation of the pathways, leading to anthocyanin or flavonol production (Soglio, *et al.*, 2009). During ripening, expression of the transcription factors *MdMYB11* and *MdMYB10* and the chalcone synthase (*MdCHS1)* genes decrease (Soglio, *et al.*, 2009). A plethora of other genes determining fruit pigmentation, cell wall structure and defence in the flavonoid biosynthetic pathway are also expressed during fruit development (Chalker-Scott, 1999). As the fruit matures, the amount of phenolic compounds synthesised by the fruit decreases (Soglio, *et al.*, 2009).

1.1.5.8 Protein fate

Protein and amino acid metabolism genes involved in ATP/ubiquitin-dependent non-lysosomal proteolytic pathway show a decrease in transcription level during apple fruit development and ripening (Klee and Clark, 2010). Ubiquitin-mediated protein degradation is a mechanism through which plants are able to quickly modulate their response to hormones such as auxin, gibberellins, abscisic acid and ethylene (Klee and Clark, 2010).

1.1.5.9 Cell wall metabolism

Besides changes in cell metabolism, fruit development and ripening involves modifications of cell morphology. This includes changes in membrane transport and cell wall metabolism (Giovannoni, *et al.*, 1989; Giovannoni, 2004; Lau, *et al.*, 2009).

Fruit development starts with an initial phase of cell division followed by cell expansion, mainly due to volume increase of the vacuoles (Li, *et al.*, 2012). There is high expression

levels of genes involved in the transport of solutes in the first stages of fruit development. The aquaporin, gamma tonoplast intrinsic protein and the potassium transporter genes, for example, are abundantly expressed in young apple fruits (Li, *et al.*, 2012). The potassium channel is also involved in potassium accumulation in the vacuole, causing an increase cell turgor and cell expansion (Li, *et al.*, 2012). Cell expansion is however, only achieved after modification of the cell wall structure and cell-cell adhesion. In the presence of auxins, expansin proteins loosen the cell wall, allowing the cells to expand (Lau, *et al.*, 2009).

Expansins are cell wall localised proteins that loosen the cell wall by reversibly disrupting the hydrogen bonds between cellulose microfibrils and matrix polysaccharides (Goulao and Oliveira, 2008). In tomato and apple fruit, *expansin (EXP)* genes are expressed during fruit development (Brummell and Harpster, 2001) and gradually decrease in expression levels during fruit development and ripening, suggesting their involvement in cell enlargement (Janssen, *et al.*, 2008; Soglio, *et al.*, 2009).

Pectin forms the middle lamellae, while cellulose or hemicellulose form the primary cell wall (Goulao and Oliveira, 2008). Depending on the functions of the group of cells, the walls may be lignified making them rigid. The lignified xylem and phloem cells of the vascular bundle are responsible for translocation of water and assimilate while epidermal cells protect the fruitlet from mechanical damage. The rigidity of the cells forming the fruit determines the firmness of the fruit (Alayón-Luaces, *et al.*, 2011).

During the ripening process, the cell wall becomes increasingly hydrated as the pectin-rich lamella is modified and partially hydrolysed (Alayón-Luaces, *et al.*, 2011). The changes in cohesion of the pectin gel governs the ease with which one cell can be separated from another, affecting the final texture of the ripe fruit (Brummell and Harpster, 2001). This process occurs late in ripening apple fruit.

The enzyme pectin methylesterase (PME), is responsible for the de-esterification of the highly esterified polygalacturonans in the cell wall making polyuronides susceptible to degradation by polygalacturonase (PG) (Soglio, *et al.*, 2009).

PG is a major cell-well polyuronide-degrading enzyme, transcriptionally activated during apple fruit ripening (Atkinson, *et al.*, 1998; Brummell and Harpster, 2001; Lau, *et al.*, 2009). The *PG* promoter sequence contains ethylene-specific control elements (Atkinson, *et al.*, 1998; Prasanna, *et al.*, 2007). Using antisense *PG* genes to delay fruit ripening and increase shelf life, a transgenic tomato FLAVR SAVR[TM] was produced (Giovannoni, 2001; Gray, *et al.*, 1992; Lau, *et al.*, 2009). There are however no apple fruittrees in cultivation to date that are equivalent to FLAVR SAVR[TM].

Based on the mode of action, PGs are classified as exo-PGs [exo-poly (1,4 α-D-galacturonide) galacturonohydrolase] or endo-PGs [endo–poly (1,4 α-D-galacturonide) glyconohydrolase]. Exo-PG hydrolyses glycosidic bonds releasing galacturonic acid and is linked to ethylene production (Prasanna, *et al.*, 2007). Endo–PG decreases viscosity by polymerising pectic acid. The composition of endo– and exo–PGs determine the rate of

fruit softening (Lau, *et al.*, 2009). PG is not the major determinant of fruit softening, but low PG fruits are more resistant to splitting, mechanical damage and pathogen infection (Alexander and Grierson, 2002; Brummell and Harpster, 2001). PG extracted from the cell wall may be associated with a PG ß-subunit converter, a 38 kDa glycoprotein distributed throughout the cell wall to control PG diffusion or action during ripening (Alexander and Grierson, 2002; Brummell and Harpster, 2001).

Early in fruit ripening, polymeric galactose within the wall is gradually broken down by ß-galactosidase into free galactose and this continues throughout ripening. The enzyme is encoded by a gene family of at least seven members in apple and tomato, which display different patterns of expression during fruit ripening (Alexander and Grierson, 2002). Strong suppression of ß-galactosidase activity early in ripening can reduce fruit softening by up to 40% and has been reported to be ethylene induced with apple fruit development (Alexander and Grierson, 2002).

1.2 Transcriptome analysis

Fruit growth and development is accompanied by physical and chemical changes, which are linked to cell specific differential gene expression resulting in different cell fates and functions (Morozova, *et al.*, 2009; Varshney, *et al.*, 2009). The approaches for gene expression studies have evolved from candidate gene-based detection of RNAs using Northern blotting (Alwine, *et al.*, 1977) to high-throughput expression profiling driven by micro arrays and most recently massively parallel DNA sequencing approaches (Morozova, *et al.*, 2009). The advancements in DNA sequencing approaches have

resulted in a new breed of technology, 'next generation sequencing' an alternative to microarray-based methods (Sung, *et al.*, 1998; Tu, *et al.*, 2008; Zhou, *et al.*, 2003).

1.2.1   Candidate gene approaches

The initial attempts to elucidate transcriptomes used total cellular RNA from either different organisms, tissues or pathological states in Northern blot analysis as reviewed by (Morozova, *et al.*, 2009). However, the technique has relatively large amounts of RNA input compared to its low throughput. The technique was therefore restricted to the detection of a few known transcripts from samples with unlimited RNA availability. With the advent of reverse transcription quantitative polymerase chain reaction (RT-qPCR) methods (Higuchi, *et al.*, 1993), the throughput and sensitivity of transcript detection thresholds improved whilst input RNA quantities were reduced(Cantacessi, *et al.*, 2011). Considering most biologists aim to approach a transcriptome-wide scale, RT-qPCR remained limited to the throughput of hundreds of known transcripts.

1.2.2   Microarray-based approaches

Complimentary DNA (cDNA) microarray technology (DeRisi, *et al.*, 1996) represented a significant advance for large-scale studies of the transcriptome of apple (Newcomb, *et al.*, 2006; Soglio, *et al.*, 2009). In cDNA micro arrays, thousands of oligonucleotide gene probes (60-70 bp) usually in the form of cDNA, Expression Sequence Tag (EST) clones or PCR products of known genes are individually spotted onto glass slides. Labelled sample genes in the form of mRNA are detected by hybridization to the array probes and transcript abundance is inferred from hybridization intensity. This is an indirect measure compromised by background noise. Experimental controls are therefore, used to ensure the reproducibility and reliable cross-sample comparisons (Morozova, *et al.*, 2009).

Complementary DNA microarray technology can be used without previous sequence information knowledge. This allows analysis of gene expression profiles in plant tissues at different developmental stages or to compare transcript level changes between different genotypes (Newcomb, *et al.*, 2006; Soglio, *et al.*, 2009). Over the years, significant strides have been made towards elucidating the apple transcriptome using microarray analysis of fruit development (Janssen, *et al.*, 2008; Lee, *et al.*, 2007). This includes studying the responses to low temperature and water deficit (Wisniewski, *et al.*, 2008), various treatments in thirty four different tissues (Gasic, *et al.*, 2009; Newcomb, *et al.*, 2006) and fire blight (Norelli, *et al.*, 2009). The growing amount of apple ESTs then necessitated the development of bioinformatics tools to facilitate analysis, hence enabling *in silico* large–scale statistical analysis of ESTs in databanks (Park, *et al.*, 2006). Such analyses have made possible genome annotation and identification of novel transcripts. Transcriptome analysis in apple using micro arrays has incorporated a variety of tissue types from root to fruit using a limited set of gene probes. Such a limited depth in gene numbers reduces tissue-specific transcript coverage. Microarray technologies, however, require high quality control and clone access. Furthermore, there are high chances of cross-hybridization of probes across gene families. The microarray technology has limitations in revealing transcriptional changes of low abundance transcripts, which is typical of transcriptional factors (Soglio, *et al.*, 2009; Soglio, *et al.*, 2007)

1.2.3 Sequence-based approaches

Sequence-based approaches, unlike Northern blots and microarray methods, can directly determine the sequence identity and abundance of a transcript (Elaine R, 2008; Mardis, 2008a, 2008b).

1.2.3.1 Sanger cDNA

The Sanger sequencing method is based on the DNA polymerase-dependent synthesis of a cDNA strand in the presence of 2'-deoxynucleotides (dNTPs) and 2', 3'-dideoxynucleotides (ddNTPs) that serve as non-reversible synthesis terminators (Sanger, *et al.*, 1977). Its discovery resulted in the first determinations of cDNA nucleotide sequences. Despite the many advances in chemistries and the robust performance of instruments like the ABI 3730xl, the application of relatively expensive Sanger sequencing and the associated complex cloning steps results in extremely high costs (Mardis, 2008b). As such routine full-length cDNA (FL-cDNA) sequencing efforts are not feasible, resulting in low coverage, insufficient to comprehensively characterize whole transcriptomes of multicellular species. FL-cDNA sequencing has thus, been limited to novel transcript discovery and annotation. This limitation has apparently been successfully addressed to varying degrees by all of the latest technology offerings.

1.2.3.2 Expressed sequence tags (ESTs)

ESTs are fragments of mRNA sequences derived from single-pass sequencing reactions performed on randomly selected clones from cDNA libraries (Adams, *et al.*, 1993; Seo and Kim, 2009). ESTs have been used to identify new members of gene families, as makers on chromosomes, to discover polymorphic sites and to compare expression patterns in different tissues or pathological states (Clifton and Mitreva, 2009; Gill and Sanseau, 2000; Parkinson and Blaxter, 2009; Seo and Kim, 2009; Sung, *et al.*, 1998). *In silico* analysis of publicly available ESTs has been used for the analysis of global gene expression in correlation to particular stages of development or induced in response to

biotic or abiotic factors (Clifton and Mitreva, 2009; Gill and Sanseau, 2000; Parkinson and Blaxter, 2009; Seo and Kim, 2009; Sung, *et al.*, 1998). The technique may have solved the cost limitations of FL-cDNA sequencing making it accessible. The sequence information is however incomplete and expensive for routine transcriptome-wide scale analysis. In addition, due to the low redundancy of sequence reads, EST data is not suited for estimating transcript abundance (Parkinson and Blaxter, 2009). This may be solved by full-length cDNA cloning and sequencing, but costs are high compared to short-tag approaches such as serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS).

More than 261 000 apples ESTs are available in the Genebank (Park, *et al.*, 2006; Soglio, *et al.*, 2009), a repository for gene mining. If aided by micro arrays, the databases can be used to select candidate genes important in particular crop traits. The growing numbers of the ESTs has increased the number of simple sequence repeats (SSRS) and single-nucleotide polymorphisms (SNPs), which are invaluable markers for marker-assisted selection (MAS).

1.2.3.3 Short tag sequencing

Serial Analysis of Gene Expression (SAGE) and Massively Parallel Signature Sequencing (MPSS) methods measure short-tag sequences and facilitates the use of Sanger sequencing for gene expression profiling (Hu and Polyak, 2006; Velculescu, *et al.*, 1995). In SAGE and MPSS, cDNA fragments (short tag sequences 14 bp to 21 bp per transcripts) are extracted, concatenated, cloned and sequenced. However, unlike SAGE, MPSS uses an *in-vitro* cloning process and hybridization-ligation based parallel

sequencing (Meyers, *et al.*, 2004; Reinartz, *et al.*, 2002). Long-tags from MPSS and SAGE allows formore specific read mapping directly to genome sequences (Velculescu, *et al.*, 1995), and data comparison across different experiments and the discovery of novel alternative splice isoforms. However, SAGE involves a laborious cloning step and has a slow linear sequencing process. The process is costly and produces short tags (14 or 21 bp) difficult to use in resolving transcripts with high sequence identity (Hu and Polyak, 2006).MPSS simultaneously access millions of transcripts, which are one order of magnitude greater compared to SAGE. However, owing to the complex processing involved, specialized instrumentation is required, which is often not feasible for standard laboratories (Albrecht, *et al.*, 2012; Reinartz, *et al.*, 2002).

Recently, the Sanger method has been partially supplanted by several "next-generation" sequencing platforms that offer a dramatic increase in cost-effective sequence throughput but at the expense of read lengths (Albrecht, *et al.*, 2012; Reinartz, *et al.*, 2002).

1.2.3.4 Next generation sequencing (NGS) approaches

A new generation of DNA sequencers that can rapidly and inexpensively sequence billions of bases are being used for whole-genome sequencing and for a variety of sequencing-based assays, including gene expression, DNA-protein interaction, genome re-sequencing and splicing studies (Fullwood, *et al.*, 2009; Schuster, 2008). To date, the best next-generation sequencing platforms are commercially available and continue to be improved. These include the Roche 454 Applied Science, Illumina HiSeq systems, Applied Biosystems SOLiD (ABI Supported Oligonucleotide Ligation and Detection) systems, and the Ion Torrent platforms. The 454/Roche platform (Margulies, *et al.*,

2005), the first next-generation sequencing technology released on the market, is based on PCR pyrosequencing technology, and circumvents the cloning requirement by taking advantage of a highly efficient *in vitro* DNA amplification method known as emulsion PCR. The 454/Roche platform produces extra-long reads of up to 1000 bp of high quality, important for *de novo* transcriptome assembly(Mardis, 2008a). The Illumina HiSeq systems use reversible-terminator sequencing by synthesis (SBS) technology. The 454 uses emulsion bead micro-reactors whilst the latter incorporates a parallel solid state sequencing technique, which amplifies the product on a glass plate.

Massively parallel sequencing by hybridization-ligation, implemented in the SOLiD (Pandey*, et al.*, 2008), is based on the Pollinator technology(Mardis, 2008a; Pandey*, et al.*, 2008). The SOLiD is an open source sequencer that utilizes emulsion PCR to immobilize the DNA library onto a solid support and cyclic sequencing-by-ligation chemistry(Shendure*, et al.*, 2005). The Ion Torrent is based on pairing of semiconductor technology and simple base chemistry (Henson*, et al.*, 2012; Merriman*, et al.*, 2012; Quail*, et al.*, 2012; Rothberg*, et al.*, 2011). The high throughput capacity of these platforms has kept them at the forefront of the genomic and transcriptomic research (Mardis, 2008a), each addressing a specific niche in the genomics market.

1.2.3.4.1 Illumina

The Illumina approach achieves cloning-free DNA amplification by attaching single-stranded DNA fragments to a solid surface known as a single-molecule array, or flow cell, and conducting solid-phase bridge amplification of single-molecule DNA templates (Bentley*, et al.*, 2008)). In this process, one end of single DNA molecule is attached to a

solid surface using an adapter. The molecules subsequently bend over and hybridize to complementary adapters (creating the "bridge"), thereby forming the template for the synthesis of their complementary strands. After the amplification step with more than 1.6 billion clusters/flow cell, wherein each cluster is composed of approximately 1000 clonal copies of a single template molecule. The templates are sequenced in a massively parallel fashion. This is done using a DNA sequencing-by-synthesis approach that employs reversible terminators with removable fluorescent moieties and special DNA polymerases that can incorporate these terminators into growing oligonucleotide chains. The fluorescent terminators are labelled with four different colours to distinguish among the different bases at the given sequence position. Previous Illumina read lengths were short (~35 bases), however, new Illumina instruments such as the Hiseq and MiSeq have increased the read lengths to 150 bases paired-end and 2 x 250 bases paired-end on HiSeq and MiSeq respectively. This improvement has allowed Illumina long reads to resolve short sequence repeats. Minoche, *et al*, (2011) in his evaluation of the HiSeq and Genome Analyser IIx (GAIIx), noted the existence of different error profiles within the Illumina platform family, but with sequence quality filtering, downstream analysis artefacts were minimised (Minoche*, et al.*, 2011). Use of strand specificity assists in distinguishing sequencing errors from low abundance polymorphisms (Minoche*, et al.*, 2011). However, errors are not exclusive to the Illumina platform family, but are found across platforms (Mardis, 2008a).

This technology has been considered ideal for re-sequencing projects, targeted sequencing, SNP analysis, and gene expression studies(Cantacessi*, et al.*, 2011). The Illumina platform has been used for transcriptome analysis of sesame (*Sesamum indicum*

L., Chickpea (*Cicer arientinum* L.), radish, *Arabidopsis*, soybean, *Brassica juncea*, sorghum and grapevine as well as apple with considerable success (Fasoli, *et al.*, 2012; Hiremath, *et al.*, 2012; Krost, *et al.*, 2012; Le, *et al.*, 2011; Mizuno, *et al.*, 2012; Sun, *et al.*, 2012; Torti, *et al.*, 2012; Wang, Wang, *et al.*, 2012; Zhang, *et al.*, 2012).

1.2.4   *M. x domestica* genome sequencing: Perspectives and progress

Most cultivated apples are diploids ($2n = 34$), self-incompatible, and display juvenile periods of more than six years (Gasic, *et al.*, 2009). *M. x domestica* Borkh has been classified into the Pyreae tribe, considered by some schools of thought to be an example of an allopolyploidisation between Spiraeoideae (x=9) and Amygdaleoideae (x=8) (Velasco, *et al*., 2010). Nonetheless a within-lineage polyploidisation event has also been hypothesized with evidence of recent and old genome wide duplications (GWDs), as well as three major rearrangements in the past 20-30 million years (Velasco, *et al.*, 2010). The chromosome duplication and subsequent fusion events resulted in the distinct x=17 apple chromosome number state. According to Velasco, *et al*., (2010) duplication, rearrangement and polyploidisation state of the apple genome provides challenges in sequencing and assembling the genome.

Over the period 2007 to 2009, a consortium of research institutes coordinated by the Foundation E. Mach – 1st Agrario San Michelle All 'Adige, Research and Innovation, sequenced and assembled around 13 billion nucleotides of diploid 'Golden Delicious' apple using Sanger and 454 technologies (Velasco, *et al.*, 2010). Though the sequenced nucleotides provided a 17-fold genome coverage only 742.3 million nucleotides were assembled into the apple genome contiguous sequences (contigs). About 1643

metacontigs made up of 600 million nucleotides were anchored to the 17 apple chromosomes. The remainder of the contigs constituting 20% of the assembled contigs assembled into a mixture of small contigs that could not be anchored to specific chromosomes and repetitive sequences. Using various gene predication tools about 57 thousand genes were predicated to reside on the apple genome. However, only 90.2% of the genes could be assigned to the chromosomes (Velasco, *et al.*, 2010).

In a second attempt at improving the draft apple genome, the Institute Agrario Di San Michelle All' Adige in collaboration with NCBI reassembled the original *M. x domestica* genome v1.0 into pseudo haplotype assemblies, primary and alternative 1, 2 and 3. Reassembly of the apple genome into haplotypes was intended to divide overlapping contigs in the original draft genome into consensus sequences representing four different haplotypes present in the apple genome v1.0. Currently, however, only the primary pseudo haplotype is available on GDR website (www.rosaceae.org/). During the construction of the primary pseudo haplotype assembly, small contigs were removed from the dataset, and contigs with large gaps were split. A weighted network graph was constructed with contigs as nodes and distances between contigs as edge weights. Dijkstra's algorithm was then used to find the shortest path between all nodes along the genome without overlaps. The primary assembly represents the minimal number of gaps between non-overlapping contigs used to reconstitute the pseudo molecules of the genome (www.rosaceae.org/).

The non-inclusion of the smaller contigs as well as the continued presence of unanchored contigs represents a significant gap in the apple genome. Thus the knowledge of the gene

distribution in the apple genome could be confined to the anchored contigs. Gaps of 200 000 'N's were used to space scaffolds in the pseudo haplotype assemblies, however in as much as they allow estimation of chromosome size, potentially significant information still lay in these gaps.

Initiatives in reducing the complexity of the current apple genome v1.0 have been undertaken. Several groups including the Biotechnology Platform (ARC) in South Africa have sequenced double haploid 'Golden Delicious' using Illumina technology. Single and paired end Illumina sequences have been used to fill in the genome gaps using programs such as Gapfiller. Several modifications to the apple genome repair attempts have been instituted. These modifications include increasing the read length, insert size, and read coverage as well as use of mate pair libraries to bridge the large gaps.
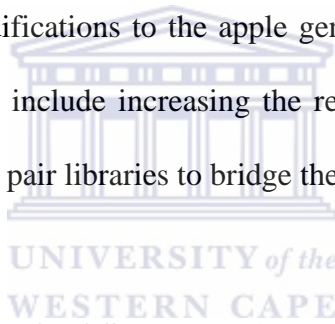
1.2.5 Applications of NGS technologies

1.2.5.1 RNA analysis – RNA-seq

RNA-seq is a technique for analysing RNA species using deep-sequencing approaches (Wang*, et al.*, 2009). The RNA-seq technique has all the advantages of the NGS approaches and, unlike microarray tiling and Sanger sequence based techniques, RNA-seq is high throughput and has single nucleotide resolution of transcript structure, splicing patterns, relative expression and sequence variation (McIntyre*, et al.*, 2011; Minoche*, et al.*, 2011; Wang*, et al.*, 2009). Transcript coordinates with respect to start and end site, as well as the number and length of the exons constituting the transcript determines the structure of the transcript in RNA-seq. The variations in the number and lengths of exons constituting a transcript result in different species of a transcript being transcribed. These

different species of a transcript are referred to as splice variants, and have varying spatial and temporal effects on physiological processes. Simply put, the use of splice variants is one of an organism's ways of making optimum use of changing conditions. The use of microarray tiling and qPCR in estimating transcript abundances in given samples requires careful selection of "ideal" reference genes. The high background noise in microarray tiling and qPCR techniques reduces their sensitivity to quantifying RNA species. As a result of high background noise, microarray tiling and qPCR technique are less sensitive to lowly and highly expressed transcripts (McIntyre, *et al.*, 2011).

1.2.5.2 RNA-Seq basics and experimental design

Setting-up an RNA-Seq experiment is crucial to the interpretation of the data. An ideal RNA-Seq experiment is designed such that the measurements taken reflect as much as possible differences due to treatment. To reduce unexplained variability in the measurements of treatment effects, technical and biological replications as well as positive and negative controls are included in the design. Biological replicates consist of different biological samples that are processed through the sequencing process separately. Technical replicates are replicates where biological material is the same in each replicate but the technical steps used to measure gene expression are performed separately(McIntyre, *et al.*, 2011).

In a basic protocol for RNA-Seq, RNA is isolated from the cell, a library containing the isolated RNA in a form ready for sequencing is prepared and the RNA is sequenced. Each protocol varies in detail depending on the technology. Different RNA species may be targeted, for example polyadenylated RNA, micro-RNAs and total RNA(McIntyre, *et*

al., 2011). The isolated RNAs may contain contaminants, which may be reduced by removing ribosomal (rRNAs) using Ribo-zero (Epicentre), DNAse treatment to remove genomic DNA, and size selection to target small RNAs before library preparation. The RNA may be copied to cDNA and may or may not be amplified via PCR. Sequencing may be single end, producing one read, or paired end, producing two reads, which may be separated by a region which is not sequenced if the reads do not overlap(McIntyre*, et al.*, 2011).

1.2.5.3 Analysing and interpreting RNA-seq data

The increasing read throughput from NGS, as well as the number of ESTs databases in public databases has been accompanied by an expansion of bioinformatics tools for analysis of sequence datasets, at cDNA and protein levels. As a result, a number of web-based programs and/or stand-alone pipelines have been developed.

Raw sequence reads from NGS platforms are screened for PCR duplicates, contaminants, adaptor sequences, ambiguous bases, and low base quality (in PHRED score at a set percentage of the read length) (McIntyre*, et al.*, 2011). The quality filtered reads are then assembled into contiguous sequences based on similarity using *de novo* or reference assisted assemblers such as Velvet (Zerbino and Birney, 2008) and Bowtie (Langmead*, et al.*, 2009) respectively.

Considering the transcripts were fragmented before sequencing, the goal for assembly would be to attempt to reproduce the original gene sequence, both in terms of length and

base organisation. Thus the short reads are aligned and merged together into longer contiguous sequences (contigs) (Nagaraj, *et al.*, 2007; Ranganathan, *et al.*, 2009). The 454 platform produces long reads, while Illumina and SOLiD platforms produces shorter reads, which are assembled using the algorithm 'overlap-layout-consensus' (Myers, 1995) and 'de Bruijn graph' (Idury and Waterman, 1995; Miller, *et al.*, 2010; Zerbino and Birney, 2008), respectively. Some of the assembly algorithms are sequencing platform specific, which limit their usefulness across platforms.

In an overlap-layout-consensus algorithm, the overlaps between reads are used to construct an overlap graph, which is used to compute a layout of reads and consensus of contigs by pair wise alignments. This approach is good for sequences with limited number of reads but significant overlaps. Examples of such algorithms are Celera assembler and CAP, and are computationally intensive for short reads (Zerbino and Birney, 2008). In the 'de Bruijn graph' algorithm, read sequences are fragmented into short segments of a given length, 'k-mer', where 'k' is the number of nucleotides in each segment. Overlaps between and among k-mers are captured and stored in graphs, which are subsequently used to generate consensus sequences. Most genomes have varying numbers and lengths of repeat regions difficult to resolve using single-end data. Reads generated from either end of a cDNA (pair-end) or longer cDNA fragments better resolve long repeated DNA regions. Programs such as Velvet (Zerbino and Birney, 2008) and Trinity (Grabherr, *et al.*, 2011), among others, are *de novo* assemblers of short reads. Different short read assemblers have shown varying efficiencies in assembling short reads of either genomic or transcriptomic origin (Grabherr, *et al.*, 2011; Schulz, *et al.*,

2012; Trapnell*, et al.*, 2010). As such special programs have been selected and developed to address the demands of either genomic or transcriptomic (RNA-seq) analysis.

In eukaryotic organisms, pre-mRNA, consisting of exons and introns, is edited into mature mRNA, without introns, before translation to protein. Pre-mRNA is transitory but the majority of the mRNA is mature, which is fragmented and sequenced. The re-assembly, therefore, of the cDNA sequences could be done by *de novo* assemblers if the objective would be consensus full-length transcripts (Grabherr*, et al.*, 2011; Schulz*, et al.*, 2012). However, if the aim is to annotate the genome or identify the different forms of the genes and their location within the genome and to calculate gene expression levels, reads are mapped to a reference genome before assembly (Trapnell*, et al.*, 2012; Trapnell*, et al.*, 2010). Thus, new programs, such as Bowtie, were developed to map the short reads to the genome. However, mapping cDNA reads to the genome only disregards all read sequences overlapping into the intronic region resulting in false assemblies and subsequent calculated abundances. To address this anomaly, splice junction mappers, such as Tophat (Trapnell*, et al.*, 2009), have been developed, which split the unmapped reads into small segments, which are mapped across the junctions. The resulting read mapping by a genome mapper, such as Bowtie, followed by a splice junction mapper refines genome/gene annotation and a larger percentage of the sequenced reads are mapped.

The mapped reads, in the Sequence alignment/map (SAM)/ Binary alignment/map (BAM) format, are assembled into transcripts. The number of the non-duplicate reads

within a defined coordinate region, marking the start and end of a transcript, are counted. These are presented as normalized transcript abundance, in fragments per kilo base per million (FPKM) (Trapnell*, et al.*, 2009). FPKM values of each transcript are normalized to reduce unexplained variability by taking into account the mapped total read mass, transcript length, highly or lowly expressed transcripts above the median (Trapnell*, et al.*, 2012; Trapnell*, et al.*, 2010).

In time-series gene expression studies, the goal is to identify genes whose expression shows a statistically significant change during the period of study, which are calculated using statistical packages. However, most programs cannot be used to process the large volume of data from RNA-seq. Some programs, such as the R-package (R core team, 2013), which is based on S-plus programming language and Cuffdiff (Trapnell*, et al.*, 2010) are extensively used and continue to be improved. In addition, some programs initially designed for microarray technology data handling, such as Multiple experiment viewer (MeV) have been modified to accommodate NGS assembly and gene expression manipulation (Ermolaeva*, et al.*, 1998; Saeed*, et al.*, 2003).

Attainment of expression values, in itself, is not an end as there is need to extract and identify the transcripts. Programs such as Picard (Byun*, et al.*, 2010) or scripts that make use of the transcript coordinates on the genome with respect to the terminal ends, extract the consensus sequences between these coordinates into contigs. The sequences of the transcripts are then compared to known sequence data available in public databases. Such a comparison is done to assign a predicated identity to each query nucleotide sequence if

significant match is found (Nagaraj, *et al.*, 2007). The nucleotide query sequence may be translated into predicted proteins, using algorithms that identify protein-coding regions (ORFs) for each transcript; for example ESTScan (Iseli, *et al.*, 1999) and OrfPredictor (Min, *et al.*, 2005), allowing for amino acid sequence comparisons with publicly available databases. Such comparisons allow for inference of protein domains/motifs using programs such as Interproscan (Hunter, *et al.*, 2009). They also allow comparison of function with information available in the databases such as PROSITE (Hofmann, *et al.*, 1999), PRINTS (Attwood, *et al.*, 2003), Pfam (Bateman, *et al.*, 2004), ProDom (Corpet, *et al.*, 2000), SAMRT (Schultz, *et al.*, 2000) and Gene Ontology (GO; (Ashburner, *et al.*, 2000). Genome and gene annotation is a continuous process, reviewed each time new information on a gene in obtained.

A significant number of query transcripts may have identity hits to uncharacterized ESTs or fail to match annotated transcripts in all available public databases. Such, 'no-hit', transcripts may be regarded as novel and together with the uncharacterized ESTs need to be defined in function and relevance. Only when the function of each transcript has been confirmed, can the transcriptome be fully elucidated. Nevertheless, regulation and networking of the genes in tissues will require further information on small-RNAs involvement.

1.2.5.4 Sources of variability in RNA-seq data analysis

In depth understanding of the possible sources of variability is invaluable in RNA-seq analysis studies. In any experiment there is inherent uncertainty present in any measurement made where something is sampled and counted. This inherent variability is

based on the value of the count itself, is called Poisson variance, and is not experiment specific. The lower the count the higher the level of uncertainty for that measure. For example, gene expression levels that are supported by low read count are less likely to be true expression levels compared with genes supported by high read count. Because of this it is more difficult to detect small fold changes in genes with low read counts than in those with high read counts (McIntyre*, et al.*, 2011).

The sample handling process before, during and after sequencing introduces variation in the RNA species count as well. Thus different protocols, mixing techniques, flow-cells, and flow-cell lanes affect the error rates and reproducibility of the sequenced reads. This source of variability is called non-Poisson technical variance, and is experiment and protocol-specific and may affect different classes of genes differently (McCarthy*, et al.*, 2012). However, errors are more pronounced when very low sampling fractions (<0.00013%) are used, especially exons with less than five reads per nucleotide coverage(McIntyre*, et al.*, 2011). As such an increase in reads that can be mapped increases the number of exons detected, with substantial reduction in error rates.

Biological variance is due to the differences in the genetic make-up of organisms, tissues, and sometimes cells and environmental effects on gene expression provides an inconsistent count of RNA species in a biological sample at any given time (Minoche*, et al.*, 2011).

Minoche *et al.*, (2011) noted that there is still a lot of work needed to substantially reduce errors in RNA-seq (Minoche*, et al.*, 2011). Depending on the sample size, including both technical and biological replicates in the design may not be financially feasible (Minoche*, et al.*, 2011). Several recommendations have been put forward and serve as standards for RNA-seq analysis (Minoche*, et al.*, 2011). On Illumina platforms, it is recommended to include phiX genome spike-in for each lane as a control (Zook*, et al.*, 2012). The phiX spike-in across the lanes is used to ensure that the data generated for each run is accurately calibrated during imaging and data analysis. Multiplexing large samples into single lanes has also been proposed to remove the possible technical effects of different lanes (McIntyre*, et al.*, 2011). Views on the amount of validation required for RNA-Seq experiments differ. While it is certainly true that some runs fail, the technical reproducibility of RNA-Seq experiments suggests that RNA-Seq technical validation is may not be necessary once the sequencing instrument has been optimised(Zook*, et al.*, 2012).

1.3 Rationale

The development and maturation of apple fruit is transcriptionally controlled. Each fruit section has defined functions, which contribute to the total physiological processes governing growth and development of the fruit. The fruit section transcriptomes are a subset of the whole fruit transcriptome. Genes whose transcripts are unique to a fruit section define specific roles and functions of the fruit section. The identified gene transcripts will be used to predict proteins and such information will contribute to scientific data needed to build the proteome database for the apple fruit. It will also provide an understanding of genes/proteins expressed in response to internal and external

factors, mainly during fruit development, as well as the contribution of each fruit section to the total fruit transcriptome. Finally, genes identified from the transcriptome profile could be utilised to identify/design molecular markers required for high-resolution mapping around QTLs of economic importance.

A fully characterized apple fruit transcriptome will reveal the possible transcripts and networks that can be expressed in a particular organism. Secondly, the spatial and temporal expression patterns of all transcripts that change in levels of expression during fruit development will be established. Thirdly, the mechanisms by which the expression patterns are regulated will be elucidated. The challenges in attaining these goals include covering the complete cataloguing of transcripts, dynamic transcription profiling and understanding the transcription regulatory network (Ruan*, et al.*, 2004). With the advent of high throughput sequencing technology, such as the Illumina HiScan^{TM}SQ, a higher coverage and more representative transcriptome is obtainable. This enables a comprehensive molecular description of the biological processes occurring during apple fruit development. The information on apple fruit transcriptome is useful for understanding fruit development in apples, but also in other fruit crops (Lee*, et al.*, 2007)

The apple fruit is a unique aspect of plant development and the source of human diet. For these reasons, it is important to understand the molecular mechanisms underlying apple fruit development and ripening towards unravelling the complex interactions leading to fruit quality. A number of previous studies have partially elucidated fruit cortex, whole fruit, leaves, and peel plus cortex transcriptomes using cDNA micro arrays and other

PCR based technologies (Han, Bendik, *et al.*, 2007; Hattasch, *et al.*, 2008; He, *et al.*, 2008; Janssen, *et al.*, 2008; Mintz-Oron, *et al.*, 2008; Soglio, *et al.*, 2009). These studies, however, remain limited in capacity and specificity to revealing transcripts of low abundance. Hence, the need for high throughput techniques to elucidating the full transcriptome of the fruit peels and pulp tissues.

1.4 Aims and Objectives

Accordingly, the aim of this study was to elucidate regulatory patterns associated with fruit development in the apple cultivar 'Golden Delicious'.

To achieve this aim, the objectives below were formulated:

- To generate mRNA libraries of apple fruit peel and pulp tissues across six developmental time points from 25 days after anthesis (daa) to full ripeness (150daa);

- To sequence and reconstruct full-length transcripts;

- To annotate the reconstructed transcripts;

- To comparatively analyse transcripts differentially regulated in pulp and peel tissues;

- To comparatively analyse transcripts differentially regulated during fruit development.

2. CHAPTER 2

2.1. GENERAL MATERIALS

2.1.1. List of chemicals and Reagents

Chemical/Reagent	Company
Agarose D1 LE	Promega
ß-mercaptoethanol	Fermentas
Boric acid	Merck
Bromophenol blue	Sigma
Chloroform	BDH chemicals
Dithiothreitol (DTT)	Fermentas
DNase 1	Ambion
Ethanol	BDH chemicals
Ethidium bromide	Sigma
Ethylenediamine tetra acetic Acid (EDTA)	Merck
Formaldehyde solution	Riedel-de Haën
Formamide	Merck
Glacial acetic acid	Merck
Glycerol	Merck
Hexadecyltrimethylammoniumbromide (CTAB)	Merck
Hydrochloric acid (HCl)	Merck
Iso–propyl alcohol	BDH chemicals
Kapa SYBER Green kit	Kapa Biosystems
Lithium chloride	Merck
NucleoSpin-RNA clean-up kit	Epicentre
Poly vinyl pyrolidone K 40	Merck
Primer oligos	Integrated DNA technologies (IDT)
QIAquick PCR purification kit	QIAGEN
Qubit reagents	Invitrogen

Sodium acetate	Merck
Sodium chloride (NaCl)	Merck
Sodium dodecyl sulphate (SDS)	Bio-Rad
Sodium hydroxide (NaOH)	BDH chemicals
Sodium sulphite	Merck
Tris (hydromethyl) aminomethane (Tris)	Merck
Truseq-Kits	Illumina
Tween 20	Merck

2.1.2. Buffers and solutions

CHISAM:	96% Chloroform (v/v), 4% Isoamyl alcohol (v/v)
CTAB extraction buffer:	2% w/v CTAB (hexadecyltrimethylammoniumbromide)
	2% w/v PVP (polyvinylpyrrolidone K 40; w/v)
	100 mM Tris-HCl, pH 8.0
	25 mM EDTA
	2 M NaCl
	0.06% Sodium sulphite (w/v)
	2.64% ß-mercaptoethanol (v/v)
DNA loading buffer	0.25% (w/v) bromophenol blue, 0.25% (w/v)
	xylene cyanol, 30% (v/v) glycerol
EDTA:	0.5 M Na$_2$EDTA-2H$_2$O in RNase-free H$_2$O,
	pH 8.0
70% Ethanol (EtOH)	70% absolute EtOH in RNase-free H$_2$O (v/v)
10 x FA gel buffer:	200 mM 3-[N-morpholino] propanesulfonic acid (MOPS)
	50 mM Sodium acetate
	10 mM EDTA, pH to 7.0
1 x FA gel running buffer:	10% v/v 10 x FA gel buffer
	0.74% formaldehyde (v/v)
LiCl:	10 M LiCl.

2X RNase Buffer	20 mM Tris-HCl, 10 mM EDTA, 0.6 M NaCl (pH 8.0), dissolved in DEPC treated H_2O
RNase-free water	1% v/v Diethylpyrocarbonate (DEPC) in distilled H_2O (dH_2O)
5 x RNA loading buffer:	0.27% bromophenol blue solution (v/v) 6.8 mM EDTA, pH 8.0 4.5% formaldehyde (v/v) 34% glycerol (v/v) 52.2% formamide (v/v) 0.07% 10 x FA gel buffer (v/v)
10% SDS solution	10% (w/v) SDS in dH_2O
Sodium Acetate	3 M NaOAc.$3H_2O$, pH to 5.2 with glacial acetic acid
SSTE:	1 M NaCl 0.5% SDS (w/v) 10 mM Tris-HCl, pH 8.0 1 mM EDTA
10X TBE	0.9 M Tris, 0.89 M boric acid, 0.032 M EDTA, pH 8.3
10X TE	100 mM Tris-HCl (pH 7.5), 10 mM EDTA, pH 7.5
Tris base	1.5 M Tris, pH 8.8
Tris-HCl:	1 M Tris-HCl, pH 8.0

NB: All chemicals and reagents used were analytical grade

Table 2. 1 List of software used for bioinformatics analysis

Software	Version	URL/Reference
BEDTools	2.16.2	http://code.google.com/p/bedtools/(Quinlan and Hall, 2010)
Blast2GO	2.6.0	http://www.Blast2GO.com/b2glaunch/start-Blast2GO(Conesa, *et al.*, 2005)
Blastall	2.2.16	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/(Altschul, *et al.*, 1990)
Bowtie	0.12.7	http://bowtie-bio.sourceforge.net/index.shtml/(Langmead, *et al.*, 2009)
Cdfasta	0.99	http://sourceforge.net/projects/cdbfasta/
CLC-bio	5.5.1	www.clcbio.com/
Cuffdiff	2.0.2	http://cufflinks.cbcb.umd.edu(Trapnell, *et al.*, 2012)
Cufflinks	2.0.2	http://cufflinks.cbcb.umd.edu(Trapnell, *et al.*, 2012)
Cutadapt	1.0	http://journal.embnet.org/index.php/embnetjournal/article/view/200(Martin, 2011)
Edge R	1.6.15	http://bioc.ism.ac.jp/2.6/bioc/html/edgeR.html ((Robinson, *et al.*, 2010)
ESTScan	3.0.3	http://sourceforge.net/projects/estscan/(Iseli, *et al.*, 1999)
FASTQC	0.5.0	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/(Andrews, 2010)
FASTX-Toolkit	0.0.13	http://hannonlab.cshl.edu/fastx_toolkit/download.html(Gordon, 2011)
Genome Vx		http://wolfe.gen.tcd.ie/GenomeVx/(Conant and Wolfe, 2008)
Ggplot2	0.9.1	http://cran.r-project.org/web/packages/ggplot2/(Wickham, 2009)
GMAP	2012-11-7	http://research-pub.gene.com/gmap/src/(Wu and Watanabe, 2005)
IGV	2.1	http://www.broadinstitute.org/igv/(Robinson, *et al.*, 2011)
Interproscan	2012	http://www.ebi.ac.uk/Tools/pfa/iprscan/(Hunter, *et al.*, 2009)
KEGG	1-10-12	http://www.genome.jp/kegg/pathway.html(Kanehisa, *et al.*, 2004)
KOBAS	2.0	http://kobas.cbi.pku.edu.cn/home.do(Xie, *et al.*, 2011)
MeV4.7.1	V10.2	http://www.tm4.org/mev/(Saeed, *et al.*, 2003)
Perl	5.10.0	http://www.perl.org/(Wall, *et al.*, 2000)
Picard	1.61	http://sourceforge.net/projects/picard/(Byun, *et al.*, 2010)
Primer-blast	3.1	http://www.ncbi.nlm.nih.gov/tools/primer-blast/
Python	2.6.1	http://www.python.org/download/releases/2.6.1/(Van Rossum and Drake, 2009)
R package	3.1.1	http://www.R-progect.org/ (R Core Team, 2013)
Rotor Gene	2.0.4	http://software.informer.com/discovered/Rotor_Gene_Q_Software_Downloa

Q		d_Mac	
RSEM	1.2.3	http://trinityrnaseq.sourceforge.net/	(Li and Dewey, 2011)
Samtools	0.1.18	http://samtools.sourceforge.net	(Li, *et al.*, 2009)
StatPlus	2009	http://www.analystsoft.com	(Analyst Soft Inc)
Tophat	1.4.1	http://tophat.cbcb.umd.edu/	(Trapnell, *et al.*, 2012)
Trinity	12-10-05	http://trinityrnaseq.sourceforge.net/	(Grabherr, *et al.*, 2011)

CLC-Bio = CLC Bio Genomics Workbench, genomic map alignment program (GMAP), Multi experiment viewer (MeV), KOBAS (KEGG *Orthology* Based Annotation System), KEGG=Kyoto Encyclopaedia of genes and genomes

Table 2. 2 List of databases used for transcript annotation

Database	URL
Grape, March 2010	www.phytozome.net/grape.php
M. x domestica chloroplast DNA, v1.0	http://www.rosaceae.org/projects/apple_genome,
M. x domestica ESTs	ftp://ftp.ncbi.nih.gov.dbEST/
M. x domestica mitochondria DNA, v1.0	http://www.rosaceae.org/projects/apple_genome
M. x domestica v1.0 contigs	http://www.rosaceae.org/projects/apple_genome
M. x domestica v1.0.gff3	http://www.rosaceae.org/node/482
Malus CDS, v1.0	http://www.rosaceae.org/node/482
Malus plant unique transcripts	www.plantgdb.org/ cited 12/09/3013
Non-redundant protein database (NR)	http://www.ncbi.nlm.nih.gov/ cited 08/08/2013
Peach , JGI release 1.0	www.phytozome.net/peach.php
Pfam	ftp://ftp.sanger.ac.uk/pub/databases/Pfam/
Populus trichocarpa v3.0	www.phytozome.net/poplar
Rice, MSU release 7	www.phytozome.com/rice.php
swiss-prot	www.ebi.ac.uk/**swissprot**/
TAIR, release 10	http://www.arabidopsis.org/

2.2 GENERAL METHODS

2.2.1 Plant material

Apple (*M. x domestica* Borkh.) cultivar 'Golden Delicious' was grown on M9 rootstocks and managed according to standard orchard practices. Fruit samples were harvested at eight time points: 0; 15; 25; 35; 60; 87; 135; and 150 days after anthesis (DAA). However, owing to the difficulty of separating core from the rest of the fruit tissue in samples collected at day zero and 15, only samples collected afterwards were used for total RNA extraction and subsequent experiments.

Ten fruit samples were harvested at about 10 am at 25, 35, 60, 87, 135, and 150 DAA intervals to reduce unexplained variations in gene expression. Immediately after harvesting fruits were packed into ice containing cooler boxes and transported for one hr to the laboratory. Fruits were then immediately rinsed with RNase-free water to remove dust and surface pathogens as well as reducing fruit field heat. Ten fruits per interval were then sectioned into peel and pulp using RNase-free scalpel blades and snap-frozen in liquid nitrogen. A gram of each fruit section pooled from ten fruits was then transferred into pre-cooled sterile 1.5 ml centrifuge-tubes labelled and stored at -80°C.

2.2.2 Apple fruit total soluble solids content and apple fruit weight measurement

The Brix˚ (total soluble solids) was assayed from juice squeezed from harvested apple fruits with a refractometer (BRIX30, Leica, Bannockburn, IL). Pome fresh weight (g) was measured with a precision balance PS4500 (Radwag).

2.2.3 Total RNA extraction from apple fruit peel and pulp sections

Total RNA was extracted from one g of pulp or peel tissue from each time point ground using pestle and mortar under liquid N_2 conditions using a modified method of Chang *et al*., (Chang*, et al.*, 1993). Sterile powdered glass was added to the mortar to aid in grinding plant tissue. Powdered pulp (1 g) was transferred to 50 ml tubes containing 15 ml CTAB extraction buffer, with 2.64% ß-mercaptoethanol (v/v), pre-warmed to 65°C and mixed. Fruit cells were broken by incubating at 65°C for 10 min then phase separated twice with 15 ml CHISAM (Eppendorf 5810 R, 2655 g) for 20 min at room temperature. The supernatant was transferred to a fresh tube and LiCl added to a final concentration of 2 M then incubated for 16 hours at -20°C to precipitate total RNA. Precipitated total RNA was recovered as a pellet by centrifugation for 30 min at 4°C (Eppendorf 5810 R, 10621 g). The supernatant was removed and the pellet completely re-suspended in 60°C warmed 700 µl SSTE for 5 min. After mixing the solution was transferred to RNase-free 2 ml micro-centrifuge tubes before phase separating twice with 700 µl CHISAM centrifuged (Spectrafuge 24D, 9200 g) for 10 min at room temperature. The upper phase was transferred to a new RNase-free 2 ml micro-centrifuge tube and RNA precipitated by adding 2 x volume absolute ethanol and incubated at -80°C for 2 hours. The precipitated RNA was recovered by centrifuging (Spectrafuge 24D, 9200 g) for 30 min at 4°C, and then washed twice with 70 % ethanol with centrifugation (Spectrafuge 24D, 9200 g) for 2 min then bench air-dried and re-suspended in 50 µl RNase-free H_2O.

2.2.4 Removal of impurities from Total RNA

Isolated total RNA was DNase treated with 2 U/µl DNase in 10% volume 10x DNase 1 buffer (Ambion DNA-free kit, catalogue number 1906) incubated for 30 min in a 37°C water bath and inactivated at room temperature for 2 min with 10% volume DNase inactivation reagent (Ambion DNA-free kit, catalogue number 1906). DNase protein as well as other possible contaminates from polysaccharides and polyphenol contaminants were then separated from total RNA using the NucleoSpin-RNA clean-up kit (Epicentre) according the manufacturers' instructions prior to library preparation and RT-qPCR.

2.2.5 RNA quantity and integrity

2.2.5.1 Fluorometric RNA quantisation

The concentration of the RNA was determined using Qubit 2.0 (Invitrogen) fluorometer according to the manufacturer's instructions. Briefly, a master mix of the Qubit® RNA broad range buffer and reagent was prepared at a ratio of 199 µl to 1 µl respectively. The master mix was then dispensed into 500 µl micro-centrifuge tubes, 198 µl and 190 µl for cDNA sample tubes and standards respectively. To the 198 µl master mix, 2 µl of cDNA sample was added, while 10 µl of standard was added to the tubes containing 190 µl master mix. The samples were mixed and incubated for min at room temperature. The Qubit® fluorometer was calibrated using the standards before quantifying the cDNA concentrations in the libraries.

2.2.5.2 RNA resolution by electrophoresis

2.2.5.2.1 RNA resolution on Bioanalyser

The integrity and quality of the DNase treated total RNA were evaluated using an Agilent 2100 Bio-analyser and a RNA 6000 Nano Assay kit (Agilent Technologies, http://www.agilent.com/) according to the manufacturer's instructions. For each sample total RNA concentration of 200-250 ng and RNA 6000 ladder (Ambion, http://www.Ambion.com) were loaded. All samples and ladder were denatured at 70˚C for 2 mins before electrophoresis. The results were analysed using Agilent 2100 Expert software v1.3.

2.2.5.2.2 RNA resolution on denaturing agarose gel

Denaturing gels were prepared by dissolving 1.2 g agarose in 100 ml 1 x FA gel buffer. Agarose was melted by heating, and then cooling to 65˚C before adding formaldehyde to a final concentration of 0.7% and staining with ethidium bromide at a final concentration of 1 mg/10 ml. The cast gels were equilibrated in 1 x FA gel running buffer at 7 V/cm for 20 min. RNA samples were prepared by adding 5 x loading buffer to a final concentration of 25% of RNA sample. The sample was then incubated at 65˚C for 5 min and chilled on ice before loading onto the equilibrated denaturing gel. RNA species were separated at 7 V/cm of gel for 60 mins, and documented using a Universal Hood II (Bio-Rad, USA), and the quality of the extracted total RNA determined.

2.2.6 Sample preparation for paired- end sequencing

Libraries for sequencing on the HiScan^{TM} SQ(Illumina) were prepared according to the Illumina TruSeq RNA sample preparation version 2 kit protocol (catalogue number RS-

930-2002, Part number 15015050 Rev.A, November 2010, Illumina). Total RNA was diluted to a working concentration of 3 µg/50µl with nuclease-free water, to which 50µl RNA purification buffer was added and mixed. The RNA purification buffer contained oligo-dT beads. RNA secondary structures were then disrupted by heating the RNA samples at 65 $^{\circ}$C for 5 min before cooling on ice, to allow for binding of the magnetic oligo-dT beads to the poly-A mRNA tails. Magnetic stands were used to pellet the bead mRNA complexes, allowing for washing and removal of supernatant with non-mRNA species.

Messenger RNA molecules were chemically fragmented using the Elute, Prime and Fragmentation (EPF), which contained random hexamers for reverse transcriptase (RT) priming and serves as 1st strand cDNA synthesis reaction buffer. First strand cDNA was then synthesized from the fragmented mRNA using random hexamers, allowing for uniform priming and amplification of the entire length of the mRNA molecules. The first strand cDNA molecules were then used as templates for second strand cDNA synthesis. AMPure XP beads which bind to double stranded cDNA molecules were used to retain the double stranded cDNA to allow for cleaning up and pellet on the magnetic stand. Overhangs at the ends of the double stranded cDNA molecules were repaired of allow for blunt-end ligation of single 3′-A residues in preparation for adapter ligation. In place of the 5´-adapters an index adapter was used instead, which allowed for sample multiplexing during sequencing. Ten PCR cycles were used to enrich the cDNA library. The enriched library was purified of PCR contaminants using the QIAquick PCR purification kit (QIAGEN, www.qiagen.com) according to the manufacturers' instructions. Tween 20

was added to the purified library to a final concentration of 0.6%, before validating the library size.

2.2.7 Library validation

2.2.7.1 DNA agarose gel electrophoresis

A 2% agarose gel was prepared by melting 2 g agarose in 1 x TBE (100 ml). The dissolved agarose gel was then stained with ethidium bromide at a final concentration of 1 mg/10 ml. The cast gels were equilibrated in 1 x TBE gel running buffer at 7 V/cm for 20 min. Library samples were prepared by adding DNA loading buffer to a final concentration of 25% of the cDNA library sample. The samples as well as the GeneRuler 100 bp DNA ladder (Fermentas) were then loaded onto the equilibrated normal 2% agarose gel. The library species were then size separated at 7 V/cm of gel, and documented using a Universal Hood II (Bio-Rad, USA).

2.2.7.2 Qubit® ds DNA assay

The concentrations of the cDNA libraries were determined using a Qubit 2.0 (Invitrogen) fluorometer according to the manufacturer's instructions. Briefly, a master mix of the Qubit® ds High Sensitivity (HS) buffer and reagent was prepared at a ratio of 199 µl to one µl respectively. The master mix was then dispensed into 500 µl micro-centrifuge tubes, 198 µl and 190 µl for cDNA sample tubes and DNA standards respectively. To the 198 µl master mix, 2 µl of cDNA sample was added, while 10 µl of DNA standard was added to the tubes containing 190 µl master mix. The samples were then mixed and

incubated for a min at room temperature. The Qubit® fluorometer was calibrated using the DNA standards before quantifying the cDNA concentrations in the libraries.

2.2.8   Quality processing of sequenced reads

Sequencing of the libraries was performed using the HiScan[TM]SQ (Illumina) at the Biotechnology Platform ARC, South Africa, as non-strand specific paired-end reads 100 bases long. Read quality was assessed using FASTQC and bases less than 20 Phred score along 95% of the read length were filtered out using FASTX-Toolkit (Gordon, 2011). Possible Illumina adapter contaminations were then removed by clipping off the adapters using Cutadapt v1.0 (Martin, 2011) retaining read sequences of at least 96 bases long (Box 1).
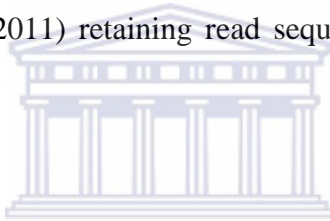
Commands executed from the UNIX shell (bash) are prefixed with a '$' character, while comments/description for each command are prefixed '#'.

<div style="border:1px solid #000; background:#c0f0c0; padding:8px;">

Box 1

#!/bin/sh

Created by Zedias Chikwambi on 2012/10/15.

Copyright 2012 __ARC-biotechnology Platform_. All rights reserved.

#Trimming terminal low quality bases

$ fastx_trimmer -f 2 -l 97 -i read_1/2.fastq -v -Q33 | fastq_quality_filter -q 20 -p 95 -v -Q33 -z -o read_1/2trm.fastq.gz

For more information refer to FASTX-Tool kit manual (version 0.013)

#Clip Truseq-Illumina adapter sequences from the 3´end

$ cutadapt-e 0.02 -O 5 -m 96-g AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC read_1.fastq > read_1_clipped.fastq

#Read 2

</div>

```
$ cutadapt -e 0.02 -O 5 -m 96 -g AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
read_2.fastq > read_2_clipped.fastq
# Adapter sequences were removed from the 3´terminal end of the reads. For more information
refer to Cutadapt manual (v1.0).
# On the terminal a software manual can be obtained by typing: $ cutadapt -h
```

The quality of the reads was then re-checked using FASTQC (Andrews, 2010). Reads that passed the quality filtering where then used for transcript reconstruction.

2.2.9   Full-length transcript assembly and abundance estimation

Reconstruction of tentative full-length transcripts was built around the genome-guided tuxedo protocol as well as the *de novo* transcript assembly (Grabherr*, et al.*, 2011; Trapnell*, et al.*, 2012). The reference-guided assemblies were divided into, read mapping and transcript assembly without gene models (*Ab initio*, Ab) and read mapping and transcript assembly with the assistance of apple genome v1.0 gene models (Gm) using Tophat-Cufflinks pipeline (Li*, et al.*, 2010; Trapnell*, et al.*, 2009; Trapnell*, et al.*, 2012).

2.2.9.1 Genome-guided transcript reconstruction

Filtered reads from section 2.2.7 box 1, were mapped to the reference genome using Tophat, as shown in Box 2. The Bowtie software was used to build an index of the apple genome v1.0 pseudo-haploid molecules to allow quick access during read mapping. The paired end reads were not size selected and had an effective insert size of 25 to 40 bases. The reads were mapped to the reference initially using bowtie, then across splice junctions using Tophat with or without the assistance of a gene model (Box 2).

The aligned reads were reconstructed into transcripts using Cufflinks with assistance of apple pseudo haploid molecules mapped gene models (Box 3). The reconstructed gene models from each of the six time points and two tissue types were merged using Cuffmerge to reconstruct a reference gene model. Aligned reads were used to construct a consensus genome excluding any sequence regions on the pseudo molecules not covered by reads. The reconstructed transcript gene models were then used to extract the respective transcripts from the consensus genome in fasta format (Box 3).

```
- | samtools-0.1.18/bcftools/vcfutils.pl vcf2fq > All_mapped_accepted_hits_Ab3.fastq
```
#Convert fastq to fasta using fastq_to_fasta in FASTX Toolkit 0.0.13
```
$ fastq_to_fasta -i All_mapped_accepted_hits_Ab3.fastq –o
All_mapped_accepted_hits_Ab3.fasta
```
#Now extract transcript sequences from the pile-up consensus. The 'gffread' program which comes with cufflinks package is used for fasta extraction
```
$ gffread -w All_mapped_accepted_hits_mapped_transcripts_Ab3.fa -g
All_mapped_accepted_hits_mapped_Ab3.fasta peel_pulp_merged_pseudoAb3.combined.gtf
# use the gtf file to denote the regions corresponding to the transcript exons
# At this stage transcript sequences, exons corresponding to an mRNA are extracted and merged
together in their chronological order.
```

2.2.9.2 *De novo* transcript reconstruction

Ab initio (Ab) and gene models assisted (Gm) apple genome pseudo haploid molecules unmapped reads were pooled across the time points and tissues into one data set for each Ab or Gm approach. A third data set consisting of all reads from the six time points and two tissues (peel and pulp) where pooled into another data set. The three data sets were then *de novo* assembled into transcripts using Trinity (Box 4)(Haas*, et al.*, 2013). Transcripts with more than 95% identity were collapsed, while only transcripts longer than 300 bases were retained.

Box 4
#*De novo* transcript assembly
```
#Increase the stack
# Set stack size to maximum. Mac OS have a maximum stack size of 65532
$ ulimit -s 65532
$ perl trinityrnaseq_r2012-06-08/Trinity.pl --seqType fq --JM 90G --min_kmer_cov 2 --
inchworm_cpu 2 --left left_Ab_unmap.fq --right right_Ab_unmap.fq --output
trinity_output_tissue_Ab_unmapped --bfly_opts "--stderr --REDUCE --max_final_diffs=10"
```

```
# For more information the reader is referred to Trinity manual version 2012-06-08
#Assembly transcripts statistics
# This script calculates the count of each base including ambiguous bases (N), and reports out
the count of each base, length of transcript, % number of proper bases per transcript length,
and %GC content
$ perl fasta_sequence_stat.pl All_mapped_accepted_hits_mapped_transcripts_Ab3.fa >
All_mapped_accepted_hits_mapped_transcripts_Ab3_stat.txt
# Select against transcripts with more than 50% N content, in column 8 (this should remove
transcripts with %N content greater than 50%). Also remove sequences less than 300 bases in
length
$ awk '$8>=50' transcripts_Ab3_stat.txt | awk '$2>=300' > transcripts_Ab3_stat_less_50.txt
# Print the first column with sequence IDs of transcripts with less than 50% N content and
greater than 300 bases in length
$ awk '{print $1}' transcripts_Ab3_stat_less_50.txt | sort | uniq >
transcripts_Ab3_stat_less_50_Id_list.txt
#Extract the fasta file corresponding to the sorted ID list using "split_contigs.py"
$ python split_contigs.py transcripts_Ab3.fasta transcripts_Ab3_stat_less_50_Id_list.txt
transcripts_Ab3_stat_more_50.fasta transcripts_Ab3_stat_less_50.fasta
```

2.2.9.3 Creating gene models from *de novo* reconstructed transcripts

The pooled Trinity transcripts (PTTs) were mapped to the pseudo molecules using
GMAP creating gene models for the mapped PTTs (Wu and Watanabe, 2005).
Transcripts reconstructed from Ab unmapped (Ab_unTTs) and Gm unmapped
(Gm_unTTs) reads were mapped against the unanchored *M. x domestica* v1.0 contigs
using GMAP, to observe if the apple genome gene space might be extended to include
the chromosome unanchored contigs (Box 5).

```
Box 5
$ gmap_build -d <Genome_db> -k 15 <genome.fasta>
$ gmap -D gmap-2012-01-11/gmapdb/ -d Genome_db -k 15 -f 3 -b -O --nofails Trinity.fasta
```

```
> Trinity_gmap.gff3
#Transcript mapping against Malus ESTs
$ gmap -D gmapdb/ -d Malus_genes -k 15 -f 3 -b -O --nofails Trinity_min300.fasta >
Trinity_min300_Mdo_gmap.gff3
```

2.2.10 Evaluating the accuracy of the reconstructed transcripts

Reconstructed transcripts less than 300 bases long were filtered out using a custom script (Box 4). Mapped and filtered transcripts from each approach were evaluated for breadth coverage of the already existing pseudo haploid mapped gene models v1.0, as well as a set of 15243 greater than 500 bases long *M. x domestica* ESTs (Box 6). The breadth coverage of gene model exons and ESTs features by the reconstructed transcripts were then calculated using BEDTools' coverageBed (Quinlan and Hall, 2010). The number of transcripts with greater than 0-79.9, 80, 90, 95, and 100% breadth coverage in each of the gene models or *M. x domestica* ESTs gene set breadth coverage were illustrated using a histogram plotted in ggplot2 (Wickham, 2009) (Box 6).

```
Box 6
#Calculating breadth coverage of gene models
$ coverageBed -hist -s -a Trinity_gmap.gff3 -b Malus.gff3 > Trinity_gmap_exon_cov.bed


#Enumerate the covered features
# Exon feature coverage
$ awk '$15>=0.80' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=0.90' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=0.95' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=1' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=0' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | grep -c 'MDP'
```

```
# mRNA feature coverage
$ awk '$15>=0' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | sed 's/_/    _/g' |
awk '{print $1}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=0.95' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | sed 's/_/    _/g' |
awk '{print $1}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=0.90' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | sed 's/_/    _/g' |
awk '{print $1}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=0.80' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | sed 's/_/    _/g' |
awk '{print $1}' | sort | uniq | grep -c 'MDP'
$ awk '$15>=1' Trinity_gmap_exon_cov.bed | awk '{print $10}' | sort | uniq | sed 's/_/    _/g' |
awk '{print $1}' | sort | uniq | grep -c 'MDP'
```

2.2.11 Visual comparison of the reconstructed transcripts

Reconstructed transcript gene models were visualized using Integrative Genomics Viewer
(IGV) (Robinson*, et al.*, 2011) to illustrate full transcript reconstruction and alternative
splicing. The gene models obtained from the apple genome pseudo haploid molecules and
the reconstructed transcripts (Ab, Gm and PTTs mapped to genome) were uploaded on to
the IGV browser as tracks to compare the features in each gene model. A read alignment
file was also uploaded to give an indication of the read coverage across the features.
Feature patterns were noted as, an extension at either 3´or 5´ends, insertion, or deletion
splicing patterns.

2.2.12 Mapping reconstructed transcripts to chloroplast DNA and mitochondrial DNA

The filtered transcripts from section 2.2.9 were mapped to the chloroplast DNA (cpDNA)
and mitochondrial DNA (mtDNA) using GMAP to produce cpDNA and mtDNA gene
models (Box 7). The gene models as well as the cpDNA and mtDNA length coordinates

were then used to draw cpDNA and mtDNA plasmids with gene annotations using GenomeVx (Conant and Wolfe, 2008).

Box 7

 # Creating index

$ gmap_build -d <mtDNA> -k 15 <mtDNA.fasta>

$ gmap_build -d <cpDNA> -k 15 <cpDNA.fasta>

#Mapping transcripts to Malus cpDNA and mtDNA

$ gmap -D gmapdb/ -d MtDNA -k 15 -f 3 -b -O --nofails Ab_combined.fasta >
Ab_combined_mtDNA_gmap.gff3

$ gmap -D gmapdb/ -d CpDNA -k 15 -f 3 -b -O --nofails Ab_combined.fasta >
Ab_combined_cpDNA_gmap.gff3

#Coverage analysis was done as shown in Box 6.

2.2.13 Functional annotation

Pooled data sets of the genome-guided (Ab and Gm) and their respective transcripts reconstructed from reads unmapped to the 17 pseudo molecules using Trinity were annotated by BLASTn/x against swiss-prot, *Malus* CDS, *Malus* plant unique transcripts (Dong*, et al.*, 2004), *Malus* ESTs, Rice*, Populus,* Peach*,* Tair, and Grape using local BLAST as well as against the NCBI non-redundant protein database (NR), using Blast2GO(Conesa*, et al.*, 2005) (e-value 1e-05; Box 8).

Box 8

#Annotating by Blastall using BLASTn or BLASTx against:

Databases: #1) a combined database of rice, grape, peach, *Arabidopsis*, *Malus* EST, *Malus* PUT and *Populus* #2) swiss-prot #3) non-redundant (nr) protein database

a) formatdb -i <database.fasta> -p F # for nucleotide sequence databases

---

blastall -p blastn -a 2 -d <database.fasta> -e 1e-05 -v 1 -b 1 -m 8 -i <Ab3_transcripts.fasta> -o

<Ab3_transcripts_bn.blast   #output BLAST result as in tabular format

b) formatdb -i <database.fasta> -p T # for amino acid sequence databases

blastall -p blastx -a 2 -d <database.fasta> -e 1e-05 -v 1 -b 1 -m 8 -i <Ab3_transcripts.fasta> -o

<Ab3_transcripts_bx.blast   #output BLAST result in tabular format

#the filter script "blast_best.pl" retains only the best scoring BLAST hits

#Blastn filtering

$ perl blast_best.pl <Ab3_transcripts_bn.blast><Ab3_transcripts_bn_best.blast>

#Blastx filtering

$ perl blast_best.pl <Ab3_transcripts_bx.blast><Ab3_transcripts_bx_best.blast>

Merge all BLAST results (BLASTn/x) to include all the sequences in the transcripts.fasta file

by using the <IDs_less_50_list.txt> from above.

$ perl merge_blasts.pl Ab3_transcripts_bn_best.blast

All_mapped_accepted_hits_mapped_transcripts_Ab3_stat_less_50_Id_list.txt >

Ab3_less_50_merged.blast

Repeat the step above till all high scoring BLAST hits have been in co-operated

Get the TAIR and UniprotKB accession equivalent of the BLAST hits by submitting to

UniprotKB ID mapping, and ID retrieval. Sometimes some accessions are deleted, merged and

replaced.

#Retrieve the replaced/merged accessions: paste the replaced/merged accessions in the script

(**get_replaced_uniprotKB.sh** (Appendix 8.4)). This script writes the replaced accession

followed by the accession that replaced it.

Make the necessary changes to the replaced/merged accessions

$ perl replace_merged_uniprotKB.pl replaced_merged_uniprotKB.txt

UniprotKB_transcript_list.txt > UniprotKB_transcript_list_replaced.txt

#Re-submit the UniprotKB list, now with an updated list of accessions to retrieve gene names

and descriptions. Download the tab-delaminated file of the UniprotKB accessions

#Merge the UniprotKB annotations with the merged blast, to annotate all the transcripts.

$ perl merge_UniprotKB_annotations.pl Ab3_less_50_merged.blast

Ab3_uniprotKB_annotations.tab > Ab3_less_50_merged_uniprotKB.txt

The output file at this stage is now comprehensively annotated with respect to the databases

used.

Putative functions of the reconstructed transcripts were assigned by similarity to the main functional categories of the Munich Information Centre for Protein Sequences (MIPS, v2.0) *Arabidopsis* accessions with greater than 80% identity and an e-value less than 0.05 (Mewes*, et al.*, 2008) (Box 9). Twenty-two main functional categories of the funcat 1.3 classification scheme were used to classify the putative functions of the reconstructed transcripts (Ruepp*, et al.*, 2004). The main categories used were, metabolism, energy, cell fate, storage proteins, development, interaction with environment, cell defence, communication, protein fate, protein synthesis, transcription, transposable elements, cell cycle and DNA processing, protein with binding function, protein activity regulation, transport regulation, systemic interaction with the environment, biogenesis of cellular components, cell type differentiation, organ differentiation and unclassified transcripts (Mewes*, et al.*, 2008) (Box 9). The frequencies of each category per transcript data set were then compared across tissues and time of fruit harvesting (Box 9).

Box 9

#Functional categories

#The MIPS functional categories used in this protocol were *Arabidopsis* greater than 80% best hits. Tair accession equivalents were thus retrieved for each gene transcript and annotated for function

$ perl merge_category.pl functional_category.txt Tair_accession_list.txt | awk '{print $2}' > classified_accession(1-20).txt

#Append the columns of the each functional category into a functional category merged file

$ paste -d "\t" Tair_accession_list.txt classified_accession(1-20).txt > func_all.txt

---

The reconstructed transcripts were assigned KEGG Orthology (KO) terms by mapping the respective UniprotKB accessions for each transcript to KO terms using UniprotKB or KOBAS. The list of KO terms for each time point and tissue type were then mapped to

KEGG pathways for enzyme function and pathway annotation. Frequency of the mapped KO terms within a pathway constituted the level of enrichment of the pathway as well as the enzyme in question. The transcript pathway information was then appended to the rest of the annotations, creating a comprehensive transcript annotation database. "No-Hit" transcripts were annotated by domain/motif search using CLC-Bio Genomic Workbench 5.5.1 against the Pfam domain database.

2.2.14 Transcripts abundance estimation and differential expression analysis

Genome-guided reconstructed transcripts were estimated in abundance using Cufflinks with expression values in fragments per kilo base of exon per million mapped sequence reads (FPKM) normalized for total mapped fragments and exon length for all six biological conditions replicated twice for both peel and pulp apple fruit tissues.

The effects of rRNA *in-silico* removal, mapped read sub-setting, and mapped read de-duplication were evaluated on genome-guided mapped reads. The analysis of these factors was done only for Ab read alignments. The experiment was designed as a four (rRNA, de-duplication, sub-setting and no-treatment) by one (Ab) factorial one-way analysis of variance (ANOVA) and analysed using StatPlus software package. Any one of the factors were considered to be significantly different at $p < 0.05$. The effect of the treatments on the number of exons detected and read support for each exon detected were used as parameters to pass a treatment.

2.2.14.1 Decontaminating reads of ribosomal reads

Reconstructed transcripts with best scoring BLAST hits against rRNA ESTs were

selected from the three transcript reconstruction approaches. The set of rRNA ESTs were then re-blasted against the NCBI, non-redundant protein sequences using Blast2GO at an e-value of e^{-10} (Box 10). The set of transcripts satisfying the stringent conditions and having hits to rRNA ESTs were selected and converted into a database of rRNA transcripts (Box 10), against which the rest of the reconstructed transcripts were blasted to de-contaminate the transcripts of rRNA transcripts (Box 10).

Box 10

#Extract transcripts with hits to ribosomal protein (from Box 7)

$ grep 'ribosomal protein' transcripts.blast | awk '{print $1}' | sort | uniq > ribosomal_transcripts_IDs.txt

#Extract the ribosomal_transcripts.fasta using "split_contigs.py"

$ python split_contigs.py reconstructed_transcripts.fasta ribosomal_transcripts_IDs.txt ribo_free_transcripts.fasta ribosomal_transcripts.fasta

#Confirm the ribosomal_transcripts.fasta annotations by re-blasting against the NCBI's nr, protein database using BLAST2GO at e-value= e-10

#Ribosomal read de-contamination

#Prepare a database of the ribosomal transcripts set

$ formatdb -i ribosomal_trancripts.fasta -p F

#Using blastn, BLAST the transcripts against the ribosomal transcripts retaining only the best hits.

$ blastall -p blastn -a 2 -d ribosomal_trancripts.fasta -e e-10 -v 1 -b 1 -m 8 > ribo_transcripts.blast

#Filter the BLAST output and retain only the best scoring hits:

$ perl blast_best.pl ribo_transcripts.blast ribo_transcripts_best.blast

#Re-filter the best-blast output to retain only high identity and more than 300 base length and high score hits

$ awk '$3==100' ribo_transcripts_best.blast | awk '$4<=3' | awk '$11==0.0' > ribo_transcripts_only.blast

#Column $3 = identity; $4 = number of mismatch; $11 = e-value

#Print the list of ribosomal transcript IDs for use in extracting ribosomal free transcripts

$ awk '{print $1}' ribo_transcripts_only.blast | sort | uniq > ribo_transcripts_only_list.txt

#Extract the ribo_free transcripts using "split_contigs.py"

$ python split_contigs.py Ab.fasta ribo_transcripts_only_list.txt Ab_ribo_free.fasta

```
Ab_ribo_only.fasta
# concatenate the ribo_free Ab or Gm fasta
#Ab and Gm unmapped transcripts were generated by de novo assembly with Trinity
$ cat Ab_mapped_transcripts_ribo_free.fasta Ab_unmapped_ribo_free.fasta >
Ab_combined_All.fasta
```

Mapped reads were used to create new fastq files corresponding to the mapped read
sequences (Box 11). The mapped read sequence files were then used to create a list of
rRNA read IDs by aligning the reads to a set of rRNA transcripts. The list of rRNA read
IDs was then used to extract rRNA-free mapped reads (Box 11). The de-contaminated
rRNA-free aligned reads were then used to estimate transcript abundance and differential
expression.

Box 11
```
## Covert the bam file to a SAM file format then make a header list of the sequences:
$ samtools view -h -o accepted_hits.sam accepted_hits.bam
#Create a list of the transcripts IDs from aligned reads
$ grep 'H-126' accepted_hits.sam | awk '{print $1}' | sort | uniq > accepted_hits_list.txt
# This list will be used to extract fastq sequences for read 1 and read 2 respectively
#Create a fastq index of the clipped reads.
$ cdbfasta -Q pe1_clip.fastq
$ cdbfasta -Q pe2_clip.fastq
# -Q option was used because the input was a fastq
#Retrieve fastq sequences with cdbyank corresponding to read 1 and read 2
$ grep 'H-126' accepted_hits_list.txt | cdbyank pe1_clip.fastq.cidx -d pe1_clip.fastq -o
pe1_mapped_cd.fastq
$ grep 'H-126' accepted_hits_list.txt | cdbyank pe2_clip.fastq.cidx -d pe1_clip.fastq -o
pe2_mapped_cd.fastq
$ perl trinityrnaseq_r2012-06-08/util/alignReads.pl --max_dist_between_pairs 129 --
retain_SAM_files --left pe1_mapped_cd.fastq -o ribo_aligned --right pe2_mapped_cd.fastq --
target ribosomal_trancripts.fasta --aligner bowtie --seqType fq
```

```
$ grep 'H-126' ribo_aligned.nameSorted.sam.+.sam | awk '{print $1}' | uniq >
ribo_aligned.sam_list.txt
#Remove ribosomal reads from the alignment file (SAM)
$ perl mapped_ribo-free_reads_sam.pl accepted_hits.sam ribo_aligned.sam_list.txt > ribo-
free_accepted_hits.sam
#Get the @SQ header for the sam file in the subsequent steps. The script disregards the @SQ
headers and these will be needed in the SUB-set SAM file:
$ head -n (size of Ab_sample) accepted_hits.sam > accepted_hits_header.txt
#Append Header to .sam
$ sed '$r end' accepted_hits_header.txt ribo_free_accepted_hits.sam >
ribo_free_accepted_hits_sot.sam
$ samtools view -Sh -o ribo_free_accepted_hits_sot.bam ribo_free_accepted_hits_sot.sam
# Estimate differential transcript expression using Cuffdiff
$ cuffdiff -o peel_pulp_SUB_Ab -T -L <label>Malus.ggf3 ribo_free_accepted_hits_sot(25-
150).sam
```

The effect of read count uniformity across time points on transcript abundance estimation
was evaluated by sub-setting the aligned reads to a uniform read count (Box 12).

Box 12
#Sub-setting mapped reads
#Count the number of unique transcript IDs in the IDs list files created in Box 10 and establish the
benchmark for use in sub-setting
$ grep -c 'H-126' accepted_hits_list.txt
# Subset the list of transcript IDs
$ for i in `cat accepted_hits_list.txt`; do echo "$RANDOM $i"; done | sort | sed 's/^[0-9]+ //' | awk
'{print $2}' | head -n (benchmark) > accepted_hits_SUB.txt
#Now extract the subset from the aligned reads "accepted_hits.sam"
$ perl mapped_SUB_reads_sam.pl accepted_hits.sam accepted_hits_SUB.txt >
accepted_hits_SUB.sam
#Get the @SQ header for the sam file in the subsequent steps. The mapped_SUB_reads_sam.pl
script dis-regards the @SQ headers and these will be needed in the SUB-set SAM file:
$ head -n (size of Ab_sample) accepted_hits.sam > accepted_hits_header.txt
#Append @SQ Header to SUBset.sam

```
$ sed '$r end' accepted_hits_header.txt accepted_hits_SUB.sam > accepted_hits_SUB_sot.sam
# Estimate differential transcript expression using Cuffdiff
$ cuffdiff -o peel_pulp_SUB_Ab -T -L <label>Malus.ggf3 accepted_hits_ SUB_sot (25-150).sam
```

Potential PCR duplicates were removed from the aligned reads using Picard (Box 13). De-duplicated reads were then further subset as outlined in Box 12. Transcripts abundance was then calculated using both subset and none subset de-duplicated aligned reads (Box 13).

```
Box 13
#De-duplicate and subset
$ java -Xmx2g -jar picard-tools-1.61/picard-tools-1.61/MarkDuplicates.jar
INPUT=accepted_hits.sam OUTPUT=accepted_hits_rdp.sam METRICS_FILE=PCR_duplicates
REMOVE_DUPLICATES=true VALIDATION_STRINGENCY=SILENT
#Create a list of the transcripts IDs from de-duplicated aligned reads
$grep 'H-126'accepted_hits_rdp.sam | awk '{print $1}' | sort | uniq > accepted_hits_rdp.txt
#Count the number of unique transcript IDs and establish the benchmark for use in sub-setting
$grep -c 'H-126' accepted_hits_rdp.txt
$ for i in `cat accepted_hits_rdp.txt`; do echo "$RANDOM $i"; done | sort | sed 's/^[0-9]+ //' | awk
'{print $2}' | head -n (benchmark) > accepted_hits_SUB_rdp.txt
#Subset de-duplicated reads
$ perl mapped_SUB_reads_sam.pl accepted_hits_SUB_rdp.txt accepted_hits_rdp.sam >
accepted_hits_rdp_SUB.sam
#Append Header to SUBset.sam : Use the headers created in Box 9
$ sed '$r end' accepted_hits_header.txt accepted_hits_rdp_SUB.sam >
accepted_hits_rdp_SUB_sot.sam
# Estimate differential transcript expression using Cuffdiff of the de-duplicated and subset SAMs
$ cuffdiff -o peel_pulp_SUB_rdp_Ab -T -L <label>Malus.gff3 accepted_hits_rdp_SUB_sot(25-
150).sam
# De-duplicate only
# Estimate differential transcript expression using Cuffdiff of the de-duplicated SAMs
$ cuffdiff -o peel_pulp_rdp_Ab -T -L <label>Malus.gff3 accepted_hits_rdp(25—150).sam
```

The effect of each of the treatments on transcript abundance estimation and transcript detection were then analysed using a one-way ANOVA using StatPlus software package.

2.2.15 Pooled transcript abundance estimation and differential expression

The expression levels of the reconstructed rRNA-free transcripts were estimated by mapping reads to the combined transcript set of genome-guided and *de novo* reconstructed transcripts using RNA-seq by Expectation Maximisation (RSEM) (Box 14). A single-end read set was constructed by simply excluding the second read of each pair. RSEM aligned reads to transcripts using Bowtie, then computed maximum likelihood (ML) abundance estimations using the Expectation-Maximisation (EM) algorithm(Li and Dewey, 2011). In order to treat each transcript individually the "--no_group_by_component" option was used. Transcript differential expression across time points for each fruit tissue were estimated using Edge-R (Robinson*, et al.*, 2010). Transcript differential expression was considered significant if the calculated p-value was less than 0.05. Transcript abundance (counts) estimations were normalized using trimmed means of M-values (TMM) according to Robinson, *et al*., (2010) to account for differences in the read mass composition of the RNA-seq samples. The effective library size (total mapped reads) for each sample generated by TMM was then used to calculate normalized FPKM. All statistically significant differentially expressed transcripts were further analysed for functional annotation using Blast2GO, MIPS and KEGG for enzyme systems and pathways as outlined in section 2.2.12.

Box 14

#Reconstructed ribosomal-free transcript abundance estimation using RSEM

#Align reads to transcripts and estimate abundance

$ perl trinity-r2013-02-16/util/RSEM_util/run_RSEM_align_n_estimate.pl --prefix Daa(25-150)_RSEM --thread_count 10 --no_group_by_component --transcripts transcripts.fasta --seqType fq --single Daa(25-150).fastq

'run_RSEM_align_n_etimate.pl' outputs abundance estimates per transcript in the Daa(25-150)RSEM_isoform.results file

#Create sample FPKM file

$ awk '{print $1}' sample_expression.results > transcript_ids.txt #This command creates a list of transcript IDs as ordered in the sample_expression file

$ awk '{print $7}' sample_expression(1-24).results >sample_FPKM(1-24).txt **#Create**

sample.matrix file

#Append the FPKM value columns from each sample_expression.results file to the transcript IDs list file separated by tab

$ paste -d "\t" transcript_IDs_list.txt sample_FPKM(1-24).txt >transcripts_counts.matrix

#Estimate differential transcript expression using EdgeR, bundled in Trinity

$ perl../../../Analysis/DifferentialExpression/run_EdgeR.pl --matrix transcripts_counts.matrix --transcripts Ab_combined_All.fa

#Computing normalized expression values

$cat sampleA.RSEM.isoforms.results | cut -f1,3,4 > feature_lengths.txt

$TRINITY_HOME/Analysis/DifferentialExpression/run_TMM_normalization_write_FPKM_matrix.pl --matrix transcripts_counts.matrix --transcript_lengths feature_lengths.txt

This command outputs normalized_transcripts_FPKM files that can be used for cross tissue and harvesting time-point transcript expression comparisons

#For more information refer to trinity-2013-02-25 manual in the doc/ folder

2.2.16 Gene level Expression estimation using Real time quantitative PCR (RT-qPCR)

2.2.16.1 Designing primers for the genes

Gene sequences were obtained from the NCBI-Genebank *M. x domestica* ESTs database. Gene specific primers corresponding to *anthocyanidin synthase (ANS); glucose-1-phosphate adenylyltransferase (ADG); NAD$^+$-dependent sorbitol dehydrogenase (SDH); pectate lyase A10 (PL); sorbitol transporter 5(SOT5)* genes were then designed against the identified gene sequences using NCBI-Primer-Blast to a stringent set criteria, including a melting temperature (Tm) of between 57.5°C and 60.5°C, an amplicon length of between 70 to 100 bp, limited predicted secondary structure or dimer formation, as well as avoiding GC clamping at the 3′ end. The designed primers were then synthesized at Integrated DNA technologies (IDT, www.idtdna.com/).
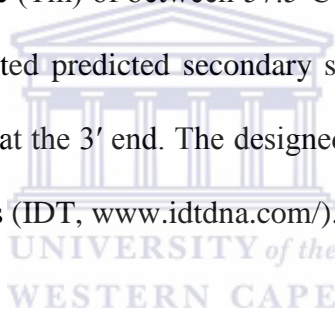
2.2.16.2 RT-qPCR

RT-qPCR was performed on de-contaminated total RNA (section 2.2.4) to detect relative transcript expression levels of the selected five genes. To verify reaction specificity, RT-qPCR reactions were carried out according to the manufacturer's instructions using Kapa SYBER FAST one-step RT-qPCR Universal kit (Kapa Biosystems), with a thermal profile as follows; pre-incubation at 95°C for 5 min followed by 25 cycles of 95°C (30 s), 60°C (30 s) and 72°C (30 s) with a final extension at 72°C for 5 min. Amplification and analysis by qPCR was carried out using the Rotor Gene (Rotor Gene Q, Qiagen). All reactions were performed with the Kapa SYBER Green FAST one-step RT-qPCR Universal Master Mix (Kapa Biosystems) with a thermal profile as follows: pre-

incubation at 95˚C for 5 min followed by 25 cycles of 95˚C (5 s), 60˚C (5 s) and 72˚C (8 s). Reactions were performed in triplicate using 2 µl 5x Master Mix, 0.5 µM each primer, 1 µl diluted cDNA and nuclease-free water (Kapa Biosystems) to a final volume of 20 µl. A negative no template control (NTC) was included in each run. Fluorescence was measured at the end of each extension step. Amplification was followed by a melting curve analysis with continual fluorescence and data acquisition during the 65˚C to 95˚C melt. PCR products were visualised by gel electrophoresis as described in section 2.2.7 The raw data was analysed with the Rotor Gene Q software version 2.0.4–build 3 and gene expression was normalised to *ß-actin* (CN938023), *tubulin*(GO568904.1) reference genes as well as the geometric mean of the *ß-actin* and *tubulin* (equations 1 and 2) according to Vandesompele, *et al*., (2002).
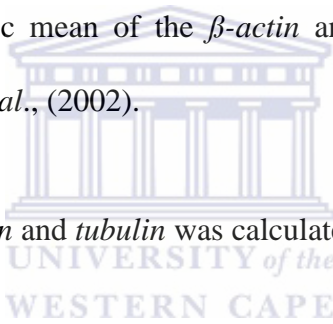
The geometrical mean of *ß-actin* and *tubulin* was calculated as:

$$GM = \sqrt[2]{(ß\,exp \times Texp)}$$

#Where: $ß_{exp} = ß$-*actin* expression, and T_{exp} = *tubulin* expression (Vandesompele, *et al.*, 2002).

Calculating target gene expression:

1) Mean sample expression:

$$Mtg = \Sigma \frac{(rep1 + rep2 + rep3)}{(number\ of\ reps)}$$

#Where: $[M_{tg}]$ = mean target gene concentration (copies/ul)

M_{tstd} = STDEV (rep1; rep2; rep3) #Where: M_{tstd} = mean gene concentration standard deviation; STDEV = standard deviation.

2) Normalising gene expression to *ß-actin* / *tubulin* or the geometric mean of *ß-actin* and *tubulin*

a) $[N_{tg}] = \dfrac{Mtg}{Mref}$

#Where N_{tg} = normalized target gene concentration (copies/µl); M_{tg} = mean target gene concentration (copies/µl); M_{ref} = mean reference gene concentration (copies/ul). The reference gene concentration could be any one of the *ß-actin, tubulin* or the geometric mean of the two reference genes

b) $N_{tg\_std} = \dfrac{Mtg_{std}}{Mref_{st}}$

#Where: N_{tg_std} = normalized target gene standard deviation; M_{tg_std}= mean target gene standard deviation; M_{ref_std} = mean reference gene standard deviation.

3) **Transformed target gene expression** $= \dfrac{\log_2 (Ntg)}{\log_2 (Ntg_{daa25})}$

#Where: N_{tg} = normalized target gene concentration (copies/µl); N_{tg_daa25} = normalized target gene concentration (copies/µl) at daa25.

The normalized expression values were log transformed to base two and the expression values of the five genes at daa25 were used as calibrators and assigned a nominal value of 1 (equation 3). Thus the expressions of the genes across the time points were regarded as down or up-regulated with respect to daa25.

2.2.17 Transcripts expression profiling

Differentially expressed transcripts were obtained as mentioned in section 2.2.15. To each transcript expression value (FPKM) of the differentially expressed transcripts a numerical value of one was added, then $\log_2$ transformed to normalise the data (Equation below).

Normalised expression (FPKM) = $\log_2(x + 1)$

\# Where x = RSEM calculated FPKM value.

Transcript expression values (FPKMs) for all the time points and tissue types were analysed for significant expression in the two tissue types as well as across the six fruit harvesting time points, using a two-way ANOVA. Tissue and time point responsive transcripts were then separated into tissue types, and clustered into closely related expression profiles using *k-means* clustering method in the MeV software package (Saeed*, et al.*, 2003). An estimate of the predictive power of the *k-means* clustering algorithm was done using Figure of merit (FOM) (Figure 2.1).

The FOM optimal cluster number was however adjusted to allow for closer grouping of the transcripts with similar expression patterns. A *k-means* clustering method, which uses a Pearson correlation was used to cluster the transcripts into 20 closely, correlated groups at a false discovery rate of 0.05. The 20 clusters were noted to exhibit eleven expression trends across the harvesting time points, hence clusters showing similar general expression trend were grouped together into one main expression trend group. The clusters from each expression trend group were modified to include the cluster numbers and expression trend group then concatenated into one data set using a custom script.

Mean Adjusted FOM values (± SD) vs. Number of Clusters
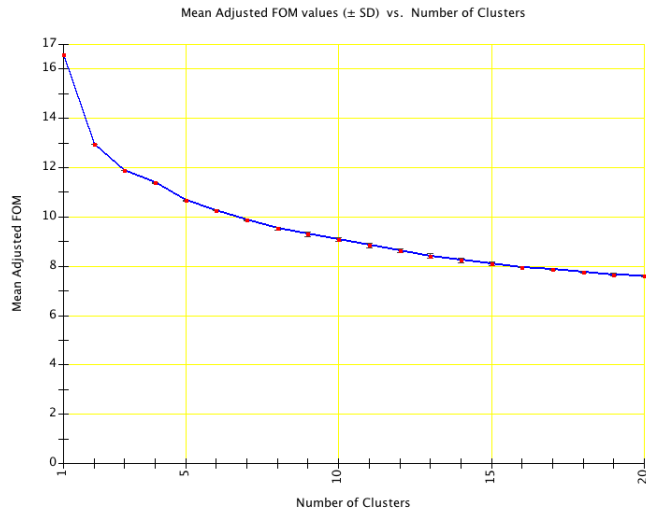
Figure 2. 1 Estimating the number of clusters to group transcripts using Figure of merit (FOM).

The transcript UniprotKB accessions, obtained by blastn/x in Box 8, were then mapped to KEGG Orthology (KO) terms and submitted to KEGG database to search for pathways and enzyme functions as outlined in section 2.2.12. Enriched pathways for each cluster were then noted, and pooled into cluster groups with similar general trend of transcript expression profiles. Comparisons across each group of clusters were made to give an overview of the main physiological processes showing the general expression patterns as represented by the group clusters.

Transcript profiles for tissue type comparisons were obtained by filtering the tissue and time point responsive transcripts to retain only transcripts with up-regulation in either peel or pulp tissues by more than 2 fold change using the MeV software package.

3 CHAPTER 3

M. x domestica **cv. 'Golden Delicious' pome fruit full-length transcript reconstruction and transcript abundance estimation from RNA-Seq data: genome-guided vs** *de-novo* **assembly**

Abstract

Full-length transcript reconstruction is the foundation for comprehensive cataloguing of gene and isoform species constituting a tissue's transcriptome. In this study the accuracy of *de novo* (Trinity) and genome-guided (Cufflinks) assemblers to reconstructing transcripts from RNA-seq data sets pooled from peel and pulp 'Golden Delicious' fruit tissues harvested at 25, 35, 60, 87, 135, and 150 days after anthesis (daa) was evaluated. Genome-guided reconstructed transcripts had higher count and breadth coverage of *M. x domestica* genome v1.0 gene models and ESTs gene set compared to *de novo* reconstructed transcripts. Both approaches extended 83% of the apple genome's gene models in both the 3´and 5´ends further annotating the genome. Approximately 30% of the combined peel and pulp transcriptome had no hits to nucleotide and protein databases. Among the transcripts with known function, transcripts encoding proteins with binding function were predominant. "No-hit" transcripts were composed mostly of leucine rich repeat; WD domain-G beta repeat and teratricopeptide repeat domains. Some of the reconstructed transcripts had similarity to sequence regions on the chloroplast DNA (cpDNA) and mitochondria DNA (mtDNA). Transcript abundance estimates using RSEM positively correlated to RT-qPCR estimates across the six time points, when multiple reference gene geometric means were used. Transcript levels of expression were affected by removing ribosomal, duplicate, as well as sub-setting read count. This study validated existing gene models and reconstructed new gene models residing on the pseudo molecules, unanchored apple genome v1.0 contigs as well as providing new transcripts with no known genomic positions. The data generated in this study constitutes a comprehensive catalogue of a combined peel and pulp developing 'Golden Delicious' fruit transcriptome.

Key words: Full-length transcript, transcript annotation, gene model, 'Golden Delicious'

3.1 Introduction

Various apple transcriptome sequencing projects have been done using both conventional Sanger sequencing (Baldi, *et al.*, 2004; Dal Cin, *et al.*, 2009; Takos, *et al.*, 2006; Ubi, *et al.*, 2006) and micro arrays (Janssen, *et al.*, 2008; Schaffer, *et al.*, 2007) in apple. Transcriptome sequence data is essential for molecular based assays necessary for apple crop improvement (Naganeeswaran, *et al.*, 2012). The advent of high throughput sequencing technologies has allowed quicker generation of transcript data in most organisms including apple. One of the biggest challenges in the utilization of such large volume of data is finding suitable bioinformatics analysis tools.

Northern blots, Sanger sequencing, ESTs and micro arrays are traditional methods that have been used in transcriptome analysis. These methods however are limited in throughput as well as detection ranges of lowly and highly expressed transcripts. The advent of high throughput next generation approaches resulted in sharp increases in transcript detection ranges, throughput, as well as significant reduction in clonal and amplification steps increasing sensitivity of detection. The next generation approaches with the assistance of Sanger sequencing have made possible the sequencing and assembly of the first *Malus x domestica* genome (Velasco, *et al.*, 2010). The availability of an acceptable apple genome version makes reference-guided-analysis of transcriptome data a possibility. However, the current apple genome version 1.0 shows high levels of heterozygosis and genome-wide duplication, therefore imposing technical challenges in whole genome and transcriptome assembly and annotation (Velasco, *et al.*, 2010). Furthermore, some 17 pseudo-haploid molecule assemblies of version 1.0 remain

complex with 37% of the genome unknown, many overlaps in scaffolds and 421 gaps on which potential protein coding and non-protein coding transcripts could reside. The *in-silico* predicted gene space and gene models of the apple genome in its current state are not representative of the entire transcriptome space of the actual apple genome. The high fragmentation state of the version 1.0 apple genome necessitates for the use of a genome-guided as well as *de-novo* assemblers towards full-length transcript re-construction.

With the objective of reconstructing and identifying full-length functional gene transcripts expressed in developing 'Golden Delicious' fruit, a genome-guided and a *de-novo* assembler, Cufflinks (Trapnell*, et al.*, 2012) and Trinity (Grabherr*, et al.*, 2011) respectively were compared. While Cufflinks has been adopted widely for genome-guided transcript analysis, a comparison of RNA-seq assemblers by Grabherr, *et al* (2011) reports Trinity to be among the best in *de-novo* assemblers with the ability for reconstructing paralogous transcripts.

This study evaluated the genome-guided Cufflinks as well as the *de novo* Trinity algorithms for their ability to reconstruct full-length transcripts from RNA-seq data obtained from six biological conditions, two tissue-types, replicated twice. A Unigene build was created a using *de novo* transcriptome assembly, which will be useful for finding genes not present in the current gene models and for the discovery of novel genes, as well as validation of existing gene models. The total transcripts discovered in this study will constitute an apple fruit transcriptome database essential for annotation, gene mining and profiling as well as molecular marker design for future apple crop improvement.

3.2 RESULTS AND DISCUSSION

3.2.1   Total RNA integrity

The integrity of the extracted total RNA was evaluated as described in section 2.2.5 using a Bio-analyser (Agilent). The RNA integrity (RIN) values give a measure of the respective intactness of an RNA sample. The RIN values for all the RNA samples were above 6 (Figure 3.1). Total RNA samples, which had RIN values greater than 5, were regarded as intact and suitable for RNA-Seq and RT-qPCR processing. The integrity of the sample RNA was also assessed by observing the distinct bends of the total RNA, which showed a 2:1 bend intensity ratio of the ribosomal RNA (rRNA) 25S to 18S with no sign of degradation and were considered intact, hence of high quality. The consistently high and almost uniform 25S to 18S rRNA bend intensity as well as RIN values greater than 6 for all the RNA samples indicated the reproducibility of the CTAB total RNA extractions. Only high quality total RNAs (Figure 3.1) were used for library preparation as described in section 2.2.6.
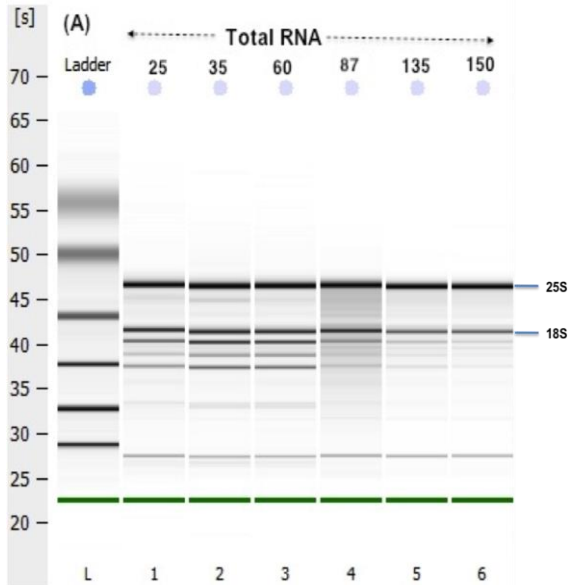
Figure 3. 1 Total RNA quality analysis by electrophoresis separation on a Bio-analyser. Total RNA from peel fruit sections sampled from fruits harvested at six time points in days after anthesis (daa), with RIN values of 7, 6.9, 7, 6.7, 8.2, and 8.4 for the samples 1, 2, 3, 4, 5, and 6 respectively.

3.2.2 Transcript assembly

The protocols used to assemble, annotate and profile transcripts in apple fruit are outline in Figure 3.2, with full experimental details given in chapter 2. RNA-seq libraries were prepared and sequenced from individual fruit samples at six time points (25, 35, 60, 87, 135 and 150 days after anthesis (daa), from two tissue types (peel and pulp) and two replicates. A combined total of 815 630 045 paired end raw reads were obtained by the HiScan SQ sequencing (Figure 3.4).
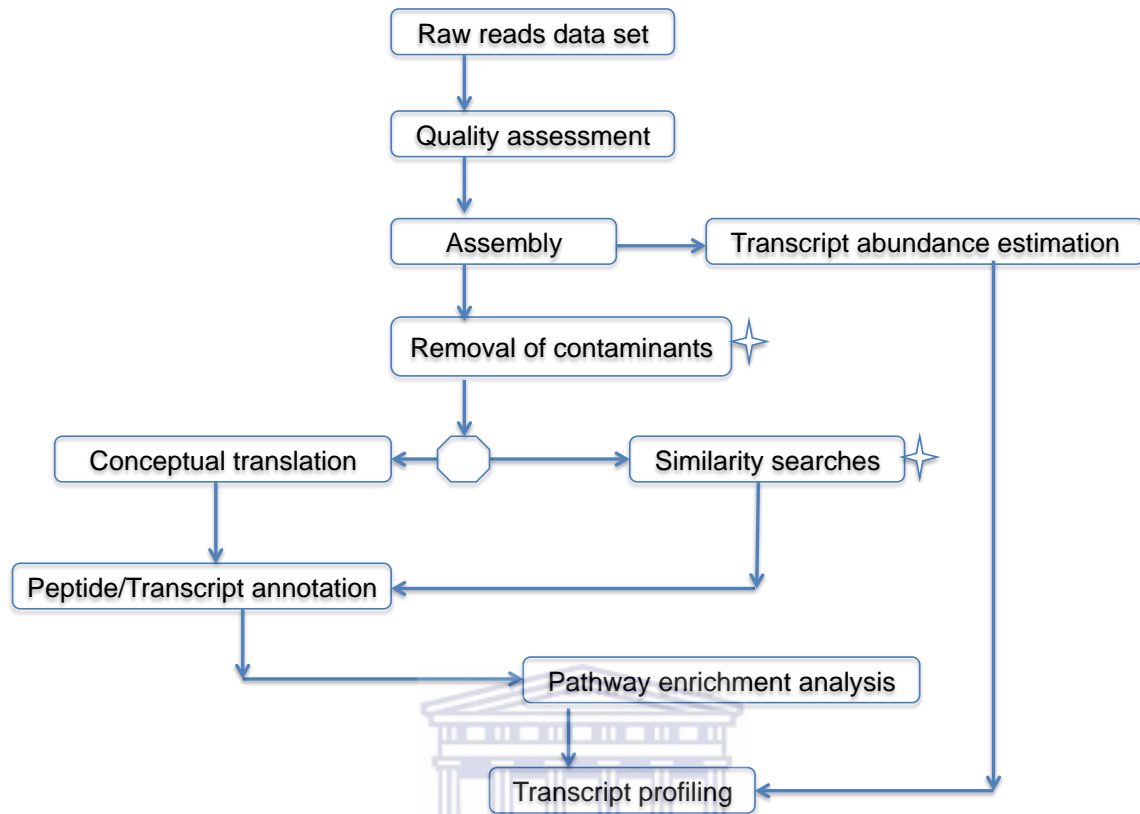
Figure 3. 2 Bioinformatics analysis of *M. x domestica* fruit peel and pulp tissue extracted mRNA data sets. The Figure shows the approaches used to assemble, annotate and profile transcripts from developing 'Golden Delicious' fruit. Stars indicate analysis performed using custom-written Perl, Python and/or Unix shell scripts. Protocols are described in chapter 2.

3.2.3   Raw reads processing

Raw reads obtained by the HiScan SQ were filtered to exclude low complexity reads and reads containing adaptor sequences, and their quality assessed using FastQC (Andrews, 2010). The protocol for quality checking and processing is as described in detail in section 2.2.7. Per base quality scores of the combined raw read data set was on average above 28 Phred score (Figure 3.3). Trimming of ambiguous bases from the 3´and 5´ends as well as clipping of adapter sequences and quality filtering significantly improved the quality scores of the reads, with per base quality score above 20 Phred score across the entire 96 base-long reads (Figure 3.3). Nine %of the reads were discarded as too short

while approximately 660 million high quality reads of 96 bases length were retained from all tissues and time points for assembly (Figure 3.3). The results suggested that the majority of the reads were high quality reads.
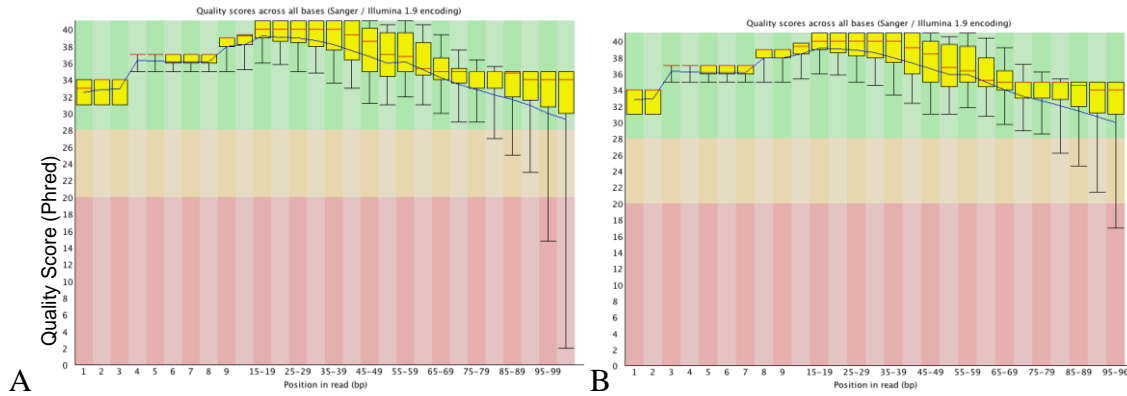
Figure 3. 3Read quality analysis before and after trimming and adapter clipping. A=analysis of raw reads, B = analysis of read quality after read processing. Read quality was viewed using FastQC version 0.50. Quality limit set at 20 Phred score, at a read percentage of 95% along 96 bp read length

3.2.4   Full-length transcript reconstruction

The workflow for full-length transcript reconstruction using genome-guided and *de novo* assemblers are outlined in Figure 3.4 and 3.5, with full experimental details given in section 2.2.8. The resulting clean reads from read quality checking and processing were pooled from six biological conditions, two tissue types and two replicates then assembled into transcripts using reference-guided assembler Tophat/Cufflinks (Trepnell, *et al.*, 2012) with and without the assistance of gene *M. x domestica* gene models as *Ab initio* and gene-model assisted assemblies respectively. The pooled data set was also used to assemble transcripts using *de novo* assembler Trinity-2012-02-25 (Grabherr, *et al.*, 2011).

Using approximately 660 million high quality filtered reads, a total of 127474 contigs were *de novo* assembled with Trinity (Figure 3.4), while 93950 and 94314 transcripts

were reconstructed using Cufflinks reference-guided assembly as *Ab initio* (Ab) and gene model assisted respectively (Figure 3.5).

Pooled paired-end and single-end reads (~81%) were used in the *de novo* reconstruction of 127474 transcripts, while 9% of the reads that were not used for assembly were discarded (Figure 3.4). The number of transcripts reconstructed by Ab and Gm were not significantly different, but were significantly different to those assembled by Trinity. Trinity had higher reconstructed transcript counts compared to Cufflinks (Figure 3.4 and 3.5), possibly because Trinity assembled transcripts from all the reads including those that could not map to the pseudo haploid molecules. As such the transcripts reconstructed by Trinity using pooled reads from both peel and pulp tissues constituted the totality of apple fruit *de novo* assembled transcripts.
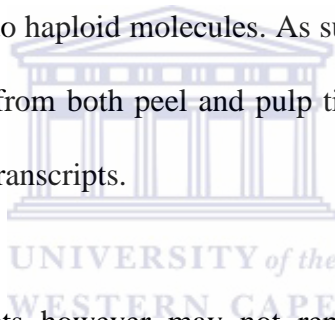
Trinity reconstructed transcripts however may not represent all the isoforms of the reconstructed genes, making genome-guided transcript reconstruction approach much more suited to identifying the possible forms of the reconstructed genes. Nonetheless, genome-guided transcript reconstruction is constrained by the completeness of the available genome. The apple v1.0 genome still has multiple gaps, which may have resulted in some reads not mapping to the reference genome (Figure 3.5).

Comparison and evaluation of only the transcripts that mapped to the 17-pseudo haploid molecules disregards a significant % of transcripts that did not map to the genome. As such, reads that did not map to the *M. x domestica* version 1.0 pseudo haploid molecules

in Ab and Gm approaches were pooled across time points and *de novo* assembled using Trinity to salvage transcripts that may have fallen into the unknown 37% of the apple genome (Figure 3.5). In order to find if the unanchored Malus genome contigs had gene space, Ab, Gm and pooled Trinity assembled transcripts that did not map to the pseudo-molecules were thus mapped to the unanchored contigs using GMAP (Figure 3.5), and the count of mapping and not mapping transcripts noted. Reconstructed transcript sequences with less than 300 bp lengths, more than 50% 'N' content were eliminated.

Of the 104559, 105754 and 88576 *de novo* assembled transcripts from reference-pseudo molecules-unmapped reads (Ab and Gm) and Trinity_unmapped approaches, 22, 22 and 26% mapped to the unanchored contigs respectively (Figure 3.4, 3.5). Considering these transcripts were reconstructed from genome-unmapped reads, they were not expected to map back to the genome or the genome-guided assembled transcripts. However, some full-length transcripts, reconstructed from the pooled unmapped read data sets, perfectly mapped back to the genome (~54%)and genome-guided assembled transcripts with high % identities of ~99% (Figure 3.5, 3.6). Only 8% of the de novo assembled transcripts had no hits to the combined genome (pseudo molecules and contigs). These unmapped transcripts possibly constitute a yet to be sequenced gene space.
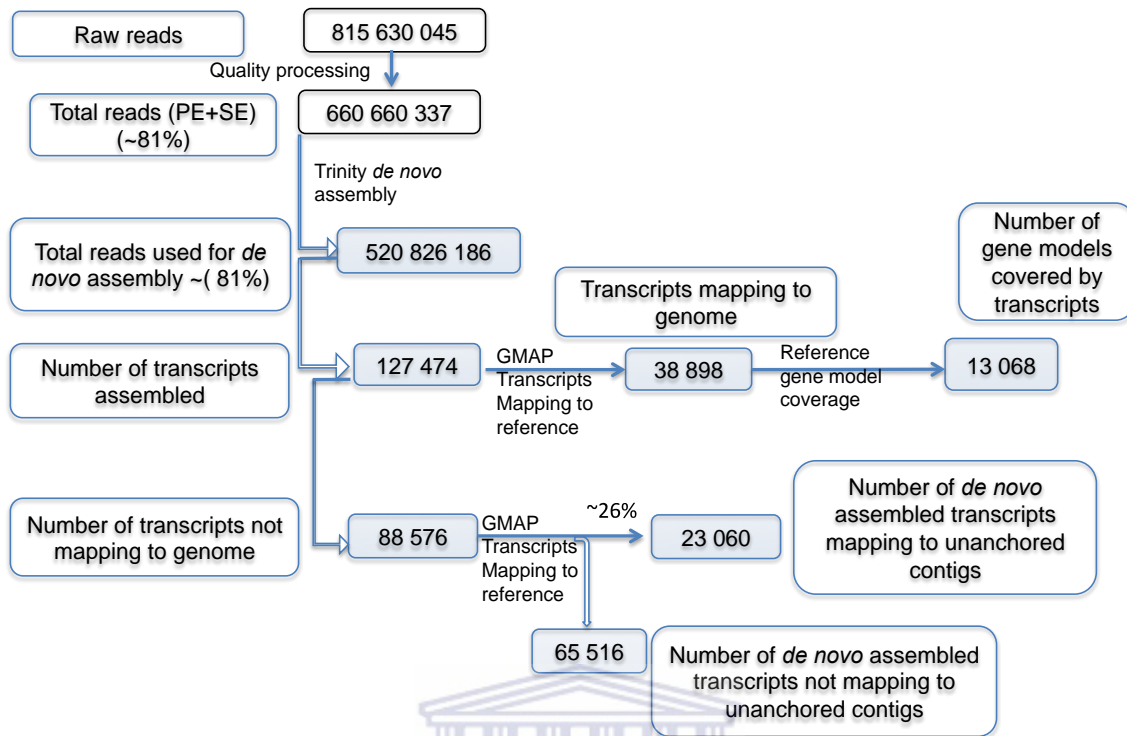
Figure 3. 4*De novo* **Assembling of sequence reads into transcripts using Trinity**. The Figure shows the read and transcript counts throughout the assembling and post assembly processing. At each stage of the process the analysis that was done is indicated. Reads were assembled using trinityrnaseq_r2012-10-05 and verified with trinityrnaseq_2013-02-25 (Grabherr, *et al.*, 2011). Transcripts mapping to the apple genome pseudo haploid molecules or unanchored genome contigs (http://www.rosaceae.org/) was done using GMAP. The percentages are indication of the proportion of reads of transcripts used.

Read mapping and transcript reconstruction by the Tophat/Cufflinks pipeline has a set threshold read count of 10, to call consensus(Trapnell, *et al.*, 2012). The continuity of the transcript during assembly is cut off each time the base coverage goes below the threshold(Trapnell, *et al.*, 2012). As such depending on the read coverage across a transcript, fragmented transcripts may be produced. All reads with low coverage below the thresholds are then reported as unmapped and are not used in transcript assembly. However, when multiple samples each with low read coverage of a particular transcript are pooled, their overall read coverage for that particular transcript may become higher than the threshold enough to be assembled into transcripts that can be mapped back to the

genome or genome-guided assembled transcripts. A combination of genome-guided and *de novo* approaches are able to salvage reads considered low coverage or falling into genome gaps and assemble them into full-length transcripts for a more comprehensive apple fruit transcriptome.
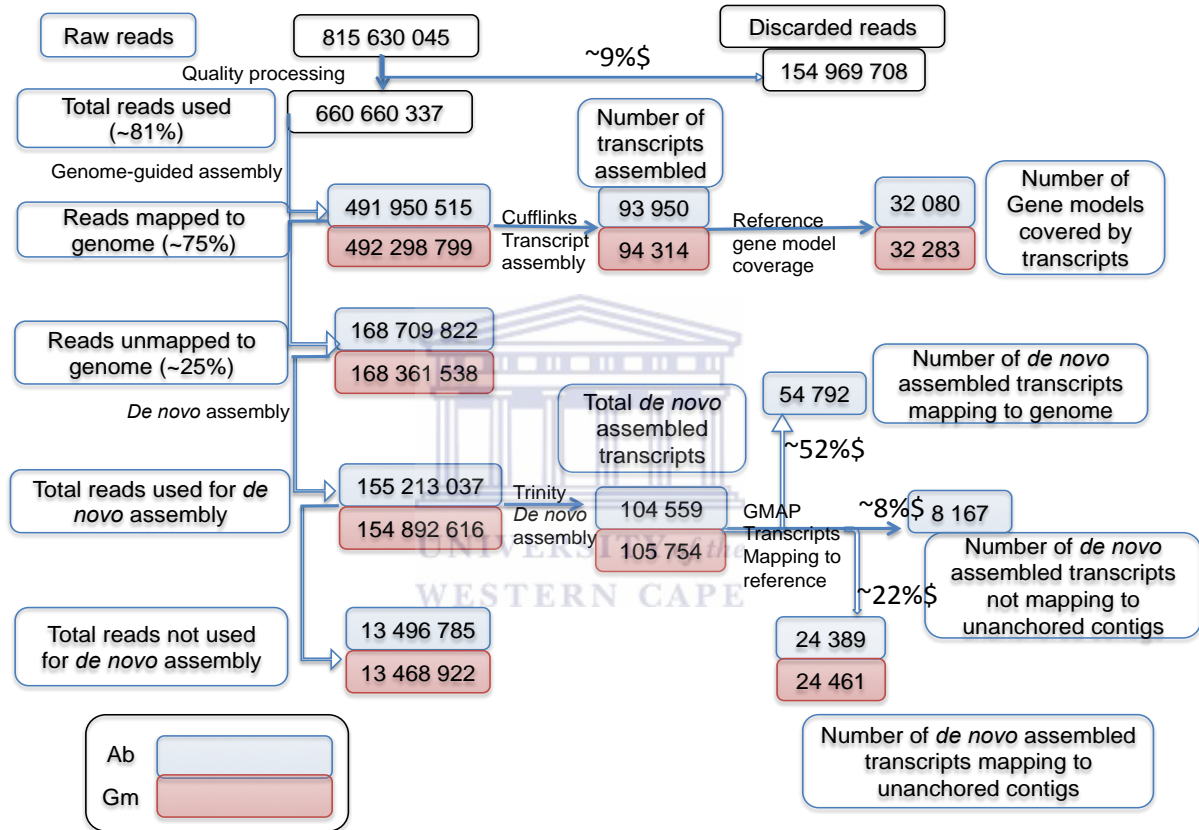
Figure 3. 5Genome-guided and *de novo* transcript reconstruction. The Figure shows the read and transcript counts throughout the assembling and post assembly processing. At each stage of the process the analysis that was done is indicated. Ab = *Ab-initio*, genome-guided transcriptome assembly without use of gene models. Gm = genome-guided transcriptome assembly with aid of gene models (Gff3 file, http://www.rosaceae.org/node/482). *De novo* assembly was done using trinityrnaseq_r2012-10-05 and verified with trinityrnaseq_2013-02-25 (Grabherr, *et al.*, 2011). Transcripts mapping to the apple genome pseudo haploid molecules or unanchored genome contigs (http://www.rosaceae.org/) was done using GMAP. The percentages are indication of the proportion of reads of transcripts used.
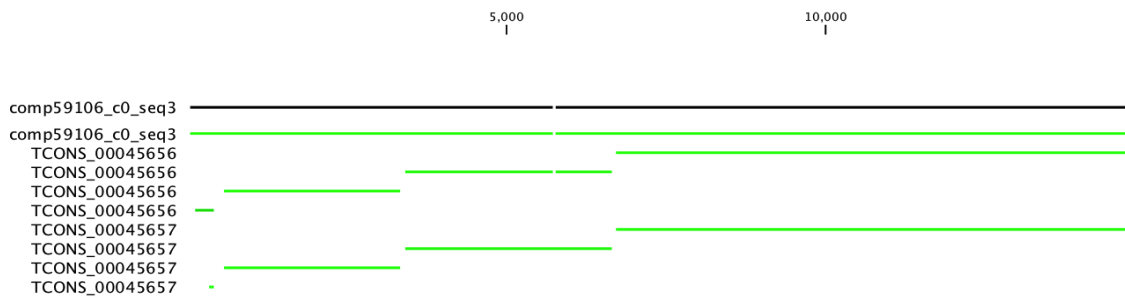
Figure 3. 6*De novo* **assembled transcripts reconstructed from genome-unmapped reads map to genome-guided assembled transcripts.**Transcripts whose name start with comp* were de novo assembled from genome-unmapped reads using Trinity, while those starting with TCONS* were genome-guided assembled. *De novo* assembled transcripts were mapped against genome-guided assemblies using BLAST.

The genome-guided assembly approaches used ~75% of the reads, 6% short of the read count used by Trinity in the pooled *de novo* assembly (Figure 3.4, 3.5). Mapping of~75% of the reads to the genome, suggests that the reads used in the assembly were high quality (Trapnell*, et al.*, 2012).However, the gapped nature of the current apple genome v1.0 suggested more might have mapped to the genome.

The high count of non-ATGC bases in the genome might also have increased the chances of rejecting the reads due to increased number of mismatches. Nonetheless, the genome was able to explain 80% of the combined reconstructed transcripts (Figure 3.5). The results thus showed that if an acceptable genome assembly is available, as with the apple genome v1.0 pseudo haploid molecules, genome-guided transcript assembly might be a better method. This was because genome-guided assembly was able to use more reads per unit transcript length than Trinity (Figure 3.5, 3.7). A substantial number of Trinity reconstructed transcripts were less than 500 bases long (Figure 3.7). The genome-guided

reconstructed transcripts clustered in the 500-2500 bases range, and a small number longer than 5000 bases (Figure 3.7).

A considerable number of Trinity-pooled reconstructed transcripts (88 576) did not map to the 17 pseudo-haploid molecules, using GMAP (Figure 3.4). These unmapped transcripts most probably reside in the un-sequenced or unassembled regions of the pseudo-molecules. Trinity like any other transcript *de novo* assembler is very sensitive to sequencing errors and the presence of chimeric molecules in the data set. Figure 3.7 shows Trinity to have assembled the largest number of transcripts, with the majority of the transcripts less than 500 bases long. Heterozygous and duplicated genomes like that of apple are complex and have multiple and some times very similar alleles. Such alleles produce closely related gene transcripts difficult to assemble *de novo*. These paralogous transcripts have high levels of identity, sometimes above 90% and are collapsed into one chimeric transcript, or may be fragment into smaller transcripts if the Trinity algorithm fails to resolve them (Martin and Wang, 2011; Strickler*, et al.*, 2012). It is also important to note that the current versions of Trinity algorithms cannot effectively separate isoforms and paralogous transcripts. A comparative study by Grabherr, *et al,* (2011) observed that reference-based assemblers discover a greater number of unique splicing patterns than *de novo* approaches, further highlighting the greater sensitivity of genome-guided assembly (Grabherr*, et al.*, 2011; Strickler*, et al.*, 2012)
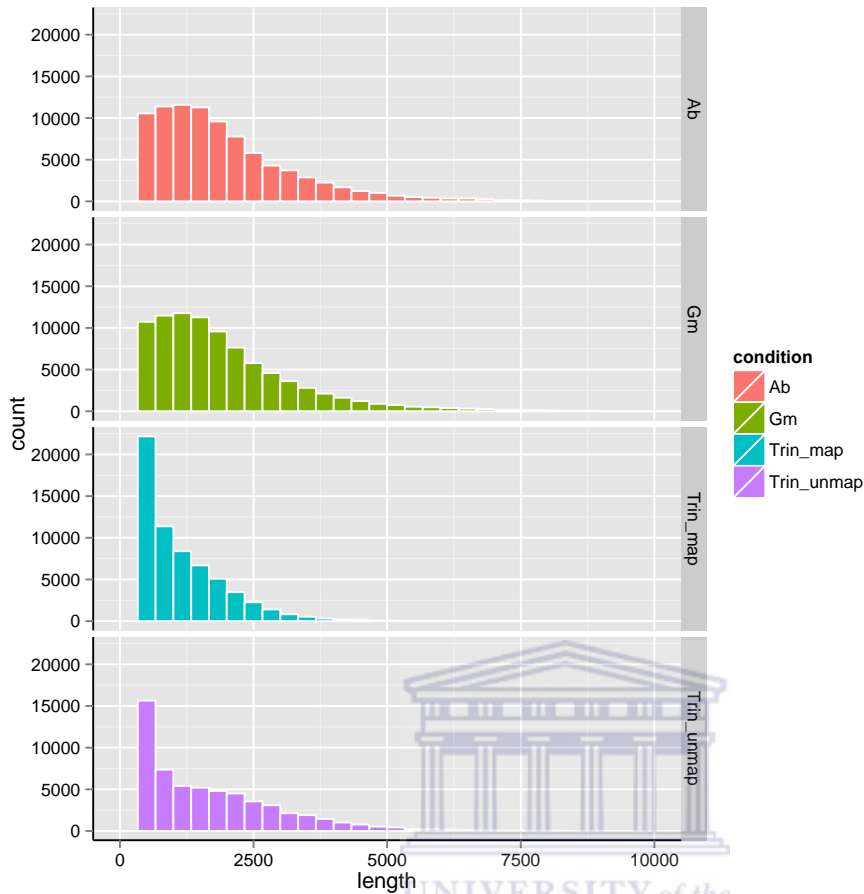
Figure 3. 7 Length distribution of transcripts reconstructed using Ab, Gm and Trinity approaches. Transcripts were assembled from a pooled data set ~660 million Illumina read sequences from six biological conditions (daa25, daa35, daa60, daa87, daa135, daa150), two tissue types (peel and pulp) replicated twice were pooled. Ab=genome-guided transcript assembly using *Ab initio*, Gm=genome-guided transcript assembly with the assistance to reference gene-models, Trin_map=Trinity reconstructed transcripts that mapped to the pseudo haploid molecules of *M. x domestica* genome v1.0 using GMAP, Trin_unmap= Trinity reconstructed transcripts that could not map to the reference genome using GMAP. Graph was drawn in R using ggplot2 version 0.9.3.1 package.

3.2.5 Reconstructed transcripts showed high breath coverage of gene models

The region of identity between two transcripts or gene models with respect to their length is called 'breadth coverage'. Breadth coverage is an indicator of the extent to which one transcript or gene model covers the length of the target transcript, hence the extent to

which the transcript has been reconstructed relative to the reference transcript. The accuracy of the reconstructed transcripts was evaluated using breadth coverage of the *M. x domestica* gene models at mRNA as well as exon level. A target gene model was covered if at least one feature from the reconstructed transcripts overlaid it. Levels of coverage ranged from greater than zero %, 80%, 90%, 95% and full coverage (100%) of the gene models by the reconstructed transcripts. The majority of Ab and Gm reconstructed transcripts had more than 80% breadth coverage of the gene models (Figure 3.8). However, a significant number of Trinity reconstructed transcripts covered less than 80% of the gene model mRNA or exon features (Figure 3.8A, B). Trinity reconstructed transcripts supported less than 15000 gene models, while Ab and Gm supported close to 25000 gene models.
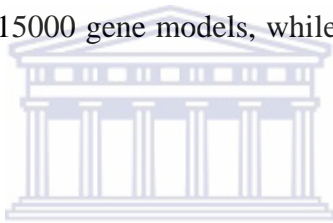
The criteria to assessing mRNA transcript assembly has recently been proposed, and basically requires assessment of the accuracy of the reconstructed transcript (Martin and Wang, 2011). According to Martin and Wang, (2011) a total count of reconstructed transcripts by each assembly approach is a subjective parameter, unless the accuracy of the putative full-length transcripts is evaluated (Martin and Wang, 2011).
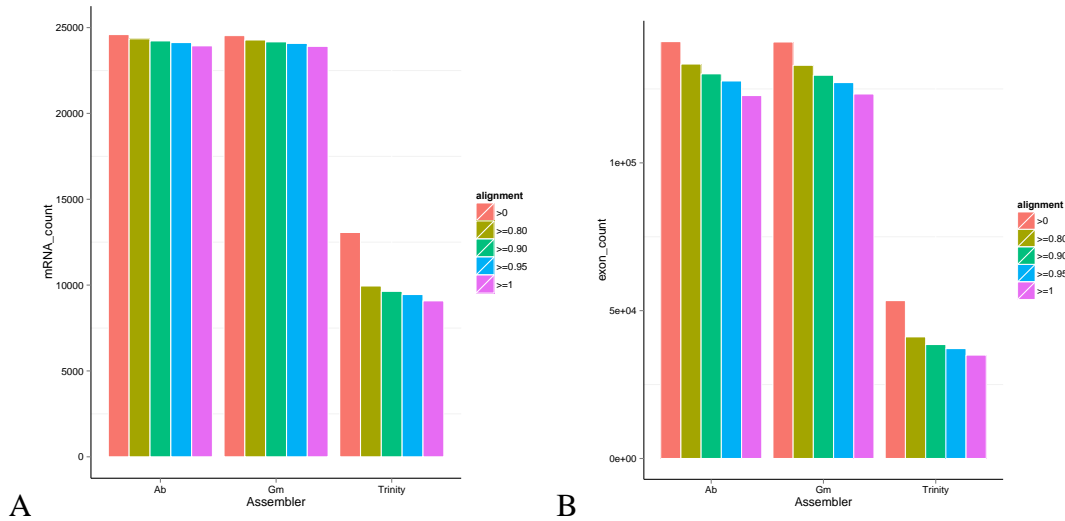
Figure 3. 8 Breadth coverage of reference gene models by transcripts reconstructed with Cufflinks and Trinity assemblers. A) Reconstructed transcripts breadth coverage of the mRNA genome gene model feature. B) Reconstructed transcripts breadth coverage of the exon genome gene model feature. The greater than zero %, >=80%, >=90%' >=95% and 100% are mutually inclusive categories. Graph plotted in R using ggplot2 R package

An assessment of both mRNA and exon level breath coverage by Cufflinks and Trinity reconstructed transcripts revealed that the former had better accuracy with no significant difference between more than 80% breath coverage and total transcripts with at least one feature supported. On the contrary a substantial number of Trinity reconstructed transcripts had less than 80% breath coverage of the gene models (Figure 3.7). As mentioned earlier the large number of fragmented and possibly miss assembled Trinity reconstructed transcripts might explain the low coverage of most of the mRNAs and exon features.

Proportional analysis of the assembly approaches revealed that Ab and Gm genome-guided transcript reconstruction approaches were at most in agreement on the number of the gene models reconstructed (Figure 3.9). Out of a total of 24837 unique gene models

reconstructed by both approaches, only 294 and 250 were unique to Ab and Gm respectively (Figure 3.9). Ab, Gm and Trinity reconstructed 24588, 24543 and 13068 gene models constituting 81.1%, 81%, and 43% of the total 30294 apple pseudo haploid molecules v1.0 gene models respectively. In addition to the gene models the transcript reconstruction approaches, Ab, Gm and Trinity reconstructed 7492, 7740 and 20742 novel genes respectively mapped to the 17 pseudo-haploid molecules, but unsupported by any one of the gene models. The current gene models of the apple genome version 1.0 were obtained through gene prediction and use of more than 300 000 apple ESTs (Velasco*, et al.*, 2010). Despite the large volume of the ESTs used, observations in this study suggested that there were still a substantial number of transcripts unknown to the current gene models. In this study, 25155 gene models among the three transcript reconstruction approaches were supported, with 12720 gene models supported by all three-assembly approaches (Figure 3.9). The transcripts constituting the intersection of the three approaches might form a set of transcripts highly expressed in the combined peel and pulp tissue transcriptome.

Assembler	Total genes reconstructed	Mapped to *Malus* gene models	Unmapped to *Malus* gene models
Ab	32 080	24 588	7 492
Gm	32 283	24 543	7 740
Trin_map	33 810	13 068	20 742
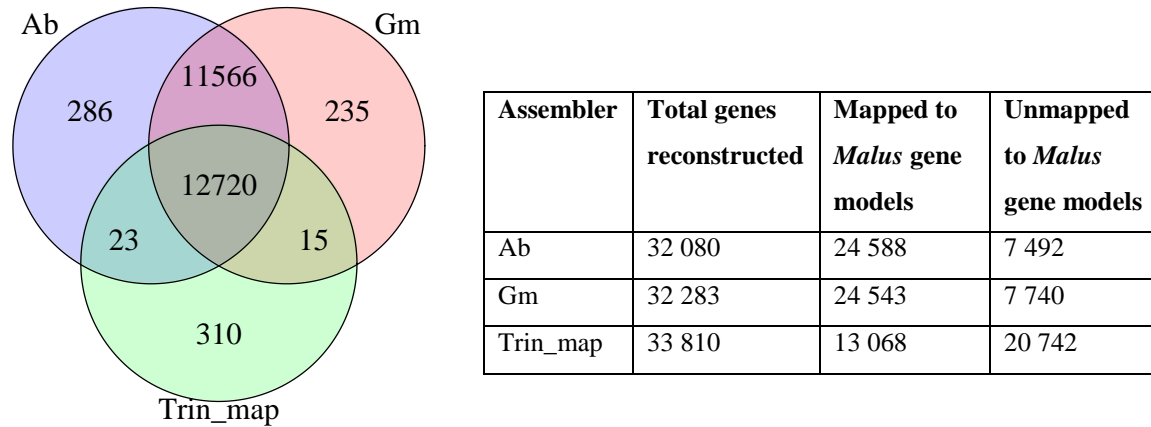
Figure 3. 9 Unique reference gene model transcripts reconstructed by Ab, Gm and Trinity methods. The table shows the number of unique genes encoded for by transcripts reconstructed by three assembly approaches and how many map or do map to *Malus* gene models. The Venn diagram on the other had shows the count and distribution of the unique genes encoded for by the transcripts mapping to *Malus* gene models. Overall the Figure explains how many unique genes were reconstructed by the combined transcript data set. Unmapped to *Malus* gene models = Number of reconstructed transcripts with no match to any gene model. Ab = Cufflinks *Ab initio* assembled transcript gene model, Gm = Cufflinks reference gene model guided assembled gene model, and Trin_map = gene models of Trinity reconstructed transcripts mapped to the genome using GMAP. Venn diagram drawn in R using Venn diagram package

3.2.5.1 Transcript accuracy with reference to a subset of *M. x domestica* genes

Alignment scores of the reconstructed transcripts to the genome alone might be misleading considering the gapped state of the current genome. To validate the quality of the reconstructed transcripts, transcripts were mapped to 15243 *M. x domestica* genes. The mapping coordinates of the transcripts to the ESTs set were then compared across assemblers and assembling approaches. The number of transcripts mapping to the *M. x domestica* ESTs set were shown to be higher in reference-guided assemblies compared to *de novo* assembly (Figure 3.10). Despite the lower number and percentage breadth-coverage of transcripts with hits to the ESTs set (Figure 3.10), the *M. x domestica* ESTs breath coverage by reconstructed transcripts was noted to be highest in Trinity reconstructed transcripts (Figure 3.10). The ability of Trinity to reconstruct full-length

transcripts was reported by Grabherr, *et al*., (2011). Several other studies corroborated the potential of *de novo* assemblers in reconstructing full-length transcripts, especially high coverage transcripts (Haas*, et al.*, 2013; Martin and Wang, 2011).
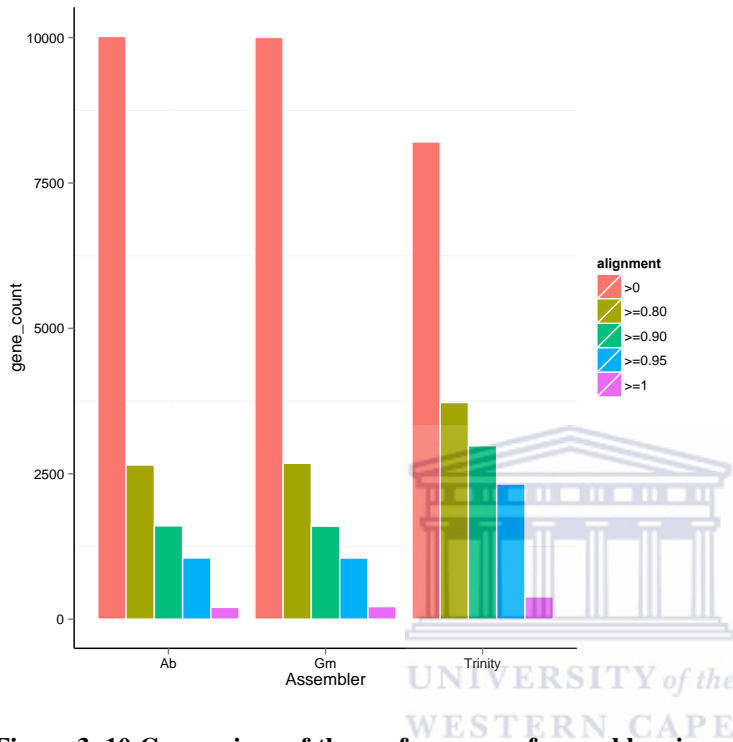
Figure 3. 10 Comparison of the performance of assemblers in reconstructing *M x domestica* gene set. The y-axis = count of unique *M. x domestica* genes aligned to reconstructed transcripts across a minimal per cent of their length. The greater than zero %, >=80%, >=90%' >=95% and 100% are mutually inclusive categories. Graph plotted in R using ggplot2

The gapped nature of the current apple genome v1.0 necessitated the use of a combined strategy to salvage, as well as optimise reconstruction of apple fruit transcriptome. Martin and Wang (2011) suggested the use of a combined approach, align-then-assemble or assemble-then-align depending on the quality of the existing reference genome. This study adopted the align-then-assemble strategy to take advantage of the high sensitivity of the reference-based assembler while leveraging the ability of Trinity to detecting novel and trans-spliced transcripts (Martin and Wang, 2011). As such, despite some of Trinity

pooled reconstructed transcripts being full-length transcripts, in this study Gm and Ab combined transcript sets were used to evaluate transcript abundance in the later stages.

3.2.6 Reconstructed transcripts improve the apple genome annotation

The reconstructed transcripts gene models were evaluated to observe if they improved the existing *Malus* gene models. To illustrate the improvement of the *Malus* gene models, a *ß-amyrin synthase* (BAS) gene model was randomly selected. According to the illustration of the BAS gene model in Figure 3.11A, the genome-guided approaches reconstructed more isoforms of the BAS gene model compared to Trinity (Figure 3.11A). Trinity and the genome-guided approaches reconstructed fully the BAS gene model as well as extending the annotation of the BAS gene model at both the 3´and 5´ends (Figure 3.11A). The existing *M. x domestica* gene models show no splicing variants, but in this study splicing variant gene models were provided as illustrated by the BAS gene model (Figure 3.11A).

Three BAS encoding transcripts from *Glycine max* (AAM23264.1), *Arabidopsis* (*AT1G78950.1*), and *M. x domestica* v1.0 (MDP0000227287) were aligned to the BAS transcript Ab reconstructed using CLC-Bio Genomic Workbench 5.5.1. The alignments revealed that the reconstructed transcript was similar to the apple gene model, *Arabidopsis*, and *Glycine max* in that order of decreasing similarity (Figure 3.11B). The reconstructed transcript extended the 3´and 5´ends, hence improving the apple genome annotation. Though the alignment analysis suggested the reconstructed BAS transcript was 4 and 9% different from AT1G78950.1 and AAM23264.1, Cufflinks was able to

98

fully reconstruct BAS as well as extend the 3´and 5´ BAS ends of the gene model. The observation corroborates the ability of Cufflinks to separate and reconstruct even closely related transcripts (Trapnell, *et al.*, 2012), as indicated by multiple isoforms in Figure 3.11A.

The 3' and 5' extending regions of the reconstructed BSA (120 and 160 nucleotides respectively) could be untranslated regions (UTRs), since they differ from the other plant gene models. The whole BAS nucleotide sequence was re-aligned against the same plant species as in the protein alignment with the same results being produced (Alignment shown in Appendix). The extending 160 nucleotides in the 5' as well as 120 bases in the 3' ends of the BAS gene were blasted against the NCBI (nr) nucleotide database (date blasted 27/08/2014). The sections of the gene suggested to be UTR regions had matches only against *Malus x domestica Beta-amyrin synthase* (OSC1), mRNA (ref|NM 001294017.1) and *Erioboryta japonica* Jiefangzhong mixed *amyrin synthase* mRNA, complete cds (gb|JX173279.2) with 100% and 94% respectively. This confirms that the 3' and 5' extending regions do indeed exist in the plant kingdom as coding sequences despite having stop codons in between and not having Methionine at the start of the sequence. The blast alignments are appended in the Appendix 8.10

Not the entire gene models were verified by either of the transcript reconstruction approaches in this study of the transcriptome of developing 'Golden Delicious' fruit. Genes not related to fruit development were not expected to be expressed during normal fruit development, except under particular conditions as elicited by the gene-environment

interaction. The study, however, provides verification to some of the existing gene models as well as elucidating to the existence of several 'novel' gene models, further annotating the apple genome.
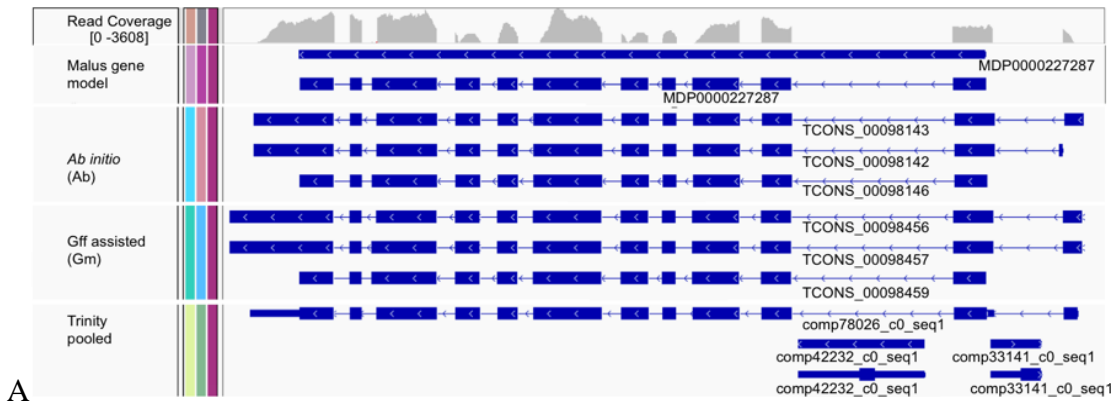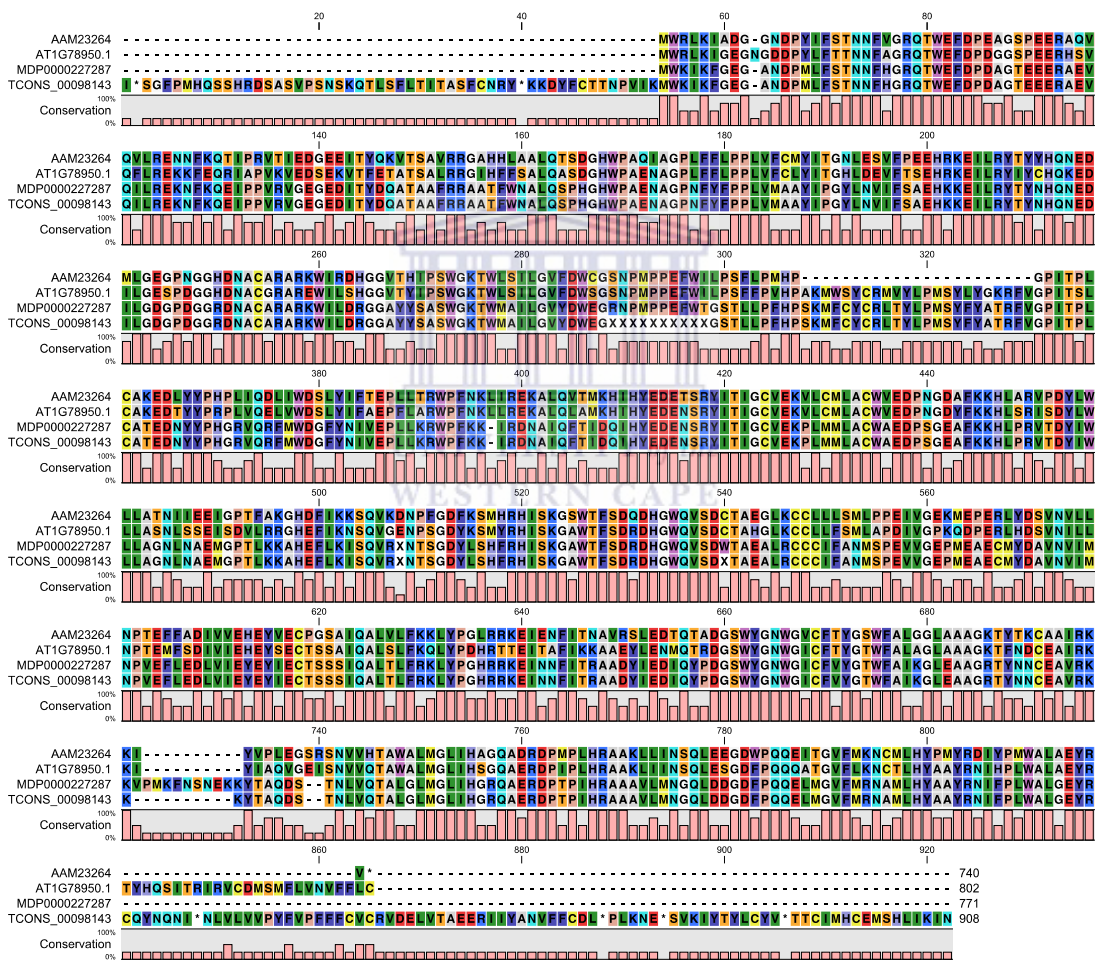
Figure 3. 11 Verification and annotation feature improvement of apple genome. A, Tracks of annotation features from genome-guided (Ab and Gm) and *de-novo* (Trinity) reconstructed transcripts were compared to *M. x domestica* reference gene models on IGV. B, amino acid alignment of *glycine max* (AAM23264), *Arabidopsis (AT1G78950.1)*, *M. x domestica* (MDP0000227287) and Ab reconstructed *ß-amyrin synthase* encoding transcript (TCONS00098143)

3.2.7   Transcript annotation

Identifying transcripts and assigning them possible function, is important in understanding physiological processes in which the putative proteins may be involved. Using the pooled transcript data set from the combined genome-guided and *de novo* assemblies, functional groups were assigned to a total of 69 and 71% Ab and Gm combined transcripts data sets respectively using the Munich Information Centre for Protein Sequences (MIPS) (Mewes*, et al.*, 2008) (Figure 3.12 A, B), as described in sections 2.2.13. However, 13 and 21% of the transcripts with annotations in Ab and Gm data sets had no classification, while 31 and 29% for Ab and Gm transcripts had no hits to any of the databases (TAIR, *Malus* ESTs, *Malus* CDS, peach, rice, *Malus* PlantGDB, populus, grape, swiss-prot and the non-redundant (nr) protein database (NCBI) used(Figure 3.12A, B). Transcripts that did not have any hits to the used databases could be classified as 'novel', requiring further elucidation in motif and domain composition, hence deducing their possible function. Figure 3.12 shows little variation in the count and proportion of identified functional categories and annotations of the Ab and Gm reconstructed transcripts, suggesting consistency in transcript reconstruction and annotation. The global analysis of the reconstructed transcripts showed enrichment of transcripts encoding protein with binding function, metabolism, and protein fate as the three major functional categories in the combined apple fruit peel and pulp transcriptome.

The nucleotide and protein databases are dynamic databases, which increase in complexity and volume with addition of new sequences. These "No-hit" transcripts constituted part of *M. x domestica*'s yet to be identified gene pool. The *Arabidopsis* (TAIR) database is well curetted and continues to be improved. However, limiting

functional classification of the apple transcripts to the TAIR accession equivalents might explain the high number of transcripts in the unclassified as well as 'No Hit' categories. A substantial number of genes have been classified as unique to *M. x domestica* (Troggio*, et al.*, 2012; Velasco*, et al.*, 2010) making it difficult to annotate such transcripts by similarity, more so when the category range is reduced to a single TAIR database.
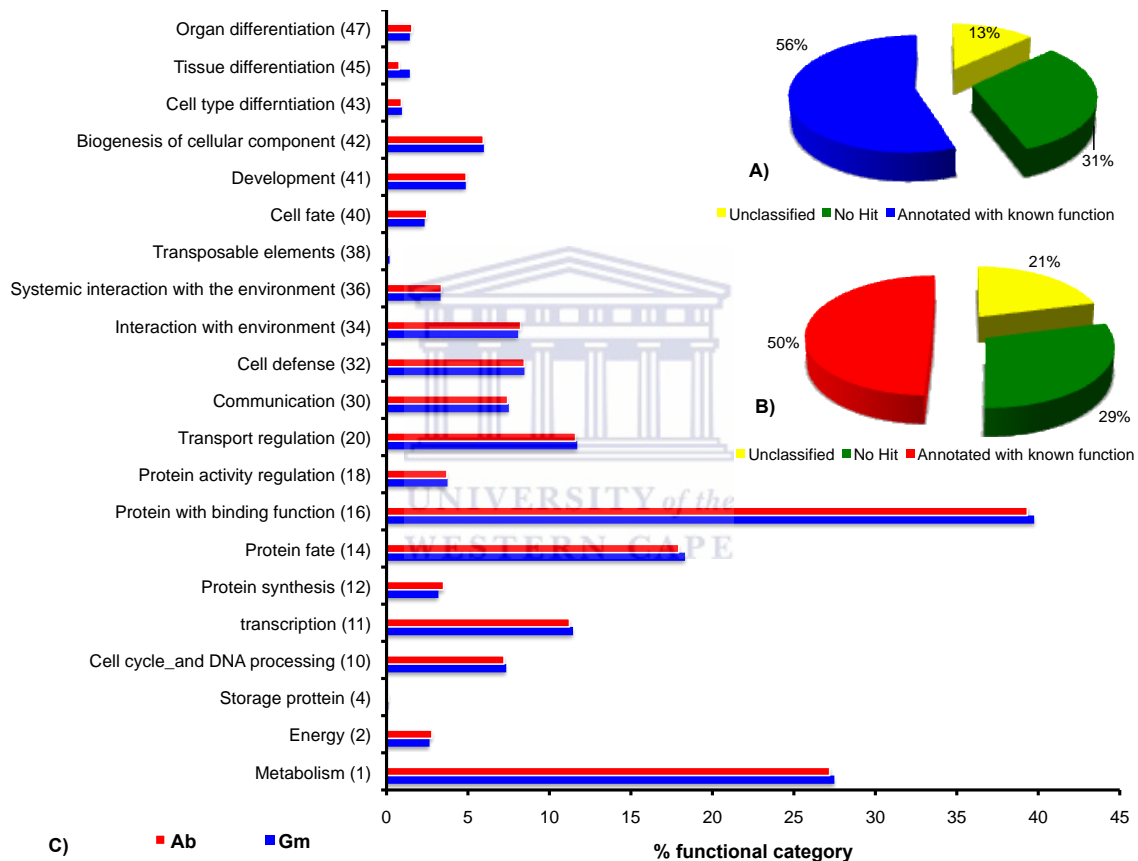
Figure 3. 12 Annotation and functional categories of transcripts reconstructed by Ab and Gm combined with their respective Trinity assemblies from reference unmapped reads. The transcripts sets from Ab and Gm were annotated using BLASTn/x. TAIR database accession equivalents were then assigned to the transcripts. A and B are pie chart plots of percent annotated transcripts of Ab and Gm assembly approaches respectively. C) Distribution of Unigenes according to the Munich Information Centre for Protein Sequences (Mewes*, et al.*, 2008). Number in parenthesis following category names indicates the MIPS number of each category.

3.2.7.1 Un-annotated transcripts codes for functional gene space (domains)

The 76479 un-annotated transcripts were translated into 20806 amino acid sequences with a minimum length of 100 amino acid residues. These amino acid sequences had hits to 98 PFAM domains. The distribution of the Pfam domains among the 20806 protein sequences, revealed that leucine rich repeat (LRRs), WD domain G-beta repeat, tetratricopeptide repeat, protein kinase domain, zinc knuckle, zinc finger, ABC transporter, ATPase family, RNA recognition motif and ankyrin repeat domains were the ten most common (Figure 3.13).

LRRs are 20-29-residue sequence motifs, for example spliceosomal protein U2A' and Rab geranylgeranyltransferase, which primarily function to provide a versatile structural framework for the formation of protein-protein interactions (Lewis and Alessi, 2012; Torti*, et al.*, 2012). WD-repeat proteins function in coordinating multi-protein complex assemblies, where the repeating units serve as a rigid scaffold for protein interactions. Examples of such complexes are E3-ubiquitin ligase and G proteins.

ATP-binding cassette (ABC) transporters are transmembrane proteins that utilize the energy of ATP hydrolysis for translocation of various substrates across membranes and non-transport-related processes such as translation of RNA and DNA repair (Fukuda*, et al.*, 2011; Krumpochova*, et al.*, 2012; Zolnerciks*, et al.*, 2011). ATPases exert their activity through the energy-dependent remodelling or translocation of macromolecules. They are involved in a range of processes, including DNA replication, protein degradation, membrane fusion, microtubule severing, peroxisome biogenesis, signal

transduction and the regulation of gene expression (Muench, *et al.*, 2011; Pedersen, *et al.*, 2012). The presence of these domains suggests that there are multiple physiological processes the identified proteins may be involved in during fruit development.

Legend:
- Leucine Rich Repeat
- WD domain, G-beta repeat
- Tetratricopeptide repeat
- Protein kinase domain
- Zinc knuckle
- Zinc finger, C3HC4 type (RING finger)
- ABC transporter
- ATPase family associated with various cellular activities (AAA)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Ankyrin repeat
- others
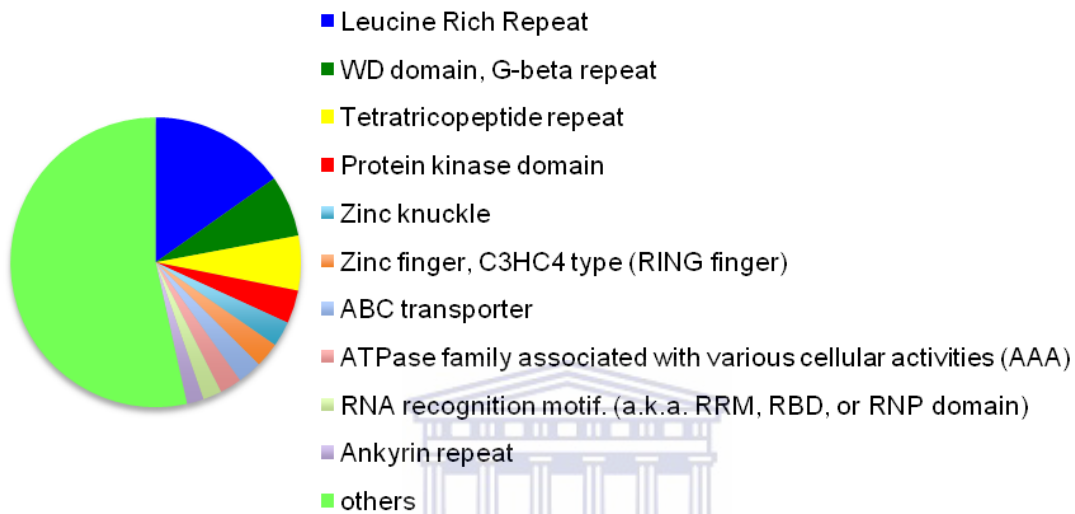
Figure 3. 13 Domain annotation distribution of un-annotated transcripts. PFAM domains of the un-annotated transcripts were obtained by submitting amino acid sequences of the transcripts to PFAM protein domain database on CLC-Bio Genomics Workbench 5.5.1. Enrichment of each domain was calculated as the total count of domain across all submitted transcripts. Translation to amino acid sequences was done using ESTScan.

3.2.7.2 Chloroplast and mitochondrial genomic DNAs have sequence regions homologous to nuclear DNA reconstructed transcripts.

Chloroplast and mitochondria organelles evolved from a cyanobacterial endosymbiont. Subsequent co evolution of the plastid organelles and the nuclear genomes produced organelles that are eubactarial at their core but with extensive plastidic-specific embellishments (Ajjawi, *et al.*, 2010; Barkan, 2011). The chloroplast genome has genes falling into three categories: those encoding 1) components of the chloroplast gene expression machinery (RNA polymerase, ribosomal proteins, tRNAs, and rRNAs), 2)

105

subunits of photosynthetic enzymes (Rubisco, PSII, the cytochrome b6f complex, PSI, the ATP synthase, and the NADH dehydrogenase), and 3) proteins involved in other metabolic processes (Ajjawi, *et al.*, 2010; Barkan, 2011).

On the other hand plant mitochondria genome has been reported to contain chloroplast and nuclear DNA sequences. Mitochondrial gene sequences have also been identified in the nucDNA as fragments or entire genes. The mitochondria genome encodes for protein expression machinery as well as respiratory machinery (ATP synthase, cytochrome bp, NADH dehydrogenase, cytochrome c oxidase, as well as cytochrome c biogenesis) (Gray, 2012; Liere, *et al.*, 2011). However, the proteome of the mitochondria and chloroplast has revealed a complexity of several proteins predominately nuclear gene products, which are synthesised in the cytosol and imported into the organelles (Schröter, *et al.*, 2010). As a result, mitochondria and chloroplast gene expression and regulation consist of proteins that are derived from physically separate genetic systems.

The use of nucDNA gene products by either the chloroplast or mitochondria as well as used of cpDNA and/or mtDNA genes products by the cytoplasm has been documented (Zhang, *et al.*, 2011). This exchange of gene products to execute physiological functions is called 'cross talk'. To understand the existence of nuclear-organellar cross talk, organellar-nuclear co evolution and gene transfer, 186498 nuclear DNA encoded reconstructed transcripts were mapped to *M. x domestica* chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA) genomes as outline in section 2.2.12. Of the 186498 transcripts reconstructed from poly-A enriched cDNA libraries prepared from peel and

pulp 'Golden Delicious' fruit tissues, 282 mapped onto 69 cpDNA loci with 149 unique transcripts (Table 3.1). Twenty-seven of the 149 unique transcripts had no hits to NCBI and Swiss-prot protein databases, and the remainder mapped to 47 unique UniprotKB accessions, for proteins involved in photosynthesis, disease resistance, membrane proteins, DNA-directed polymerase and ribosomal proteins.

Table 3. 1 Transcripts reconstructed from mRNA map to chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA)

	Mapped to cpDNA	Unmapped to cpDNA	Mapped to mtDNA	Unmapped to mtDNA
*Ab	282	186214	546	185950
*Gm	280	186216	530	185966
ᐃTrinity	170	127304	499	126975

*Ab and *Gm were a combination of transcripts reconstructed by mapping to apple genome v1.0 pseudo molecules then *de novo* assemble the unmapped reads with Trinity. ᐃTrinity, transcripts reconstructed by *de novo* assembly using Trinity from pooled RNA-seq data from peel and pulp developing 'Golden Delicious' fruit.

Five hundred and forty six nucDNA encoded transcripts (Table 3.1) mapped to 319 loci on mtDNA, 96 of the loci had no hits to UniprotKB database accessions and the remaining 223 loci had 53 unique UniprotKB accession hits. The 53 unique UniprotKB accessions mapped to six KEGG Orthology (KO) terms involved in oxidative phosphorylation, two-component system, and one carbon pool by foliate, nitrogen metabolism, purine metabolism and biosynthesis of secondary metabolism.

Nuclear DNA transcripts show homology to sequence regions on the cpDNA

Figure 3.14 depicts regions on cpDNA homologous to the transcripts reconstructed from poly-A enriched mRNA. The grey box colourations represent homologous regions between cpDNA and the nucDNA-encoded transcripts with unknown function. Coloured boxes are homologous regions with known function (Figure 3.14). Homologous regions were observed on both strands, but a small region was left which had no transcripts homologous to it (Figure 3.14).

Eighty % of the transcripts mapping to cpDNA had known function. The cpDNA mapped transcripts showed that membrane protein (*ycf*, 18%), DNA-directed RNA polymerases (*rps*, 17%), NAD(P)H-quinone oxidoreductases (*NAD*, 11%), cytochrome b6 (5%), ribosomal protein (*rpo*, 5%) and Rubisco large subunit (*rdcL*, 2%) were the main gene groups common between nuclear DNA (nucDNA) and cpDNA (Figure 3.15). The ycf membrane proteins facilitate communication and transport of molecules between chloroplast and cytoplasm (Börner*, et al.*, 2011; Peter*, et al.*, 2011; Yruela and Contreras-Moreira, 2012). The existence of RNA-polymerases and ribosomal protein gene sequences on the cpDNA, suggests that chloroplasts are semi-independent organelles. These organelles are capable of self-replication and encode some of the essential genes (Yruela and Contreras-Moreira, 2012).

It has been established that the major carboxylating enzyme, Rubisco, has two subunits, one coded by a nuclear gene *rbcS* for the small subunit, and one by the chloroplast gene *rbcL,* for the large subunit (Suzuki and Makino, 2012).
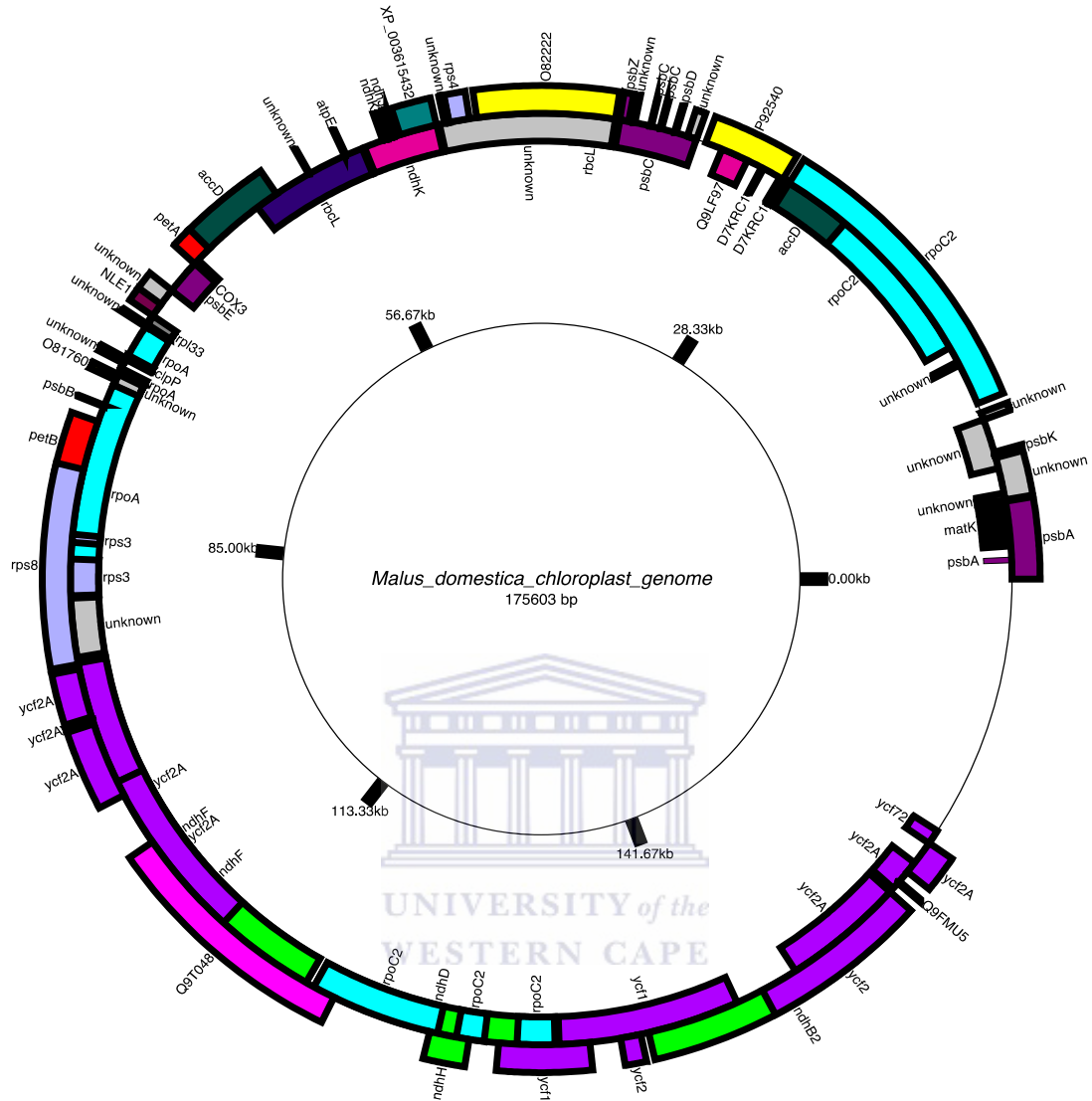
Figure 3. 14Schematic localisation of genes mapped to *M. x domestica* chloroplast genome common with *M. x domestica* nucDNA. Direction of transcription is in the counter-clockwise direction starting from the right, in both negative and plus strands. Reference and *de novo* assembled transcripts were mapped to the mtDNA using GMAP and the gene annotations were used to plot the annotated cpDNA map using an online GenomeVx programme. Unknown = regions in grey denote nucDNA-derived transcripts that had no hit to publicly available databases. Genes common between cpDNA and nucDNA: COX=cytochrome C oxidase, NLE=non-LEE effector, XP_003615432=hypothetical protein, acc=acetyl CoA carboxylase, accD= acetyl CoA carboxylase beta subunit, atp=ATP synthase, clpP=ATP dependant Clp protease proteolytic subunit, mat=maturase, ndh=nadh dehydrogenase, pet=cytochrome b6/f complex protein, psb=photosystem II reaction centre protein, rbcL=rubisco large subunit, rpl=large subunit ribosomal protein, rpo=ribosomal protein, rps=small subunit ribosomal protein, ycf=uncharacterized open reading frame.

At any given time during photosynthesis, large quantities of Rubisco are required for carboxylation. Absence of *rbcS* sequences on cpDNA, suggests the existence of cross-

109

talking mechanisms between cpDNA and nucDNA to bring the two components together to form a functional Rubisco enzyme. A mechanism known as control by epistasy of synthesis (CES) has been suggested to coordinate the synthesis of subunits of each photosynthetic enzyme complex via negative feedback loops that are triggered by specific unassembled subunits 1(Barkan, 2011; Stern, 2009).
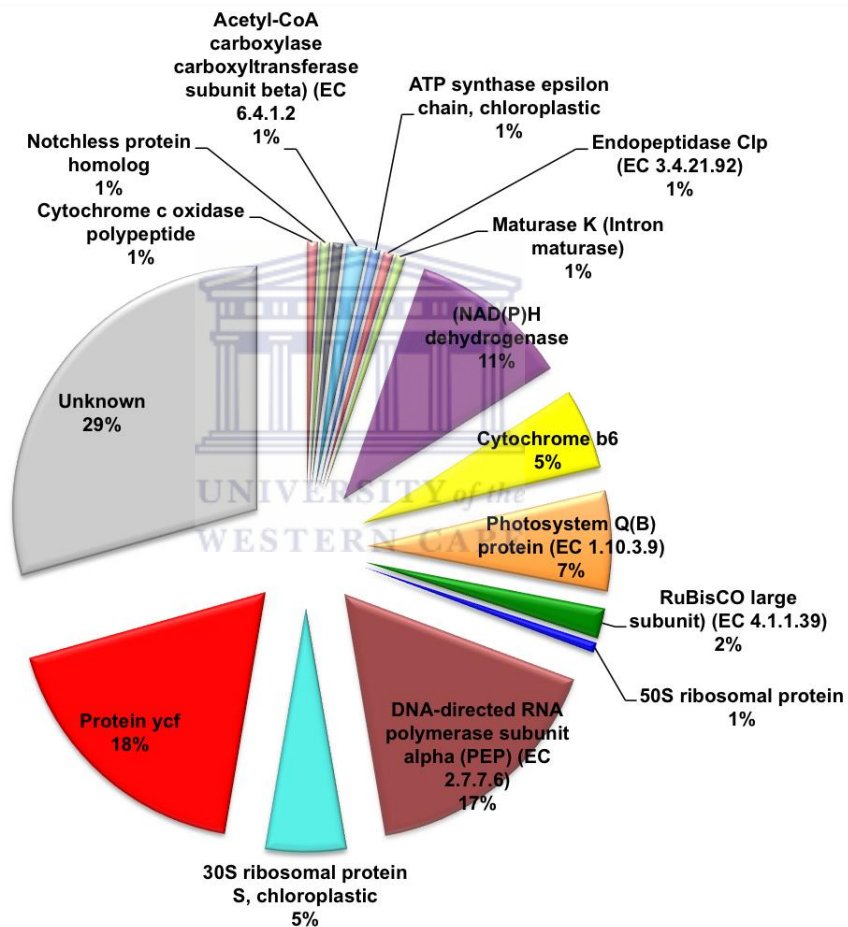
Figure 3. 15 Distribution of genes mapped to cpDNA common with nucDNA. Unknown denote nucDNA-derived transcripts that had no hit to publicly available databases. Nuclear DNA derived transcripts were mapped to the cpDNA using GMAP and the structure of the mapped transcripts captured in an exon coordinate format (gff). The gene annotations were obtained by local BLASTn/x similarity searches against nr NCBI databases.

Nuclear DNA transcripts show homology to sequence regions on the mtDNA

A large number of the transcripts that mapped to mitochondrial DNA (mtDNA, 68%) encoded for proteins with unknown function (Figure 3.16, 3.17). Transcripts encoding NADH-ubiquinone oxidoreductase (NAD) were the largest group of transcripts with known function common between nucDNA and mtDNA (Figure 3.16).

NAD enzymes are located in the inner mitochondrial membrane and constitutes complex 1, a component of the enzymes that catalyses the transfer of electrons from NADH to coenzyme Q (CoQ), a process known as oxidative phosphorylation (Bridges*, et al.*, 2011; Treberg*, et al.*, 2011). Cytochrome C oxidase (COX) encoding transcripts were however few (1%, Figure 3.17). The observation suggests that COX and NAD exist in mtDNA as unequal copies. It does not however, reflect the potential expression levels of each locus. High levels of enrichment of the two 'electron transport chain' complex proteins, COX and NAD, are characteristic of the oxidative function of the mitochondria.

Cytoplasmic male sterility (cmsT*)* in maize is a well-studied example of mtDNA and nucDNA cross talk, which illustrates the intricacies of nuclear and mitochondrial genome interactions (Chase, 2011; Zhang*, et al.*, 2011). The mitochondrial gene *cmsT* causes male sterility due to inhibition or prevention of normal stamen development, possibly as a result of failure by the mitochondria to supply enough energy to the rapidly developing stamen tissue (Zhang*, et al.*, 2011).
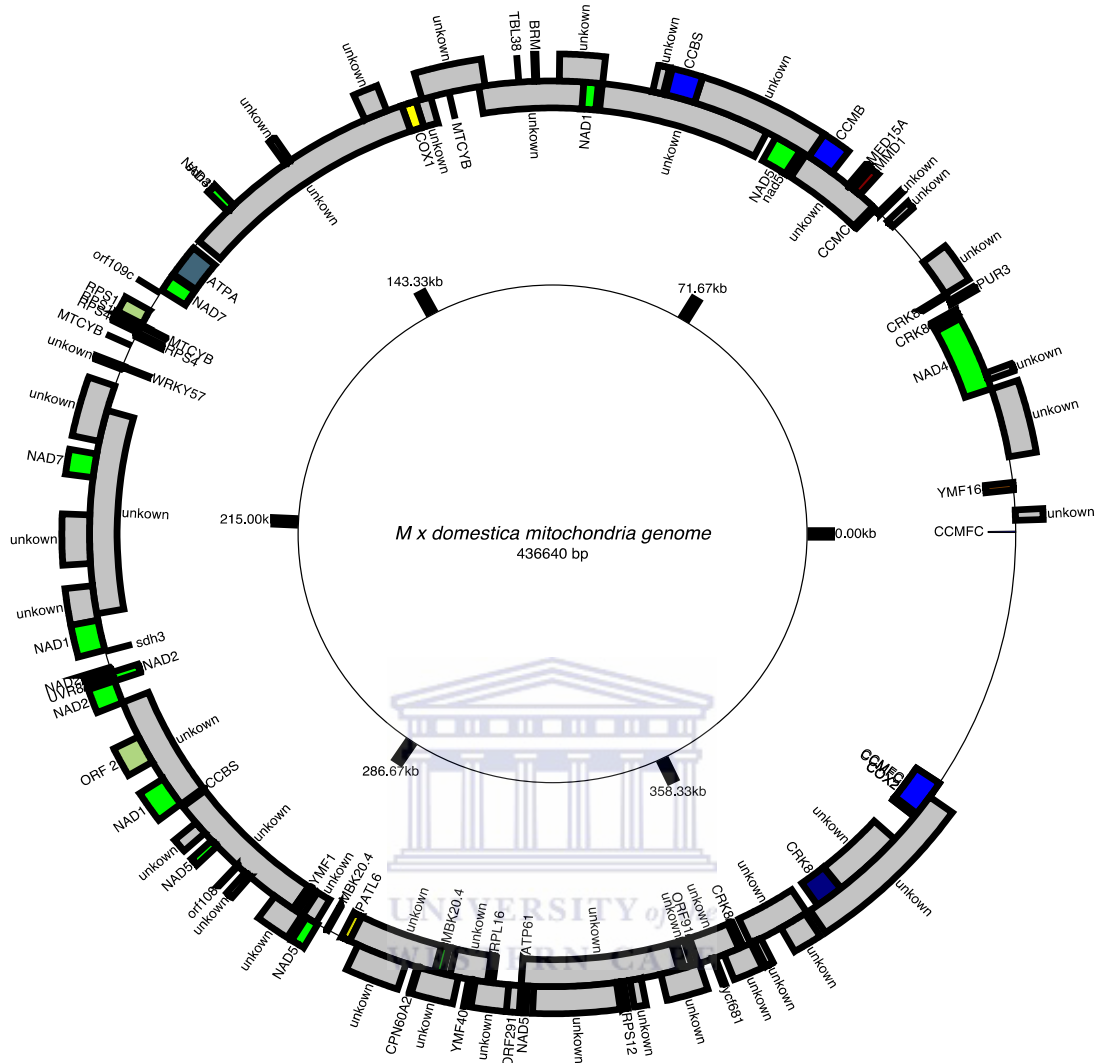
Figure 3. 16 Schematic localisation of genes mapped to *M. x domestica* mitochondria genome common with *M. x domestica* nucDNA. Direction of transcription is in the counter-clockwise direction starting from the right, in both negative and plus strands. Reference and *de novo* assembled transcripts were mapped to the mtDNA using GMAP and the gene annotations were used to plot the annotated mtDNA map using an online GenomeVx programme. Unknown = regions in grey denote nucDNA-derived transcripts that had no hit to publicly available databases. Genes common between mtDNA and nucDNA: Cytochrome C oxidase 2 (COX2), coenzyme Q (CoQ), NADH-ubiquinone oxidoreductase 2 (NAD2), UVR8=ultraviolet-B-receptor, WRKY57=transcription factor, YMF=ATP synthase, orf175=open reading frame 175 unidentified, rps4=small subunit ribosomal protein 4, ycf68=uncharacterised protein 68, ATP6-1=ATP synthase 6, ATPA=ATP synthase, BRM=Brahma protein, CCM=cerebral cavernous malformations, CPN60A2=chaperonin 60 alpha 2, CR8=cisitimine-rich receptor-like protein 8, MED15A=mediator of RNA polymerase II transcription subunit 15A, MMD1=maintenance of mitochondrial DNA 1, MTCYB= mitochondria cytochrome b, MTR_1g00g300=mitochondrial protein putative, RPL16=large subunit ribosomal protein 16, RPS1=small subunit ribosomal protein 1.

Figure 3. 17 Distribution of genes mapped to mtDNA common with nucDNA. Unknown denote nucDNA-derived transcripts that had no hit to publicly available databases. Nuclear DNA derived transcripts were mapped to the mtDNA using GMAP and the structure of the mapped transcripts captured in an exon coordinate format (gff). The gene annotations were obtained by local BLASTn/x similarity searches against nr NCBI databases. Distribution of the gene counts was plotted in Excel. Cytochrome C oxidase 2 (COX2), coenzyme Q (CoQ), NADH-ubiquinone oxidoreductase 2 (NAD2), UVR8=ultraviolet-B-receptor, WRKY57=transcription factor, YMF=ATP synthase, orf175=open reading frame 175 unidentified, rps4=small subunit ribosomal protein 4, ycf68=uncharacterised protein 68, ATP6-1=ATP synthase 6, ATPA=ATP synthase, BRM=Brahma protein, CCM=cerebral cavernous malformations, CPN60A2=chaperonin 60 alpha 2, CR8=cisitimine-rich receptor-like protein 8, MED15A=mediator of RNA polymerase II transcription subunit 15A, MMD1=maintenance of mitochondrial DNA 1, MTCYB= mitochondria cytochrome b, MTR_1g00g300=mitochondrial protein putative, RPL16=large subunit ribosomal protein 16, RPS1=small subunit ribosomal protein 1.

This study has shown the existence of five gene transcripts common among cpDNA, mtDNA and nucDNA (Figure 3.18). The existence of similar gene copies in any one of these organelles may not in itself reflect that the genes are actively expressed in these organelles. The regulation of quantity and time of expression of these common genes

could require intricate regulation networks, which may or may not be dominated by nucDNA.
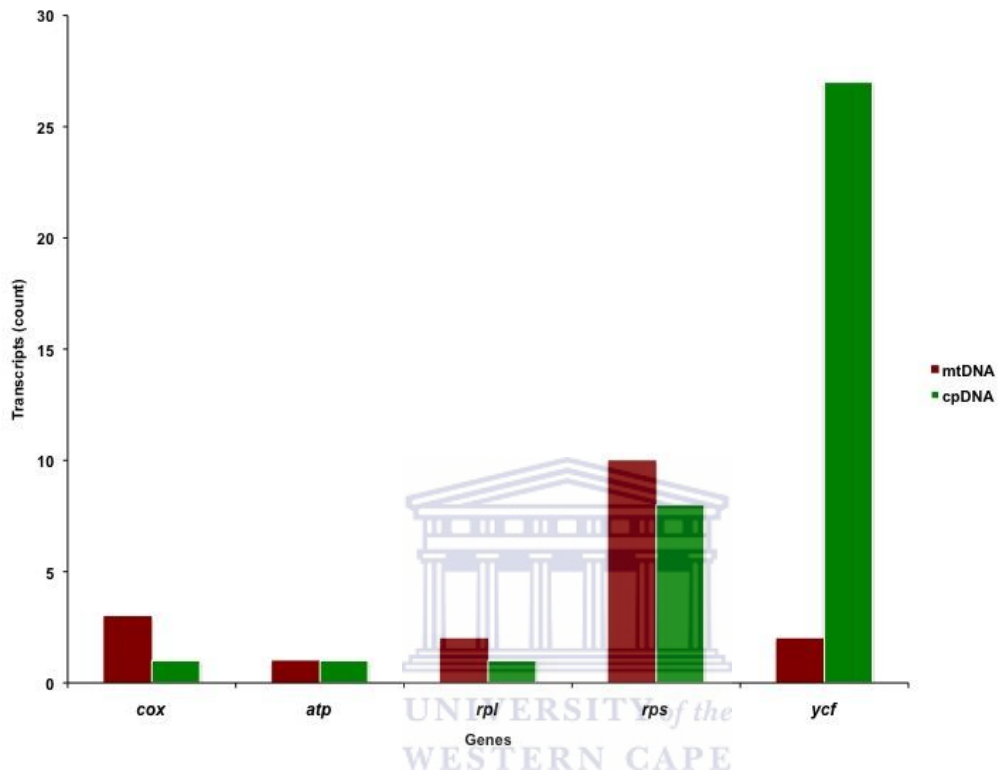
Figure 3. 18 Distribution of genes mapped to cpDNA and mtDNA that were common with nucDNA. Nuclear DNA derived transcripts were mapped to the cpDNA ant mtDNA using GMAP. The counts of the genes common among cpDNA, mtDNA and nucDNA were obtained by counting the genes that mapped to the cpDNA as well as the mtDNA and their distribution was plotted in Excel. The gene annotations were obtained by local BLASTn/x similarity searches against nr NCBI databases. Cox= Cytochrome C oxidase, atp=, rpl=large subunit ribosomal protein, rps=small subunit ribosomal protein, ycf=uncharacterized open reading frame, atp= ATP synthase.

The proportions of genes shared among cpDNA, mtDNA and nucDNA shown in Figure s 3.15, and 3.17, might not be representative of the possible full repertoire of the apple transcriptome.

It is feasible to speculate that transcripts that mapped to the cpDNA and mtDNA could have been assembly artefacts or cpDNA /mtDNA contaminants. However, transcripts

114

with homologous regions on the cpDNA/mtDNA were in this study were assigned physical locations in the apple genome when mapped to the genome. Also, when aligned against the genome, cpDNA and mtDNA had fragmented multiple hits all over the genome. This observation suggested that there were no miss assemblies in the genome used, and shows evidence of genetic transfer. The transcripts might not have come from the chloroplast genome, since chloroplast expressed genes are not poly-A tailed. As such a poly-A enrichment strongly selects for nucDNA-expressed genes. Hence the identified homology regions to the nucDNA expressed transcripts gives testimony to an on going process of horizontal gene transfer among nucDNA, cpDNA and mtDNA genomes.

Knowledge of the relationship between nucDNA and the plastid genomes will aid in optimising transgene expression in the plastid compartments.
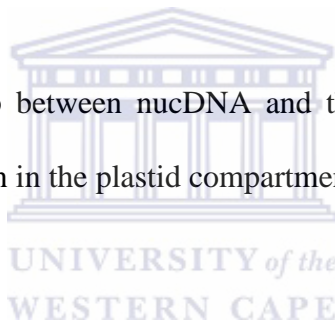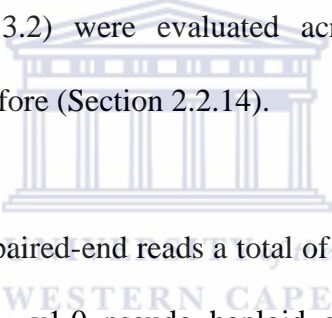
3.2.8 Transcript abundance estimation

All sampled time points had biological replicates, with each replicate pooled from ten different fruits sampled around the tree canopy. To evaluate the uniformity of the expression estimates as calculated by RSEM (section 2.2.15) six transcripts from biological replicates at daa25 time point were compared (Appendix 8.9). The comparison shows that the transcript expression level estimates were almost the same across the replicates and the six transcripts randomly selected to represent the up and down regulation. As such the expression estimates are acceptable.

3.2.8.1 Post read mapping treatments affect mapped read count

The sensitivity of any algorithm in calculating and estimating transcript abundance is mostly affected by ribosomal RNA (rRNA), PCR duplicates, and total read count uniformity (Batut, *et al.*, 2012; Jones, *et al.*, 2012). The ability of RNA-seq algorithms in correctly estimating the expression values of the transcripts is invaluable in deducing the correct expression profiles of a particular gene as well as the physiological processes influenced by such expression levels. The effects of read de-duplication, read sub-setting, and ribosomal read removal in estimating the expression values of transcripts reconstructed by Ab (Table 3.2) were evaluated across six biological conditions replicated twice as described before (Section 2.2.14).

From the pulp tissue extracted paired-end reads a total of 235 359 430 unique reads were mapped to the *M. x domestica* v1.0 pseudo haploid genome molecules (Table 3.2). However *in silico* rRNA treatment discarded 10% of the raw mapped reads as reads belonging to rRNA transcripts (Table 3.2). De-duplicating the mapped reads using external coordinates resulted in a significantly higher (48%) raw mapped read count being discarded than *in silico* rRNA treatment (Table 3.2). In establishing a uniform mapped read mass within a treatment, a significant proportion of the initial raw mapped read count was discarded. In a similar trend to the effect of the *in silico* rRNA treatment on raw read count, de-duplication of the subset reads resulted in the lowest unique read count being retained for differential expression analysis (Table 3.2).

116

Table 3. 2 Effect of *in-silico* rRNA depletion and de-duplication treatment on mapped read count

	Mapped read mass	% Ribosomal reads	% Duplicate reads
Raw mapped read count	235 359 430	(10%)	(48%)
Subset mapped read count	74 983 918	(1%)	(46%)

Read one (/1) and two (/2) of a unique pair were counted as a single read. Comparisons of the effects of the treatments on read count were evaluated on raw mapped read count as well as after sub-setting the reads to a uniform count within a treatment.

3.2.8.2 Post read mapping treatments affect transcript abundance (in FPKM) estimates

The effects of the various post-read-mapping treatments on transcript abundance estimations were evaluated across six harvesting time points, and the treatment effects resulted in a decrease in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) estimates (Figure 3.19). De-duplicating, rRNA removal, and sub-setting all affect the total read coverage per exon, thus affecting the calculations of FPKM values below threshold coverage (default for Tophat/Cufflinks =10 reads). This section will attempt to show the effect each of the treatments have on the calculated FPKM values.

De-duplicating mapped reads with identical mapping coordinates significantly reduced and showed the lowest transcript mean FPKM estimates across the six harvesting time points. Removing ribosomal mapped reads resulted in adjusted FPKM values, but was not significantly different from "No-treatment" (Figure 3.19). The absence of a significant difference in the effects of rRNA treatment and "No-treatment" (Figure 3.19)

suggested the presence of low levels of possible rRNA read contaminants, corroborating the observations in Table 3.2. Removal of rRNA reads resulted in a decrease in the number of transcripts with more than zero FPKM values (Figure 3.19). Ribosomal treatment resulted in about 1000 more transcripts with FPKM equal to zero compared to "No-treatment" and read de-duplication treatments (Figure 3.19B). Transcripts, which had an FPKM value equal to zero were considered not supported by any reads, hence were regarded as not expressed.

The apple genome v1.0 was sequenced from diploid domesticated apple, 'Golden Delicious' cultivar, whose genome is considered to have undergone recent (>50 million years ago) genome-wide duplication (GWD) resulting in the transition from nine ancestral chromosomes to 17 chromosomes in the Pyreae (Troggio, *et al.*, 2012; Velasco, *et al.*, 2010). As a result of the possible small differences in the two alleles, reads from such different maternal and paternal origin alleles at a particular locus are expected to map to the same genome coordinates and may be filtered out as duplicates, disregarding the existence of the two alleles and the slight variations in base composition (Martin and Wang, 2011). Reads that are also generated from repetitive regions are also difficult to map uniquely and may be filtered out as duplicates. The existence of duplicate reads by itself might not suggest the presence of over amplified PCR products. Transcripts might exist in the cell as multiple copies in an attempt to meet the protein demand at the particular phase of cell growth. Such high copy transcripts have high chances for part of their sequences to existing as duplicates after transcript fragmentation. Regarding such

reads in the same way as PCR duplicates distorts the high expression values of high copy number transcripts.
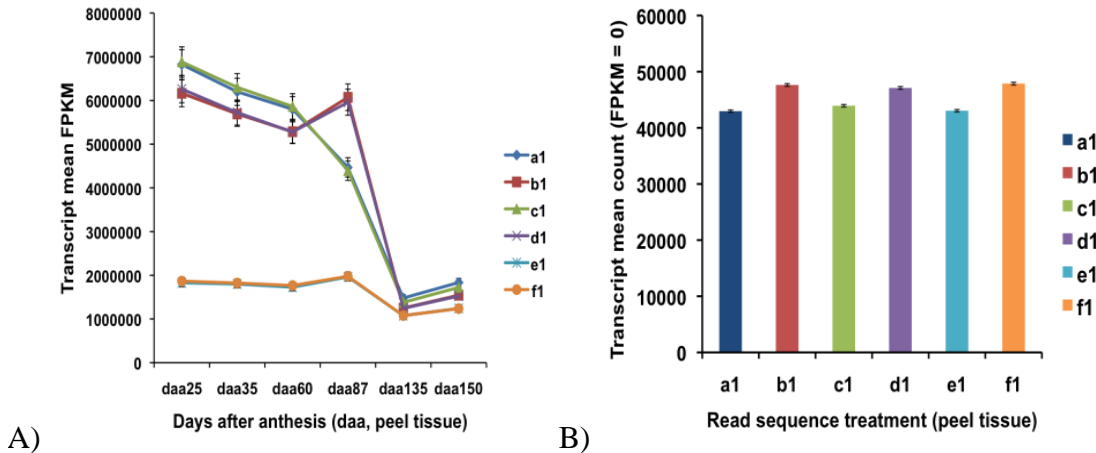
Figure 3. 19 Effect of read post mapping treatment on transcript abundance estimation. Effect of read de-duplication, ribosomal read and sub-setting treatments on the FPKM values of apple fruit peel extracted transcripts, A) per harvesting time point, and B), shows changes in the mean count of transcripts with FPKM = 0 values across the post mapping treatments. Transcripts abundance estimations were done using Cuffdiff 2.0.2 (http://www.cufflinks.cbcb.umd.edu), a1="No-treatment"; b1=sub-setting; c1=Ribosomal treatment; d1=Ribosomal treatment with sub-setting; e1=Read de-duplication; f1=Read de=duplication with sub-setting. Treatment effect p-value = 0.015, Daa effect p-value = 0.001.

Sequencing errors have been documented and are higher in certain platforms than others (Grabherr*, et al.*, 2011; Martin and Wang, 2011; Trapnell*, et al.*, 2012). Such platform specific, as well as other technically induced errors results in small differences in read sequences, which were otherwise supposed to be PCR duplicates. Thus, errors in transcript abundance estimation are introduced when such reads are retained. A possible solution would be to allow for few mismatches and/or pre-cleaning the reads. However, allowing mismatches results in different and non-PCR duplicate reads being considered as PCR duplicates, while read trimming might group different read sequences with few mismatches as PCR duplicates.

Several studies have alluded to such possible effects in the attempt to refine transcript abundance estimation (Grabherr*, et al.*, 2011; Martin and Wang, 2011; Trapnell*, et al.*, 2012). Several reports have suggested that PCR duplicates are better handled when longer and paired-end reads are used (Grabherr*, et al.*, 2011; Martin and Wang, 2011; Trapnell*, et al.*, 2012).Nonetheless the complexity of the genome has to be factored in to consider genome duplications, size and polyploidism. Careful choice of the algorithm for de-duplication is also invaluable. There is software that has been used for read de-duplication, but the preferred ones are the ones that use mapping coordinates, e.g. Picard. It is agreed that read de-duplication is not straightforward and where possible should be avoided (Martin and Wang, 2011). In an attempt to mitigate accumulation of PCR duplicates, use of minimum PCR cycles could be adopted when working with un-limiting mRNA quantities.
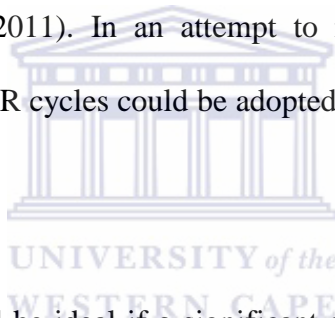
De-duplication of reads, would be ideal if a significant amount of PCR duplicates were expected, resulting from too many PCR cycles (Martin and Wang, 2011). The protocols used in this study however, used 10 PCR cycles, contrasting PCR cycles of 18-25 cycles for library enrichment when previous versions of Illumina kits were used.

Sub-setting of the mapped reads to a uniform read count as set by the lowest sample read count did not differ much for each corresponding treatment (Figure 3.19A), but resulted in significantly high count of transcripts with FPKM equal to zero(Figure 3.19B). This result suggested that by sub-setting read count the coverage per transcript was reduced possibly to below threshold of 10 reads for Tophat/Cufflinks, hence the number of

transcripts that could be reconstructed or called as being significantly expressed was significantly reduced.

According to Trapnell, *et al.,* (2012) the Cufflinks algorithm was designed to be able to normalise for read count, transcript length, highly and lowly expressed transcripts, thus eliminating the need to sub-set the mapped, reads. For the purposes of continuity, *in-silico* rRNA treated FPKM values were adopted as the accepted expression values of the reconstructed transcripts.

3.2.9 Transcripts abundance estimation using RT-qPCR.

In this study RT-qPCR was adopted to estimate and validate the expression levels of five genes, *anthocyanidin synthase (ANS); glucose-1-phosphate adenylyltransferase (ADG); NAD-dependent sorbitol dehydrogenase (SDH); pectate lyase A10 (PL); sorbitol transporter 5 (SOT5)* normalised to the geometric mean of *ß-actin* and *tubulin* reference genes. Reference genes and geometric mean normalisation steps and calculations are described in section 2.2.16.2.

These genes were expressed throughout fruit development and selected to represent changes in carbohydrate and pigment accumulation as the fruit developed. The primer sequences designed using NCBI-Primer-blast for the five target genes and two reference genes are as shown in Table 3.3.

Normalising of the expression values using individual *ß-actin* and *tubulin* reference genes revealed that *ß-actin* reference gene normalisation over-estimated the expression values of SDH, SOT5 and ADG, while underestimating those of ANS, and PL.

Table 3. 3 Primers used for RT-qPCR

Gene name	Strand	Sequence	Genebank Accession
SOT5	F	TCC CTC TCC AGG AAG ATG GCT GA	AB125648.1
	R	TGG GCT TCT GTG GGG GAT CAA	
ANS	F	GCC GGC AAT GTT CCC ACC AC	AF117269.1
	R	GCA CAA CCC GCT TCA CTT GGG	
SDH	F	CAA GGG CGG CCG CTA CAA TC	AY244807.1
	R	CAG ATC CGC GGG GTG CAC AA	
PL	F	TCA CCT CTT TGC TGG ATC TG	GO538232.1
	R	GCA GAC CCT ACA ATC AAT AGC C	
ADG	F	AGG CTG ACC CCA AAA ATG TGG CT	GO558882.1
	R	ATC GGA ACC GCC GGC TTG G	
*ß-Actin**	F	TTC CTG ATG GAC AGG TTA TTA CG	CV883188.1
	R	GAT GGT TGG AAT AGA GCC TCA C	
*Tubulin**	F	GGT CTT CTG GGC AAA CCA AC	GO568904.1
	R	GCC ACC TAC TGT TGT CCC AG	

F = forward; R = reverse strand; Genes used as references are marked *, Primers were designed using NCBI-Primer-blast, spanning exon-exon junctions were possible so that only targeted mature mRNA will be amplified.

On the other hand *tubulin* normalisation underestimated the expression values of SDH, SOT5, and ADG, while overestimating those of ANS and PL (Figure 3.20). The result suggested variation in the abilities of the selected reference genes in gene expression levels estimation.

In order to disperse the negative effect of such reference gene variations on the calculated gene expression values, a geometric mean of the two genes was calculated and used to normalise the sample gene expression values (Vandesompele*, et al.*, 2002). The

geometric mean normalisation approach showed moderation of the variation in the expression values estimated by either *tubulin* or *ß-actin* reference genes (Figure 3.20).
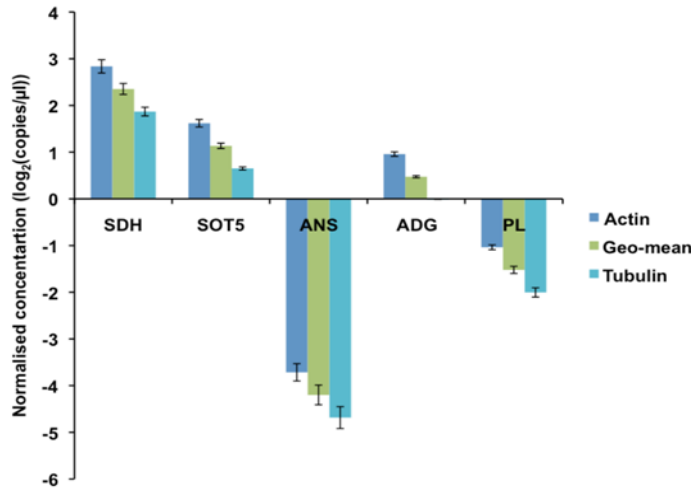
Figure 3. 20 Effect of target gene normalization using single and multiple reference gene geometric means. The graph illustrates fold changes of RT-qPCR calculated expression levels of five genes, *anthocyanidin synthase (ANS), glucose-1-phosphate adenylyltransferase (ADG), NAD-dependent sorbitol dehydrogenase (SDH), pectate* lyase A10 (PL), and sorbitol transporter 5 (SOT5) in 'Golden Delicious' fruit pulp at daa87. The RT-qPCR expression estimates were normalized to *β-actin* and *tubulin* reference genes as well as the geometric mean of *β-actin* and *tubulin* reference genes. Fold change in gene expression estimates were obtained by dividing the estimate at each time point by the estimate at daa25 for each, then log_2 transformed and plotted in a histogram.

Choice of a 'good' reference gene in RT-qPCR has been regarded as a determinant of the accuracy of the estimated gene expression levels (Vandesompele*, et al.*, 2002). A 'good' reference gene is one that is constitutively expressed, without a change in expression levels in different tissues and environmental conditions. However, such an ideal reference gene, as has been reported by Vandesompele, *et al*., (2002), does not exist. These findings suggested therefore that some reference genes might only be ideal for a certain set of genes, tissues and conditions. This variation might be ameliorated by use of more than one reference genes' geometric mean (Vandesompele*, et al.*, 2002), to disperse the variation effect of unsuitable reference genes.

A larger number of reference genes would be suitable for dispersing the over/under estimation of expression values as a result of using wrong reference genes. In addition, removing any such unsuitable reference genes might increase the accuracy of the transcript abundance estimations. The approach of using RT-qPCR expression values normalised to the geometric means of *tubulin* and *ß-actin* reference genes was therefore adopted and used for comparison with the RNA-seq estimations of the same genes.

The transcript reconstructing approaches, Ab, Gm and Trinity assembled varied in the counts of the target gene copies, *SOT5, SDH, ADG, ANS, and PL*. Gm had the highest number of total transcripts, while Trinity only assembled 4 fold less than Ab or Gm (Table 3.4). Ab and Gm had no significant difference in the count of the target genes, but were both significantly different from Trinity. As a result of the low count of the target gene transcripts, as well as the low transcript breadth coverage of *M. x domestica* gene models and *M. x domestica* ESTs gene set, Trinity pooled assembled transcripts were dropped from further analysis. To confirm the accuracy and reproducibility of the transcriptome analysis results, five selected native genes, *ANS, PL*, ADG, SOT5 and *SDH*, were selected for RT-qPCR. The correlation between transcriptome analysis and RT-qPCR was evaluated.

Scatter plots were generated comparing the $\log_2$ fold change determined by RNA-Seq and RT-qPCR. A comparison of the RT-qPCR and RNA-Seq (RSEM), calculated transcript expression estimates showed a positive correlation of the two methods with an R^2 value of 0.897 at a critical value of 5%, using Pearson Correlation Coefficient (Figure 3.21), thereby validating the observed RNA-seq expression estimates. Several studies have

alluded to the shortcomings of the RT-qPCR in estimating transcript abundance (Vandesompele*, et al.*, 2002), however in this study the RT-qPCR and RSEM estimations of transcript abundance correlated, and as such the expression levels of RNA-seq gene expression estimates were acceptable.

Table 3. 4 Map first based approaches reconstructed more target gene transcripts

	Total count	SOT5	SDH	ADG	ANS	PL	Ab	Gm
Ab	264	79	52	19	13	22		
Gm	290	81	49	23	20	19	(0.0006) R=0.9879	
Trinity	66	13	15	3	1	8	(0.07218) R=0.84404	(0.140) R=0.7552

The table shows five target genes, *anthocyanidin synthase (ANS), glucose-1-phosphate adenylyltransferase (ADG), NAD-dependent sorbitol dehydrogenase (SDH), pectate lyase A10 (PL),* and *sorbitol transporter 5 (SOT5)* reconstructed using Cufflinks (Ab initio, Ab and gene models, Gm) and Trinity. The counts of the reconstructed transcripts per target gene were measured. The transcript counts are inclusive of isoforms. Total count = total number of the reconstructed transcripts for all the five target genes. Statistical significant differences in the assemblers' abilities to reconstruct the five genes including their isoforms were assessed using Gen Stat 64-bit Release 15.1, 2012. Figure s in brackets are p values at 5% critical value, R= correlation coefficient.
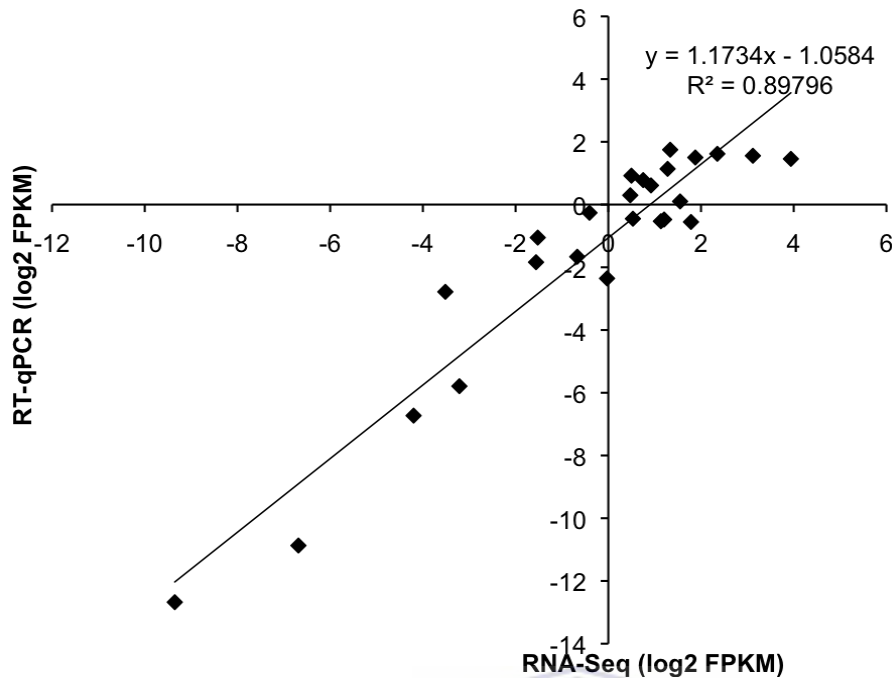
Figure 3. 21 Fold change correlation of RT-qPCR and RSEM transcript expression estimates of five genes in 'Golden Delicious' developing fruit pulp. The geometric mean (calculated as in section 2.2.16.2) of *ß-Actin* and *Tubulin* reference genes was used to normalize the expression values of the target genes *ANS = Anthocyanidin synthase; ADG = glucose-1-phosphate adenylyltransferase; SDH = NAD-dependent sorbitol dehydrogenase; PL = pectate lyase A10; SOT5 = sorbitol transporter 5* using RT-qPCR. RSEM package (li and Dewey, 2011) was used to calculate the FPKM gene expression estimates of the native genes (section 2.2.15). Both gene expression estimates from RT-qPCR and RNA-Seq RSEM were log$_2$ transformed and gene expression at Daa25 was used as the referral point for up or down expression of genes (fold change). The graph shows the correlation of the fold changes in gene expression across time points between RT-qPCR and RSEM estimates.

3.3 Conclusion

The aim of this experiment was to reconstruct full-length transcripts from RNA-seq data pooled from peel and pulp 'Golden Delicious' fruit tissue harvested at 25, 35, 60, 87, 135, and 150 days after anthesis (daa). A comparison of *de novo* (Trinity) and genome-guided (Cufflinks) assemblers revealed that genome-guided reconstructed transcripts had higher count and breadth coverage of *M. x domestica* genome v1.0 gene models and ESTs gene set compared to *de novo* reconstructed transcripts. A combination of genome-guided and de novo assembly approaches 186498 transcripts longer than 300 bases. This

transcript set constitutes a comprehensive transcript data set during apple fruit development and was used in the subsequent experiments. The comprehensive transcript data set extended 25155 *Malig*ned models in both the 3´and 5´ends further improving the annotation of the apple genome. Though 25155 gene models were verified, a substantial gene space still resides in the gaps of the genome, unanchored contigs and possibly as yet to be sequenced gene space.

Annotation of the combined (genome-guided and reference genome-unmapped *de novo* assembled) transcripts revealed that 29% of the combined peel and pulp transcriptome had hits to NCBI and Swiss-prot protein databases, and 21% had unclassified function. Among the transcripts with known function, transcripts encoding proteins with binding function were predominant. "No-hit" transcripts composed mostly of leucine rich repeat; WD domain-G beta repeat and tetratricopeptide repeat domains.

Some of the reconstructed transcripts had homology to sequence regions on the chloroplast DNA (cpDNA) and mitochondria DNA (mtDNA). These transcripts common among the nucDNA, cpDNA and mtDNA could not have been due to contamination of the samples with plastidial DNA, considering a total RNA gel analysis revealed discreet 25s and 18s ribosomal RNA bands in all samples with no indication of smearing. This homology study might therefore suggest the presence of intricate regulatory networks among nuclear DNA (nucDNA), cpDNA and mtDNA for possibly specialised functions such as photosynthesis and respiration or general functions.

Transcripts abundance estimates using RSEM positively correlated to RT-qPCR estimates across the six time points, only when multiple reference gene geometric means was used. Transcript levels of expression were affected by removing rRNA, de-duplicating, as well as sub-setting read count. Read de-duplication and sub-setting resulted in a large number of transcripts without read support, as such *in silico* rRNA read treated transcript abundance estimations were adopted as the preferred approach.

This study provides validation to existing apple genome v1.0 gene models as well as improve and provide new gene models. Transcripts not mapping to the genome pseudo haploid molecules suggests the existence of gene space unaccounted for by the current apple genome v1.0. As such the genome requires further improvements to fill in the gaps.

4   CHAPTER 4

Developmental mRNA profiling in 'Golden Delicious' apple fruit enriched for pulp

Abstract: The mesocarp constitutes the largest edible portion of the apple fruit. With fruit development morpho-physiological changes occur in this parenchymatous tissue. These changes define fruit quality and are governed by differential gene expression at particular time points as regulated by environmental and genetic factors. An RNA-seq analysis of the pulp tissue collected from six harvesting time points was performed, to gain molecular understanding of gene expression changes in developing pulp tissue. Over the course of fruit development sugar content positively correlated with increase in fruit weight, attaining a maximum of 14 °Brix score at 150 days after anthesis (daa). Out of 186498 reconstructed transcripts, 27255 transcripts were significantly expressed in the pulp tissue and, were clustered into 11 main well-correlated groups and annotated using 22 main functional categories from the Munich Information Centre for Protein Sequences (MIPS). In each of the 22 main MIPS functional categories transcript count increased with harvesting time. The top four functional categories were, proteins with binding function; metabolism; interaction with environment and cell death in that decreasing order. A significant number of transcripts remained unclassified. Calvin Cycle gene encoding transcripts showed peak expression between daa35 and daa87. High levels of sorbitol dehydrogenase (SDH) encoding transcripts were expressed from daa25 to daa87, suggesting early fruit development mostly relies on leaf-manufactured sorbitol. Gene transcripts encoding sucrose biosynthesis enzymes, sucrose phosphate phosphatase (SPP) and sucrose synthase (SUS) had peak expression at daa60. The sucrose degrading enzyme, invertase (INV), was encoded for by gene transcripts, which increased sharply from daa60 to daa150, possibly in response to the energy demanding climacteric process as the fruit ripens. Gene transcripts encoding starch biosynthesis enzymes had peak expression at daa60, while those encoding starch-degradingenzymes increased with fruit development. These finding suggests peak starch accumulation to be about daa60, after which starch degradation takes place towards physiological maturity. The increase in the

expression of genes encoding ABA biosynthesis with fruit development suggests an increase in fruit stress with development. The levels of genes encoding 1 aminocyclopropane-1-carboxylic acid (ACC) oxidase (*ACO)*, respiratory burst oxidase homolog F (*RBOHF)* and ethylene responsive factors RAP2-3 (*ERF-RAP2-3)* ethylene biosynthesis related proteins increased, while those encoding ACC synthase (*ACS)* decreased with fruit development, suggesting *ACO* is the chief driver of ethylene burst during climacteric phase. Expression patterns of genes that play key functional roles in barsinosteroid and auxin metabolism and perception were also elucidated. Differential expression of structural, transcriptional and receptor genes across gene families at particular time points are reported. The existence of multiple isoforms at any given time requires further study to elucidate their involvement in particular pathways and tissues.

Keywords: transcriptional profiling; regulatory networks; fruit development; high throughput sequencing; fruit quality; carbohydrate accumulation; hormonal metabolism and perception
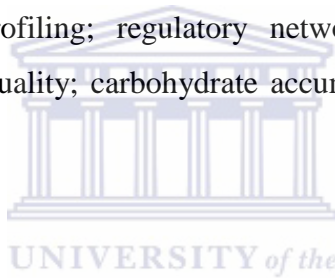
4.1 Introduction

Apple (*M. x domestica* Borkh.) mesocarp consists of fairly uniform highly vacuolated parenchymatous cells permeated by vascular bundles. The mesocarp, loosely referred to as the 'pulp', consists of the biggest portion of the edible pome. Several morpho-physiological changes take place in the pulp during development. The accumulation rates and levels of carbohydrates, organic acids and secondary compounds in the pulp under the control of environmental factors determine fruit crispiness, texture, taste and flavour. The ultimate pulp composition at physiological maturity define the pulp's nutritional and health attributes and is affected by variety, maturity, location, agronomic practices and environmental factors (Muchuweti and Chikwambi, 2008; Schaffer*, et al.*, 2007; Veberic*, et al.*, 2005).

The pulp is thus a unique aspect of pome fruit and a part of human diet, necessitating a better molecular understanding of the mechanisms underlying pulp morpho-physiological changes during fruit development and ripening determining fruit quality. Morphological, proteomic, metabolomics, genomic and studies on the complex of all transcripts in a given cell (transcriptome) have been done to elucidate gene regulation and expression during fruit maturation and ripening (Ruan, *et al.*, 2004). However, a number of previous studies have partially elucidated fruit cortex, whole fruit, leaves, and peel plus cortex transcriptomes using cDNA micro arrays and other PCR based technologies (Han, Bendik, *et al.*, 2007; Hattasch, *et al.*, 2008; He, *et al.*, 2008; Janssen, *et al.*, 2008; Mintz-Oron, *et al.*, 2008; Soglio, *et al.*, 2009), but remain limited in revealing transcripts of low abundance. A high throughput technique is thus needed to elucidating the full pulp mRNA transcriptome during fruit development.
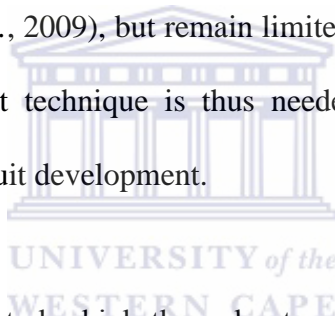
In the work described in this study, high throughput sequencing of 'Golden Delicious' apple fruit pulp mRNA transcriptome was done to elucidate carbohydrate and hormonal developmental changes during fruit development. This study reports expression patterns of the sugar accumulation, starch accumulation, hormonal metabolism and perception and CO_2 fixation encoding genes during fruit development from daa25 to daa150.

4.2 Results and Discussion

4.2.1 *M. x domestica* cv. 'Golden Delicious' pome development

Five fruits of each time point were measured for mass (g) and total soluble solids as ºBrix score. Fruit weight increased slowly from daa0 to daa35, followed by an exponential increase peaking up at daa87 with a subsequent levelling off at daa135 to daa150 (Figure

4.1B). Apple fruits develop initially by cell division and then by an increase in the volume of the cells till full cell size is reached (Fleancu, 2007; Zheng, *et al.*, 2012). The corresponding increase in fruit fresh weight with fruit development is not only due to an increase in cell volume, but also accumulation of organic compounds in the cells as well. The study has shown that maximum 'Golden Delicious' fruit weight was attained after daa87, when carbohydrate accumulation reached its maximum (Figure 4.1B). Fleancu, (2007) and Janseen *et al*, (2008) also observed sigmoid responses in fruit weight to fruit development, validating the observations. The ºBrix measurements taken from daa25 to daa150 showed a steady increase from 6 to a peak value of 14 ºBrix, respectively (Figure 4.1A). The observed ºBrix score was close to the expected ºBrix range of 12 for 'Golden Delicious' at physiological maturity (Rutkowski, *et al.*, 2008). The progressive increase in sugar formation is a consequence of a direct sugar translocation from leaves in the form of sorbitol and sucrose, carbon fixation by the green tissue, and degradation of macro-carbohydrate compounds as the fruit develops (Berüter, 1985) Though ºBrix score gives an indication of the sweetness of the fruit, other factors, such as acidity, contribute to the richness in flavour and are affected by the physiological stage, season, location and cultivar (Rutkowski, *et al.*, 2008). An array of other important physiological processes do take place to argument sugar accumulation and to sustain other plant processes.
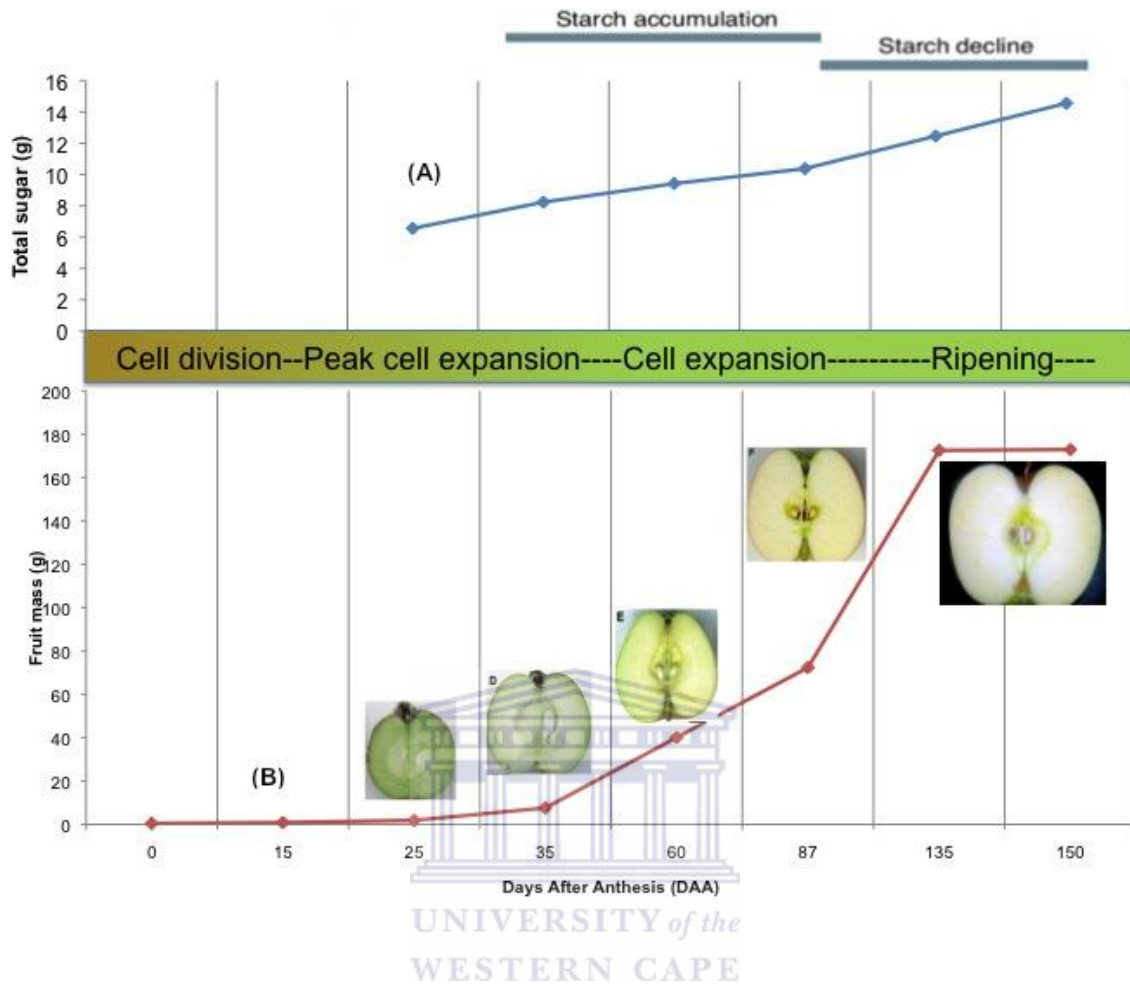
Figure 4. 1 Relation between 'Golden Delicious' fruit development and physiological changes measured during eight time points of pome fruit development. A, degree Brix (ºBrix, n=5); B, pome mass (g, n=5), periods of starch changes were obtained from (Janssen, *et al.*, 2008).

These processes are regulated by spatial-temporal gene expressions (Janssen, *et al.*, 2008). The actual structural gene expression patterns and possible mechanisms of regulations have remained elusive. Analysis of the expression patterns of both structural and regulatory genes is a starting point for inferring possible regulatory mechanisms.

4.2.2 Apple pulp significantly expressed transcripts cluster into different expression patterns

To obtain an overall view of gene spatial-temporal regulation during pome fruit development, RNA-seq analysis was performed and 186498 transcripts were reconstructed using a combination of genome-guided and de nova assembly approaches (section 2.2.91 and 2.2.9.2). The reconstructed transcripts were screened to retain only transcripts with more than one fragment per kilo base per million reads (FPKM) expression value. A total of 27255 transcripts with significant ($p<0.05$) expression in the pulp tissue were obtained by performing a two-way ANOVA analysis, with two tissues (peel and pulp) and six time points (25, 35, 60, 87, 135, and 150 daa). The 27255 pulp significantly expressed transcripts were then profiled and clustered according to their expression values using MeV 2.0 software package (Saeed*, et al.*, 2003) as described in section 2.2.17.
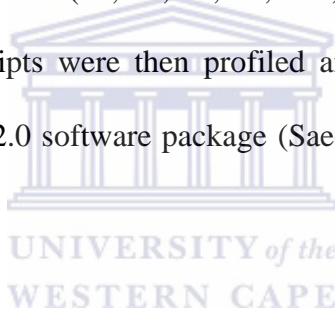
The *k-means* method within the MeV package was used for clustering (Ben-Dor*, et al.*, 1999; Saeed*, et al.*, 2003) of 27255 pulp significantly ($p<0.05$) expressed transcripts into 20 closely correlated clusters of transcripts to reflect major fruit pulp developmental patterns of transcriptional expression, from daa25 to daa150 (Figure 4.2). The choice of the number of clusters was optimized to 20 using Figure of Merit (FOM) (Yeung*, et al.*, 2001) section 2.2.17 (Figure 4.2). Within the 20 closely correlated clusters (Figure 4.2), eleven general expression trends were noted and clusters with similar expression trends were regrouped into eleven groups of distinct general expression trends (Figure 4.2). Groups one and two constituted 42% of the total transcripts (Figure 4.2).

Each cluster group was analysed for pathway enrichment using KOBAS (Xie, *et al*., 2011) to reveal the trend of physiological processes affected by enriched pathway enzymes encoded by the reconstructed transcripts (section 2.2.17). Transcripts encoding enzymes in the plant hormone signal transduction pathway were the mostly enriched in group one (45) and group two (45) (Table 4.1).

Genes encoding photosynthesis pathway enzymes were highly enriched in group-one, and decreased in expression with harvesting time (Table 4.1). Enzymes in the metabolism of porphyrin and chlorophyll pathway were also enriched in group one, suggesting increased chlorophyll degradation with fruit development. These findings agree with a decrease in fruit photosynthetic capacity and fruit chlorophyll a/b content as the fruit develops observed by Fleancu (2007). Enrichment of photosynthetic pathway enzymes in group-one suggests a significant contribution of the fruit's photosynthetic surface area to the carbon pool. It is unexpected however, that deep lying cells contribute significantly towards photosynthesis. In spite of the deep cell position, pulp mesocarp cells have chloroplasts even up to physiological maturity. However, the chlorophyll content decreases and is only confined to the vascular bundle system at physiological maturity (Figure 4.1).

The enrichment of ribosome and RNA-transport pathways early in fruit development is in agreement with the intense cell division and cell enlargement processes characteristic of early fruit development, which requires a significant amount of protein and nucleic acid for replication.
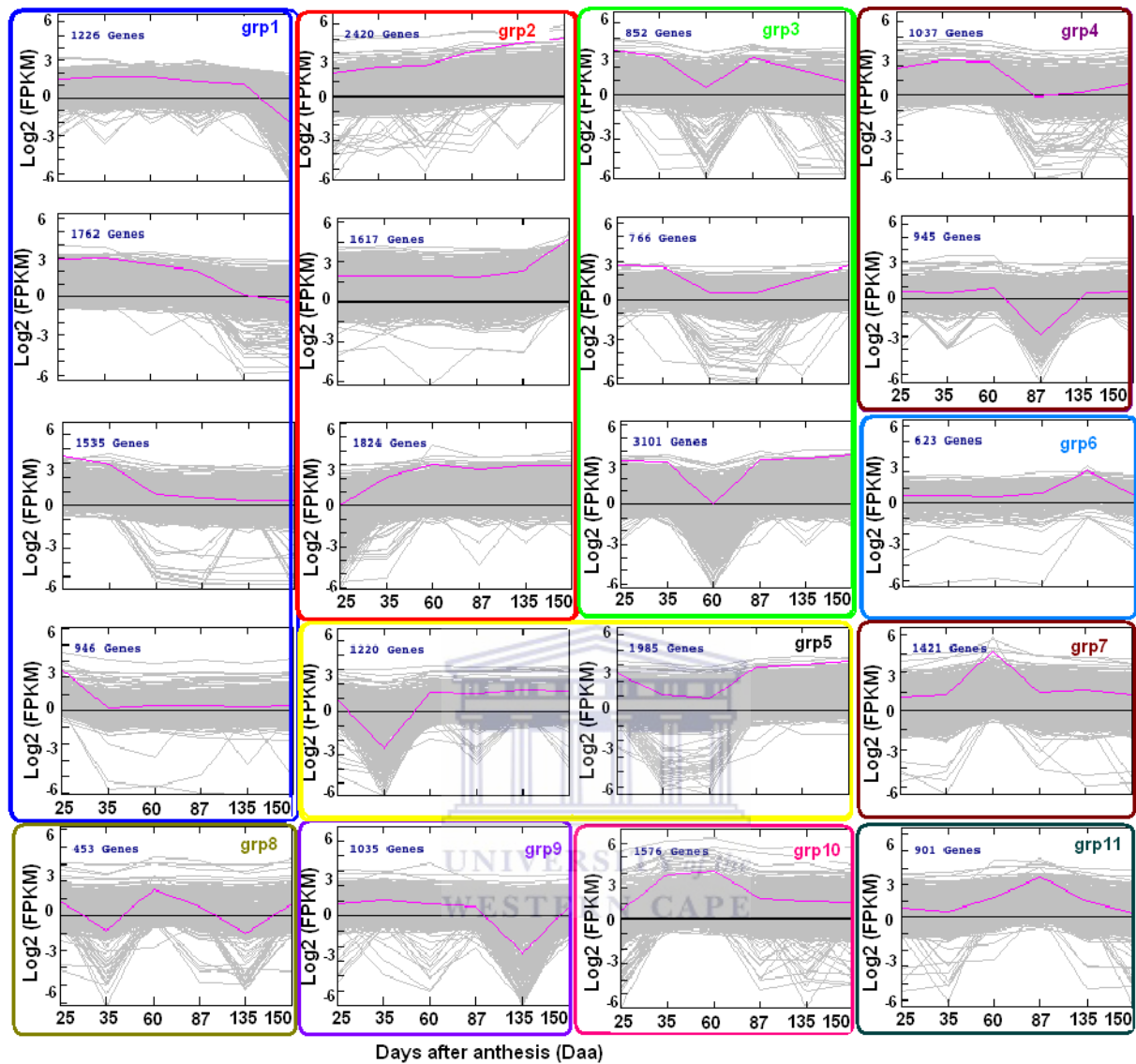
Figure 4. 2 Cluster analysis of statistically significantly pulp tissue expressed genes using *k-means*. Clustering of transcripts with respect to expression levels was done using *k-means* method and Pearson Correlation at 5% critical value available from MeV 2.0.2 (Saeed, *et al.*, 2003). Differentially expressed transcripts obtained using RSEM were subjected to a two-way ANOVA to identify peel tissue significantly expressed transcripts, which were then separated into closely correlated clusters of expression trends. The Figure of Merit (FOM) was used to optimise cluster number to twenty. Within the twenty clusters, eleven general expression trends were observed and the twenty clusters were re-grouped into the eleven general trend groups. Grp=group. The general trend groups depicted: group 1, a decrease with fruit development; group 2, an increase with fruit development; group 3, a decrease at daa60; group 4, decrease at daa87; group 5, decrease at daa35; group 6, transient peak at daa60; group 7, an increase at daa60; group 8, an increase at daa135; group 9, an increase at daa87; group 10, plateau at daa35 to daa60 and group 11, a decrease at daa135. The number of genes per cluster is shown in the left upper corner. Group number is written in the left upper corner of each colour-coded group. The horizontal axis is the harvesting time points (daa) and the y-axis is the transcript expression values. Expression values (FPKM) were $\log_2$ transformed to normalize the data (section 2.2.17). The purple line depicts the median expression level of the transcripts in a cluster.

Genes encoding 'carbon fixation in photosynthetic organisms' pathway enzymes were enriched in group two, and increased in expression with fruit development. Fixed carbon in the fruit is not only from the fruit photosynthetic machinery, but also significantly from the leaf-translocated photo-assimilates, more so with fruit development.

The oxidative phosphorylation reactions are responsible for most of the ATP, energy compounds, as well as generation of proton motive gradients in plant cells. The produced energy is invaluable to fruit cellular physiological processes, while the proton motive gradients facilitate minerals, water, proteins and photo assimilates transportation across cell membranes. Transcripts encoding enzymes in the oxidative phosphorylation pathway were enriched in groups three, two, five, and one (Table 4.1). A decrease in expression of some of the oxidative phosphorylation pathway encoding transcripts at daa35 and daa60 correlates with a decrease in fruit respiration rate as the fruit enters the pre-climacteric phase (Soglio, *et al.*, 2009; Soglio, *et al.*, 2007). Though energy is required for cell division and enlargement, an even higher amount is required for the climacteric process towards fruit maturity when a plethora of physiological processes are initiated to define fruit quality.

Terpenoids are multifamily and multifunctional molecules, which act in defence and as signal molecules (Ashour, *et al.*, 2011). For humans, the flavour of the terpenoids increases the aesthetic value of and apple fruit. The terpenoid backbone biosynthesis pathway encoding genes showed that though some of the genes may be down regulated or show transient peaks with fruit development, the majority increased in expression with

137

fruit development (group one, Table 4.1). An increase therefore of these volatile compounds with fruit development increases the attractiveness of the fruit at physiological maturity, and may be used as a bio-indicator of fruit maturity.

Peroxisomes are eukaryotic multifunctional organelles involved in numerous processes including primary and secondary, development and response to abiotic and biotic stress (Hu, *et al.*, 2012). Peroxisome related genes increased in expression with fruit development (Table 4.1), suggesting an increase in the formation of the peroxisomes as well as the metabolic processes in the peroxisomes. Common processes exist between peroxisomes and mitochondria, such as ß-oxidation of fatty acids and scavenging of peroxides (Neuspiel, *et al.*, 2008). An increase in peroxisome pathway associated transcripts would therefore suggest a possible increase in the catabolism of stored oil to provide energy and protection of fruit cell membranes from oxidizing peroxides released during the climacteric phase. Though the *k-means* clustering method minimizes intra-cluster variability, quantitative and qualitative expression differences among members of the set are likely to occur. This clustering as well as the pathway enrichment analysis showed that pome development is a progressive process as well as a dynamic process, wherein genes showed a steady increase, decrease and transient changes in expression across all stages of development.

Table 4. 1 Pathway enrichment analysis of pulp expressed reconstructed transcripts

Term	Pathway	Cluster group										
		1	2	3	4	5	6	7	8	9	10	11
Amino sugar and nucleotide sugar metabolism	Ko00520	33	44	32	4	17	2	7	2	8	9	6
Aminoacyl tRNA biosynthesis	Ko00970	12	17	14	9	12	0	12	0	2	4	2
Carbon fixation in photosynthetic organisms	Ko00710	19	30	29	8	16	2	9	2	4	6	3
Endocytosis	Ko04144	12	32	25	7	17	0	4	3	2	11	1
Oxidative phosphorylation	Ko00190	10	31	49	10	24	2	4	0	3	13	5
Peroxisome	Ko04146	7	40	31	8	13	3	4	2	3	6	4
Photosynthesis	Ko00195	31	8	6	3	3	1	0	0	4	1	4
Plant hormone signal transduction	Ko04075	45	45	36	10	14	4	6	4	8	14	8
Plant pathogen interaction	Ko04626	37	41	34	14	24	4	10	2	5	10	11
Porphyrin and chlorophyll metabolism	Ko00860	10	16	5	0	14	1	3	0	1	4	5
Proteasome	Ko03050	10	7	36	8	6	1	3	0	1	5	1
Protein processing in endoplasmic reticulum	Ko04141	10	27	52	6	31	1	12	3	8	11	2
Ribosome	Ko03010	20	4	21	5	3	4	2	2	1	0	2
RNA degradation	Ko03018	5	24	20	11	12	2	7	1	3	2	5
RNA transport	Ko03013	43	33	27	12	32	1	14	6	4	17	4
Spliceosome	Ko03040	15	17	22	15	28	1	15	5	4	13	14
Terpenoid backbone biosynthesis	Ko00900	7	13	6	3	4	3	0	2	3	1	0
Ubiquitin mediated proteolysis	Ko04120	12	15	17	1	11	2	0	3	2	3	3

The most enriched pathways per cluster pooled into groups with similar expression pattern and changes thereof across the eleven expression patterns. Cluster group = a collection of clusters of similar expression trend, Pathway enrichment was calculated as the total gene count per cluster group. Pathway = KEGG accession of each pathway, Term = KEGG pathway description of the pathway accession.

4.2.3 Functional annotation of reconstructed transcripts across different stages of development

Annotation of transcripts for function allows for insights into the particular molecular functions, cellular and biological processes in which putative proteins are involved. In this study functional categories were assigned to differentially expressed transcripts with

pulp tissue significant change in transcript abundance over the course of six developmental time points using the Munich Information Centre for Protein Sequences (MIPS) (Mewes, *et al.*, 2002)catalogue with annotations of the top *Arabidopsis* BLAST hits (80% identity and e-value 0.05).
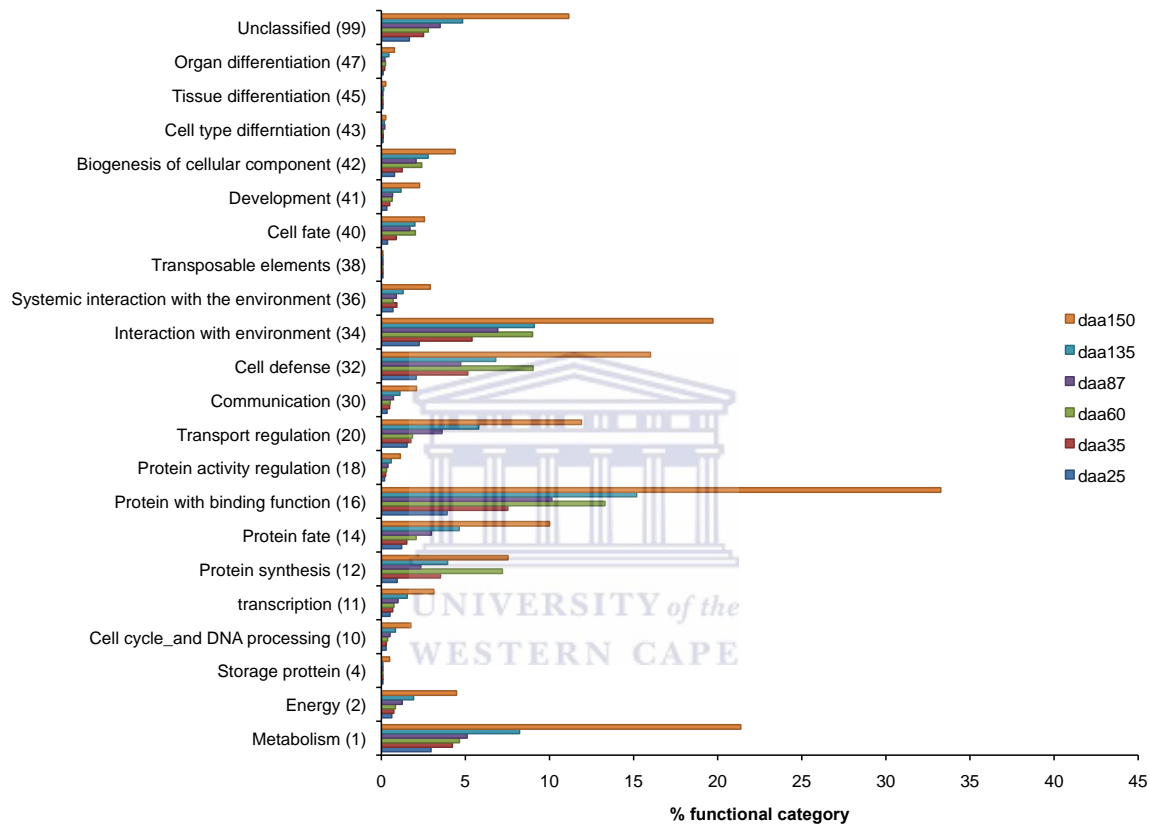
Figure 4. 3 Functional analyses of transcripts with at least two-fold change in abundance throughout pome development. Distribution of Unigenes according to their MIPS functional categories (MIPS v2.0) (Mewes, *et al.*, 2002) within the stages of pome development. Percentages are based on the total number of transcripts (27255 transcripts). Number in parenthesis following category names indicates the MIPS number for each category.

The distribution of transcripts within each of the twenty functional categories was determined and profiled for each time point throughout fruit development (Figure 4.3). Statistically significant differences in the qualitative distribution of genes between functional categories amongst these developmental stages were observed (Figure 4.3, p-value > 0.05). Distinctively, the housekeeping processes, such as cell fate, tissue

140

differentiation, cell type differentiation, organ differentiation, development, communication, transcription, cell-cycle and DNA processing, storage protein, protein synthesis and energy were under-represented across all time points (Figure 4.3). In contrast, functional categories of proteins with binding function, metabolism, interaction with the environment, cell defence, transport regulation and protein fate were over-represented across all time points (Figure 4.3).

Among the over-represented categories, proteins with binding function had the highest percentage of transcripts across all time points. The onset of the climacteric phase is accompanied by increases in respiration, volatile compounds, reduced xylem conductivity, and cell-wall degradation. These changes result in a plethora of metabolic reactions, which define fruit quality at harvest. Proteins are synthesized as per requirement but are also degraded as a way of regulating their activity throughout fruit development. Metabolic products and minerals are uploaded into the fruit from the xylem and transported across cells to sustain fruit growth throughout fruit development hence the constant high enrichment. The developing fruit is under constant challenges from biotic and abiotic factors necessitating constant high enrichment defence related proteins to overcome the adverse effects of biotic and abiotic factors (Janssen*, et al.*, 2008).

A global view of the pulp transcript set reveals that daa150 expression profile presented a distinct fingerprint, wherein daa150 had the highest enrichment across all 22 categories. In the metabolism, energy, transcription, cell cycle and DNA processing, protein fate, transport regulation, protein activity regulation, communication and development

141

functional categories, a proportional increase in transcript abundance was observed with increase in fruit maturity. However, protein with binding function, protein synthesis, cell defence, interaction with environment, systemic interaction with the environment, cell fate, and biogenesis of cellular components functional categories show a higher enrichment during daa60 than daa87. The cell defence category was more enriched at daa35 compared to daa87, and daa60 more than both daa87 and daa135. The decrease in the count of transcripts in the defence category with fruit development was expected. Developing apple fruits have been reported to resist infection better during early fruit development, however this potential decreases with fruit ripening making fruits more susceptible late during fruit development. Oxygen-scavenging enzymes, superoxide dismutase, catalase and peroxidase and polygalacturonase-inhibiting proteins (PGIPs) were shown to be more active early during fruit development (Abassi, *et al.*, 1998). As such, apple fruits are more venerable after the climacteric phase (De Lorenzo, *et al.*, 2001; Janisiewicz and Korsten, 2002).

4.2.3.1 Carbohydrate accumulation during apple fruit development

M x domestica fruit is one of the economically important sink organs in which sugar accumulation is a major determinant of yield and quality. The partitioning of the sugars is governed by several physiological processes, including photosynthetic rate, phloem loading in the source leaf, long-distance translocation in the phloem, phloem unloading in sink organs, post-phloem transport, and metabolism of imported sugars in sink cells (Gao, *et al.*, 2003; Gucci, *et al.*, 1995; Lee, *et al.*, 2007; Zeng, *et al.*, 2010; Zhang, *et al.*, 2004). Sugars (fructose, glucose, sucrose) and polyols (sorbitol, mannitol) are derived from

triose phosphate, and the absolute amounts and relative proportions of the sugars in the fruit determine the sweetness of the fruit (Gucci, *et al.*, 1995; Lee, *et al.*, 2007). In Rosaceae, photosynthesis-derived carbohydrates are transported mainly as sorbitol, constituting more than 50% of the total soluble solids (Zeng, *et al.*, 2010). The biosynthesis of sugars, starch and polyols (sorbitol and mannitol) are closely related and depend on some important enzymes (Teo, *et al.*, 2006).

Much work has been done to elucidate the individual pathways and their interactions using computational and experimental analysis (Zeng, *et al.*, 2010). To date most enzymes in these pathways have been assigned putative function or have not been identified in apple.
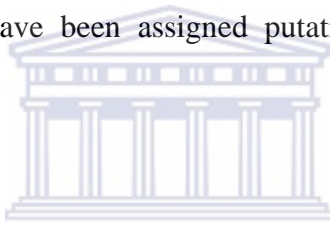
4.2.3.2 Carbon assimilation in the apple fruit pulp tissue

Plants increase in biomass by assimilating carbon from the atmosphere into carbon products. Apple trees are C_3 plants, producing glyceraladehyde-3-phosphate through the Calvin cycle as their first fixed carbon product. Genes encoding Calvin cycle enzymes are highlighted in Figure 4.4 and showed peak expression between daa60 and daa87 (Figure 4.5), suggesting daa60 to be the point of maximum carbon assimilation.

Genes encoding phosphoglycerate kinase 1 (PGK1, EC 2.7.1.4) had transient peak expression at daa60, and increased from daa87 to daa150. The continued increase in PGK1 abundances, despite other genes encoding carbon-fixing enzymes decreasing, may not be associated with carbon fixation in the pulp. Rather, the increase in genes encoding

PGK1 from daa87 could be related to glycolysis as the climacteric process reaches its peak during which sugar accumulation in the fruit pulp increases (Figure 4.1). Palma *et al.,* (2011) suggested high expression levels of genes encoding Rubisco-small subunit (rbcS, EC 4.1.1.39) could be an attempt by the fruit to increase the expression levels of this relatively inefficient carbon-fixing enzyme (Palma*, et al.*, 2011). The expression profiles of the genes encoding Calvin cycle enzymes suggested the fruit's photosynthetic capacity might be highest at daa60 (Figure 4.5), after which fruit no longer has net carbohydrate accumulation and is preparing for or has entered the pre-climacteric phase. Photosynthesis has been reported to decrease with fruit growth and maturation, owing to the degradation of chlorophyll from the onset of pre-climacteric ripening (Fleancu, 2007).

As the fruit develops, chlorophyll is gradually degraded resulting in a diminishing supply of reducing equivalents for the dark reaction. Fruit photosynthesis has also been reported to contribute between 5 and 15% of the total fruit carbon requirements (Fleancu, 2007). The existence of photosynthesis related genes in the pulp are not unreasonable considering the pulp colour is initially green turning to off-white with maturity (Figure 4.1). Though a significant proportion of the pulp turns off-white at physiological maturity, the vascular bundles retain a green colouration, suggesting the presence of chlorophyll. This presence of chlorophyll as well as the expression of photosynthesis related transcripts in the pulp suggests that the pulp is an active contributor to carbon fixation.
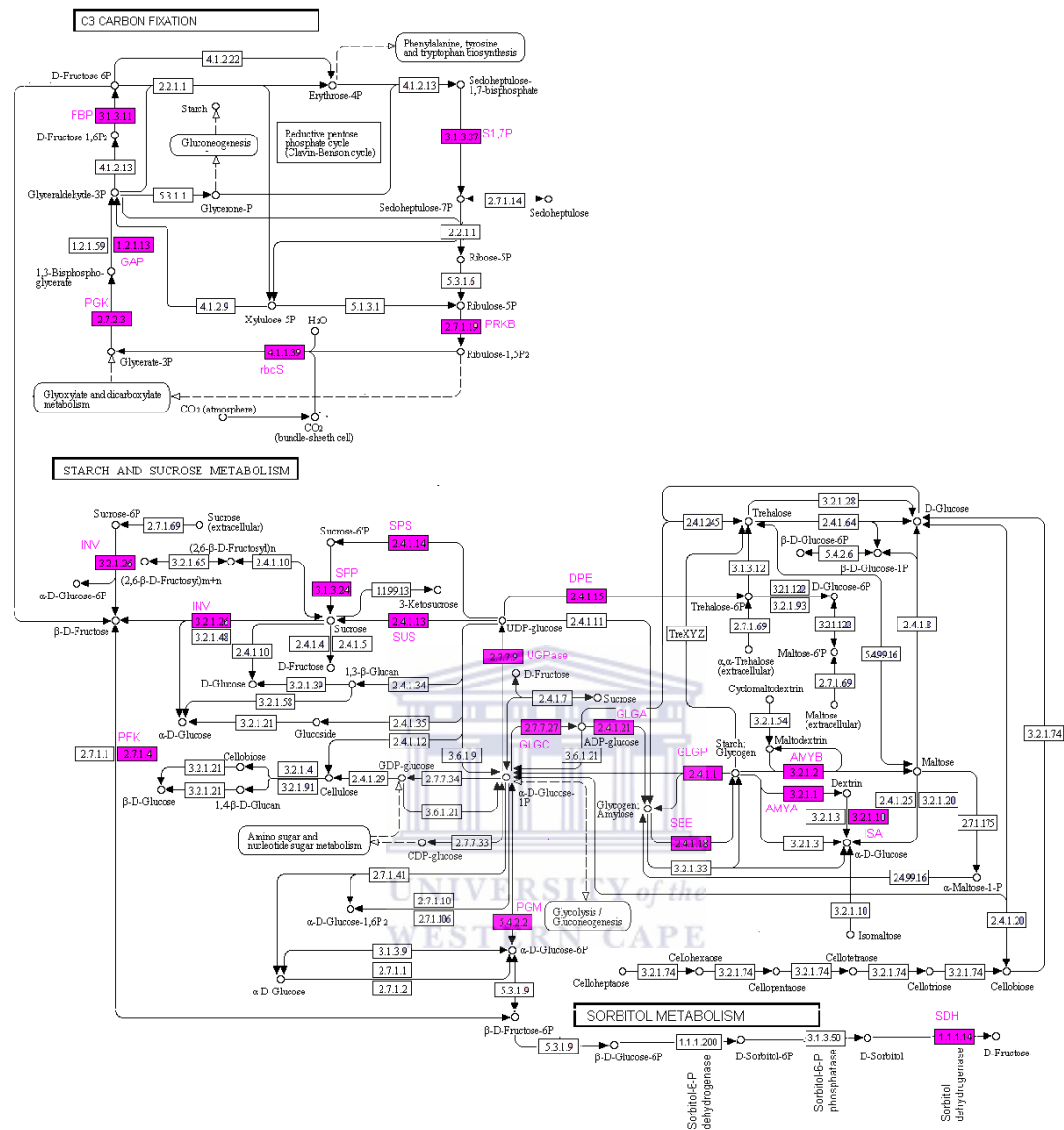
Figure 4. 4 Localisation of carbon assimilation pathways enzymes enriched in apple fruit pulp tissue.
The Figure shows a modified combined carbon fixation and starch and sucrose metabolism KEGG pathway reference map. The pulp tissue shows enrichment of sucrose, starch and sorbitol metabolism pathway enzymes. Boxes coloured pink represent sub pathway genes enriched in 'Golden Delicious' fruit pulp tissue. Names of tissue-enriched enzymes are written close to the EC number of the enzyme in pink.
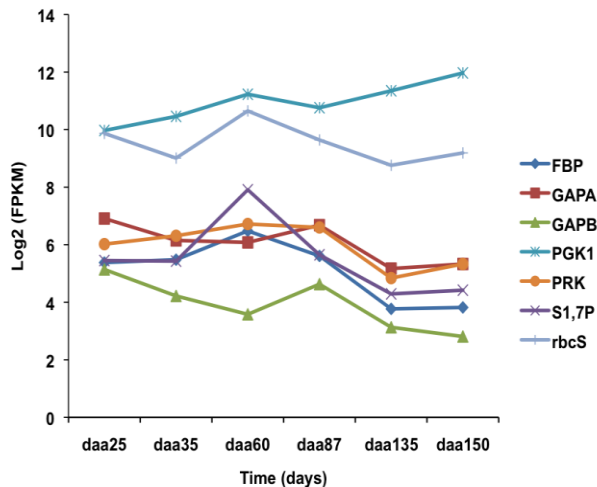
Figure 4. 5 Expression profiling of Calvin cycle encoding transcripts expressed in the pulp during pome fruit development. The Figure shows expression trends of more than two-fold up regulated genes expressed in the apple pulp tissue encoding enzymes in the Calvin cycle pathway. Expression levels are totals of transcript expression (FPKM) per locus, at each time point, from daa25 to daa150. Expression values (FPKM) were $\log_2$ transformed to normalize the data (section 2.2.17).

A proteomics study of ripe apple pulp by Marondedze and Thomas (2011, 2012) identified several photosynthesis related proteins, suggesting the transcripts identified in this study are translated into active protein. The observations in this study, unlike those of Marondedze and Thomas (2011) provides and further refines the profiles of the genes encoding photosynthesis pathway proteins, highlighting daa60 to daa87 as the region of possible high pulp photosynthetic activity (Marondedze and Thomas, 2011, 2012). The fruit total-carbon contribution, like that of the leaf is affected by the biological clock, which slows it down towards the climacteric. The levels of rbcS in the fruit at any one phase may therefore indicate the levels of carbon fixation at that time point.

4.2.3.2.1   Sorbitol metabolism and transport

Sorbitol is one of the several sugar alcohols found in higher plants and is the main photoassimilate produced in apple leaves from which it is translocated to sinks (Gao, *et*

146

al., 2001; Kanayama*, et al.*, 1992; Zeng*, et al.*, 2010; Zhou*, et al.*, 2008). NADP-dependent sorbitol 6-phosphate dehydrogenase (S6PDH, EC 1.1.1.200)is rate limiting for sorbitol biosynthesis (Nookaraju*, et al.*, 2010). The activities of S6PDH produces precursors for sorbitol synthesis from glucose-6-phosphate resulting in the accumulation of sorbitol in the plant (Nookaraju*, et al.*, 2010). In this study no genes encoding S6PDHwere significantly expressed in the pulp tissue. This might suggests that the fruit-produced sorbitol makes no significant contribution to total fruit sorbitol. As the fruit develops it may mostly rely on translocated leaf produced sorbitol with increase in fruit sink strength. The observations in this study are in agreement with, Zing*et al*., (2010)'s findings using a proteomic approach from which they noted sole expression of S6PDH in the adult leaf during vegetative phase (Zeng*, et al.*, 2010). During fruit development, as in developing leaf, the fruit is a net importer of photosynthates to meet its metabolic demands, for energy and structure, as such, there would be very little or no S6PDH expression in the fruit. The fruit manufactured photosynthates also do not need to be translocated far, hence there was no need for conversion of photosynthates to sorbitol for transport.

NAD^+-dependent sorbitol dehydrogenase (SDH, EC 1.1.1.14) converts sorbitol to fructose, and is responsible for all the unloading of sorbitol in fruits (Li*, et al.*, 2012). High levels of genes encoding SDH were expressed from daa25 to daa87 followed by a decrease to daa135 (Figure 4.4, 4.6). The initial high levels of SDH genes suggest the presence of high levels of sorbitol in the pulp, possibly from leaf-translocated sorbitol early in fruit development. Thus the fruit, early in development, mostly relies on leaf-

manufactured sorbitol. Li, *et al.* (2012) observed high levels of *MdSDH2-9* and *MdSDH1* isoforms in immature apple fruit compared to mature fruits, emphasizing the importance of SDH in providing simple sugars for metabolism and structural building blocks as the fruit develops. It is possible that after attaining maximum cell number and size as well as carbohydrate accumulation, the fruit loses its sink strength reducing photo-assimilate translocation from the leaf to the fruit.
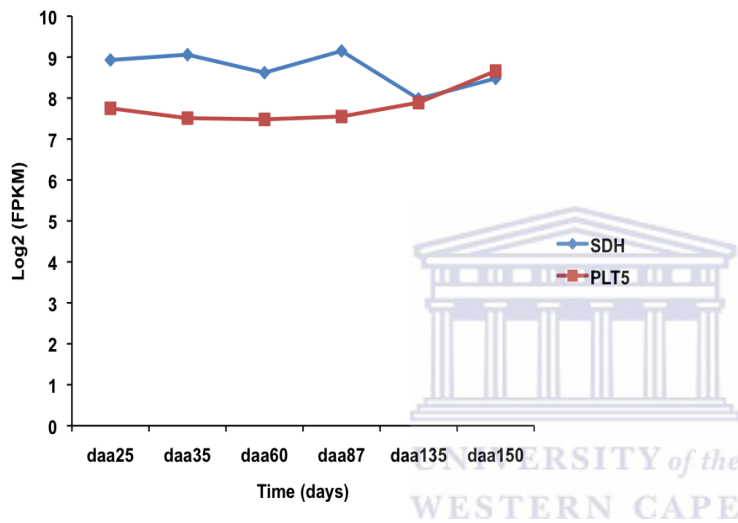
Figure 4. 6 Expression trends of pulp expressed genes encoding sorbitol metabolism and transportation proteins. The Figure shows expression profiles of more than two-fold pulp up regulated transcripts encoding sorbitol degradation and transportation in apple fruit pulp tissue. Polyol symopter 5 (PLT5) pumps out sorbitol from the sieve-element-companion cell complex as it is being translocated into the pulp tissue cells, while NAD^+-dependent sorbitol dehydrogenase (SDH, EC 1.1.1.14)converts sorbitol to fructose, and is responsible for all the unloading of sorbitol in fruits. Expression levels are totals of transcript expression per locus, at each time point, from daa25 to daa150. Expression values (FPKM) were log_2 transformed to normalize the data (section 2.2.17).

The transportation of sorbitol across membranes is facilitated by a polyol symporter (PLT5), which pumps out sorbitol from the sieve-element-companion cell complex as it is being translocated. Genes encoding PLT5 did not change in expression till daa135 (Figure 4.6). From daa135 to daa150 PLT5 encoding genes increased sharply in a similar pattern to SDH. This delayed positive relationship could reflect a close relationship

148

between the level of sorbitol degradation and its 'unloading' in apple fruit (Figure 4.6), and that in the final stages of ripening energy demand is augmented by leaf-translocated sorbitol.

4.2.3.2.2   Sucrose metabolism

The fate of sucrose is mostly dependant on the activities of sucrose phosphate synthase (SPS, EC 2.4.1.14), sucrose phosphate phosphatase (SPP, EC 3.1.3.24), and sucrose synthase (SuS, EC 2.4.1.13) for its formation and invertase (INV, EC 3.2.1.26) for its degradation (Figure 4.4, 4.7). The increase in the expression of genes encoding sucrose biosynthesis may result in sucrose accumulation, which may trigger increased transport and degradation of sucrose, as indicated by increases in genes encoding INV and SUT (Figure 4.7).

Figure 4. 7 Expression profiles of pulp tissue expressed genes encoding sucrose metabolism enzymes. The Figure shows expression trends of more than two-fold up regulated pulp expressed genes encoding sucrose synthesis, degradation and transportation pathway proteins. Expression levels are totals of transcript expression per locus, at each time point, from daa25 to daa150. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

SUS has a dual role in producing both UDP-glucose (for cell wall and gluco-protein biosynthesis) and ADP-glucose (for starch biosynthesis) (Figure 4.4), as such it may

determine the rate of starch and sucrose synthesis (Wang*, et al.*, 1993; Zheng*, et al.*, 2011), regulating the import and compartmentalization of sucrose in the early stage of fruit development (Demnitz-King*, et al.*, 1997). Genes encoding SPP and SUS had peak expression at daa60 followed by gradual decrease to daa150, giving a possible indication of an increase in sucrose formation from daa25 to daa60 (Figure 4.7). The pattern of expression of genes encoding SPP and SUS was similar to the expression profiles of genes encoding starch biosynthesis (Figure 4.8). The increase in genes encoding SUS, similar to starch biosynthesis genes, may also be in response to the increasing demand for ADP-glucose residues for starch biosynthesis. The accumulation of sucrose is however affected by the activities of sucrose degrading INV. According to Nookaraju*, et al,* (2010) the activity of SUS affect sucrose accumulation more than SPS activity in Japanese pear (Nookaraju*, et al.*, 2010). The acid INV and SUS catalyse the cleavage of sucrose in growing fruit. Genes encoding INV decreased from daa25 to daa60 then sharply increased to daa150 (Figure 4.7). The rising levels of INV are in response to the energy demanding climacteric process, hence the need to rapidly degrade sucrose. With fruit maturity, less sucrose is translocated to the fruit, forcing the fruit to rely on degrading in-fruit stored starch for energy. The increase in INV may suggest the exclusive degradation of the already accumulated sucrose to provide energy to the highly respiring fruit tissue resulting in further accumulation of simple sugars (Figure 4.1). Studies on tomato, cucumber and grape fruit development have indicated that INVs are most probably the primary determinants of sucrose levels in tomato, cucumber and hexose levels in grapes (Palma*, et al.*, 2011)

150

The enzymes catalysing the reactions upstream of SPS, SUS, and SPP help in sucrose accumulation, but are not the major determinants of the rate of sucrose formation. In this study, transcripts encoding ß-phosphoglucomutase (ß-PGM, EC 5.4.2.2), phosphoglucomutase (PGM) and UGP-glucose pyrophosphorylase (UGPase, UGP2 EC 2.7.7.9) enzymes showed similar expression profiles, which were also similar to INV (Figure 4.4, 4.7). The two forms of PGM act on distinct substrates, ß-PGM catalyses the inter-conversion of only ß-D-glucose-1-phosphate to ß-D-glucose-6-phosphate, while PGM acts on α-D-glucose-1-phosphate (Uematsu, *et al.*, 2012). There were higher levels of *PGM* encoding transcripts than *ß-PGM* (Figure 4.), suggesting that D-glucose-1-phosphate exists predominantly as α-D-glucose-1-phosphate in apple fruit.

Genes encoding sucrose transporters (SUT2, SUT3 and SUT4)present in developing 'Golden Delicious' fruit. The sucrose transporters however showed different expression patterns, with genes encoding SUT4 being highly expressed with peak expression at daa87. Genes encoding SUT2 declined in expression throughout fruit development, while SUT3 increased (Figure 4.7). Considering SUS determines sink strength of developing fruit, the close relationship of SUS and SUT confirms the role of SUS as a determinant of the rate of sucrose import in fruit (Li, *et al.*, 2012).

4.2.3.2.3   Starch metabolism

The amylose and amylopectin are α-1,4 and α-1,6 glucosyl linked starch polymers respectively, constituting the major reservoir of readily available energy and carbon compounds in plants (Zeeman, *et al.*, 2010). It is generally accepted that four enzymes play key roles in starch biosynthesis: ADP-glucose pyrophosphorylase (AGPase, glgC,

151

EC2.7.7.27), starch synthase (WAXY/SS, glgA, EC 2.4.1.21), starch branching enzyme (SBE, EC 2.4.1.18) and starch de-branching enzyme (SDE, EC 2.4.1.15) (Zeeman, *et al.*, 2010).

AGPase is considered the rate-limiting enzyme in starch biosynthesis. AGPase modulates photosynthetic efficiency in source tissues and also determines the level of storage starch in the sink tissues, thus influencing overall fruit starch content at physiological maturity. AGPase is heterotetrameric, consisting of two large regulatory subunits (AGP-L) and two small catalytic subunits (AGP-S) encoded by two distinct genes (Geigenberger, 2011; Kaushik, *et al.*, 2012). Ten transcripts encoding AGP-Ls and four encoding AGP-Ss were expressed in the pulp during apple fruit development. *AGP-L* genes increased in expression from daa25 peaking-up at daa60 followed by a gradual decrease, with the lowest expression at daa150 (Figure 4.4, 4.8). The small subunit *(AGP-S),* on the other hand, had peak expression at daa87 (Figure 4.8). Throughout fruit development the expression of the large subunit was higher than that of the small subunit. The reason for the disparity in the stoichiometric counts of the AGP subunits is not clear, considering they form a 2:2 heterotetrameric AGPase enzyme (Zeeman, *et al.*, 2010). It is possible that the regulation of one unit also controls the assembly of the heterotetrameric enzyme, thus in a way control its activity. The two mRNAs of the AGP subunits might be translated with different efficiency, which could result in the stoichiometric counts of the subunits balancing out after translation.

Starch synthase isoforms, WAXY and SS were encoded for by 8 and 6 transcripts respectively. Starch synthase (EC 2.4.1.21) catalyses the transfer of gylcosyl groups from ADP-glucose, into a growing chain of glucosyl residues (amylose) (Uematsu, *et al.*, 2012). The two starch synthase isoforms showed similar expression profiles, but WAXY was the dominant isoform with higher expression levels throughout fruit development (Figure 4.8). Like AGPase, the expression levels of starch synthase peaked at daa60, possibly, a period of maximum starch accumulation. The pre-climacteric and the climacteric phases are however associated with rapid metabolism (Figure 4.2) of carbohydrates, hence the observed decline in expression. The starch branching enzyme (SBE), catalyses the formation of branched amylopectin from amylose and genes encoding SBE had a transient peak expression at daa60, otherwise increased throughout fruit development (Figure 4.8). Interestingly isoamylase (ISA3, EC 3.2.1.10), an enzyme involved in the degradation of starch gradually increased in expression throughout the development of the fruit (Figure 4.8). The gradual increase in ISA3 throughout fruit development could indicate to the ever-rising demand for energy as the fruit ripens.

The activity of the fructokinase 2 (PFK2, EC 2.7.14) results in the formation of D-fructose-6-phosphate, used in starch biosynthesis (Passam, *et al.*, 2011). As such PFK2 is important in maintaining the flux of carbon towards starch formation. The degradation of starch results in the accumulation of nucleosides. A salvage pathway for the conversion of the nucleosides back to the high-energy nucleotides may be catalysed by fructokinase 2 (PFK2). The expression of PFK2 encoding transcripts decreased with the accumulation of starch, and increased with starch degradation, affirming the importance role of PFK2

in the starch biosynthetic pathway (Figure 4.8). The negative correlation of PFK2 expression to other starch biosynthesis genes could be in support of the second role of PFK2 in salvaging the nucleosides. The expression profiles of starch-biosynthesis-encoding transcripts are in accordance with the period of net starch accumulation in apple fruit (daa35 to daa150) (Figure 4.1).

Transcripts encoding starch-degrading genes: isoamylose 3 (ISA3, EC 3.2.1.20), α-amylose (AMYA, EC 3.2.1.1) and inactive ß-amylose (AMYB, EC 3.2.1.2) gradually increased in expression throughout fruit development, with peak expression at daa150 (Figure 4.8). Genes encoding AMYB however, increased in expression sharply from daa25 peaking up at daa60, followed by a sharp decrease in expression to daa150, a trend similar to WAXY (Figure 4.8).

The calcium-metaloenzyme, AMYA act randomly on starch, unlike AMYB, which hydrolyses starch from the non-reducing end to yield two glucose molecules (maltose) at a time (Rajjou, *et al.*, 2012). The increase in starch degrading enzymes in fruit during ripening is important in giving the fruit its characteristic sweet taste at physiological maturity. The high levels of AMYB throughout fruit development could be proof of their dominance in starch degradation.

Despite AMYB having the highest levels of expression, its sharp decrease after daa60, when expected to further increase, cannot be explained. Interestingly, with decrease in AMYA levels, the levels of the AMYB increased (Figure 4.8). This phenomenon has

been observed in maturing seeds, which increased in the levels of inactive ß-amylase with maturity. The inactive ß-amylase has been reported to be important in the initial starch degradation during seed germination (Rajjou, *et al.*, 2012). The other amylases, α-amylase and γ-amylase, have been reported to only accumulate after germination to argument ß-amylase in providing energy and carbon building blocks (Rajjou, *et al.*, 2012).
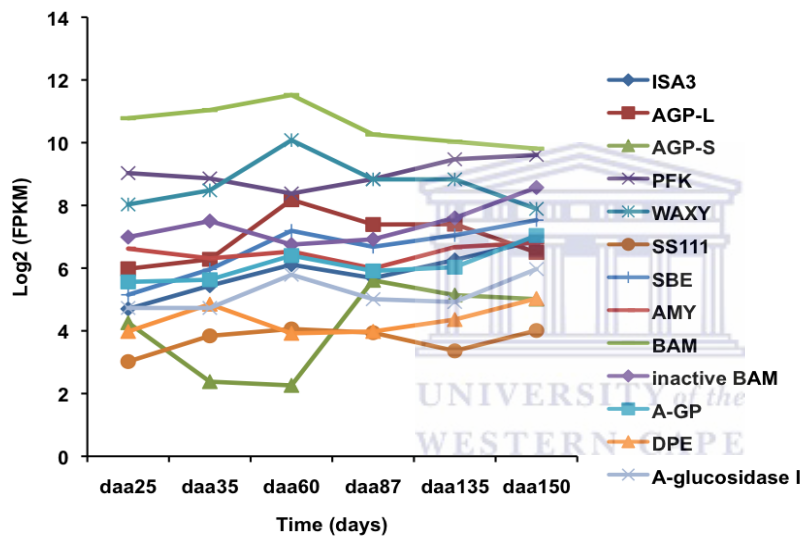
Figure 4. 8 Expression profiles of pulp tissue expressed genes encoding starch metabolism enzymes in developing 'Golden Delicious' fruit. The Figure shows expression trends of more than two-fold up regulated pulp tissue expressed genes encoding starch synthesis and degradation across six harvesting time points. Expression levels are totals of transcript expression per locus, at each time point, from daa25 to daa150. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

Genes encoding desproportionating enzyme (DPE, EC 2.4.1.15), α-glucosidase, and α-1,4-glucan phosphorylase (glgP, EC 2.4.1.1) had similar expression pattern, with transitory peak at daa60 followed by a sharp increase from daa135 to daa150 (Figure 4.8). After physiological maturity has been attained, at daa135, the rate of carbohydrate degradation increases rapidly. In this study the rapid increase in degradation of starch was

collaborated by the increase in expression of genes encoding starch-degrading enzymes with fruit maturity. At any given time point during fruit development, starch is formed, while some of the already formed starch is degraded to provide the growing cells with building carbon blocks and energy to meet the demands of the various physiological processes. Net starch accumulation will however, be determined by an interplay of starch biosynthesis and degrading enzymes under regulation of internal and external factors (Zeeman, *et al.*, 2010).

4.2.3.3 Phytohormone biosynthesis and response

Plant hormones, abscisic acid (ABA), auxin (indole-3-acetic acid, IAA), brassinosteriods (BR), ethylene, and gibberellic acid (GA) control pome development and ripening (Perez-Gil, *et al.*, 2012).
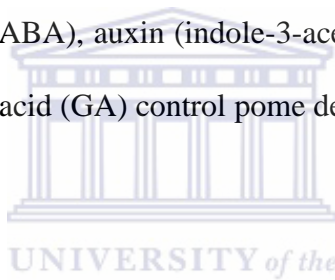
4.2.3.3.1 Abscisic acid

ABA is a terpenoid formed in the chloroplasts from a terpene, isopentenyl diphosphate (IPP), produced from either the mevalonate (MVA) or the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathways (Perez-Gil, *et al.*, 2012).

The expression of the rate-limiting-cleaving enzyme, 9-cis-epoxycarotenoid dioxygenase (NCED) encoding transcripts gradually increased with fruit development (Figure 4.9). In *Arabidopsis*, 9 members of the NCED family are known; five of them (AtNCED1, 2, 3, 5, 6 and 9) are thought to be involved in ABA biosynthesis (Finkelstein, *et al.*, 2002). In this study, however, five NCED gene family members were differentially expressed during the time series under study (i.e. NECD1, NCED3, NCED4, NCED5 and NCED9).

156

The levels of NCED genes are know to increase in response to drought stress (Finkelstein, *et al.*, 2002; Finkelstein and Rock, 2002), mostly AtNCED3 and to a lesser extend AtNCED9. Both members were detected in this study, possibly suggesting fruit stress just before harvesting. It is also known, that there are other factors besides dehydration, such as low or high temperature, salt and flooding conditions that may cause increases in ABA (Finkelstein, *et al.*, 2002).
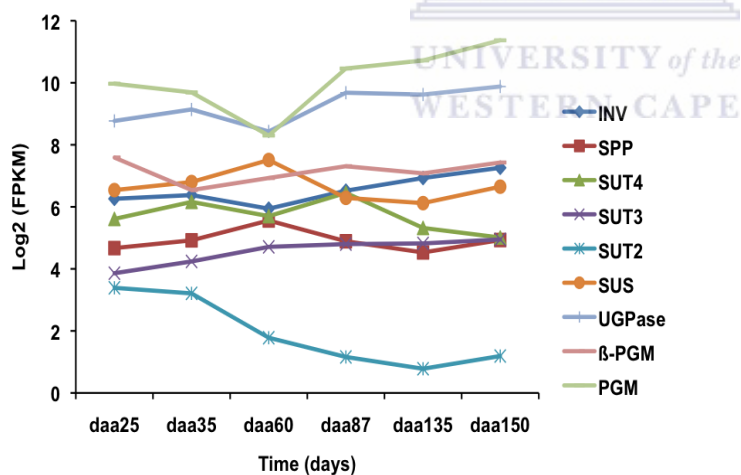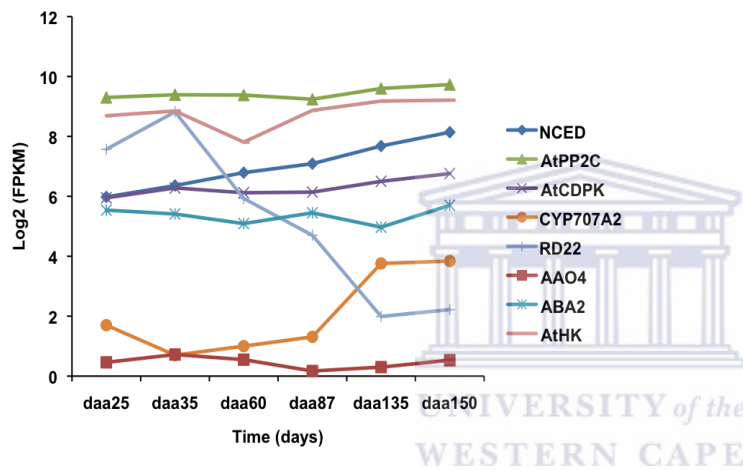

Figure 4. 9 Expression profiles of pulp expressed genes encoding ABA metabolism enzymes. The Figure shows expression trends of more than two-fold up regulated pulp expressed genes encoding ABA metabolism pathway enzymes in developing 'Golden Delicious' fruit at six time points in daa. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

An *Arabidopsis* transmembrane two-component histidine kinase (AtHK1) is a functional osmosensor, which links ABA signalling and osmostress. The expression of AtHK1 showed delayed response to NCED expression (Figure 4.9). Its expression is known to increase with increasing levels of ABA, drought, hypotonic solutions, cold and salt stress, conditions that might arise during fruit development, due to agronomic practices or environmental factors (Finkelstein, *et al.*, 2002).

Xanthanin, a product of NCED is converted by a short-chain alcohol dehydrogenase (ABA2) into abscisic aldehyde, which is oxidized into ABA by an abscisic aldehyde oxidase (AAO4) (Milborrow, 2001). Other than the Arabidopsis AtABA2 genes, there are no other similar genes in different species that have been identified. Here the expression of ABA2 encoding transcripts in apple pulp with almost equal expression throughout fruit development was reported, suggesting ABA2 might not be responsive to fruit development. AAO4 encoding transcripts were lowly expressed with highest expression at daa35, and a general decrease up to daa150 (Figure 4.9). The AAO encoding transcripts expression profile concur with the levels of AAO activity reported to initially increase early in fruit development, followed by a decrease thereafter (Masia*, et al.*, 1998). Four AAO isoforms (AAO1, AAO2, AAO3 and AAO4) are known to exist, however, AAO1 and AA4 have been shown to have an insignificant contribution towards endogenous levels of ABA (Seo*, et al.*, 2004; Seo*, et al.*, 2006) suggesting AAO3, but not AAO (1 and 4) plays a more dominant role in ABA biosynthesis (Seo and Koshiba, 2002). Nonetheless, the observations in this study suggest AAO4 has a significant expression profile, hence possible significant physiological role.

Aside from NCED and ABA 8'-hydroxylase main regulatory steps in the ABA metabolic pathway, metabolic steps upstream of ABA metabolism also contribute to ABA accumulation. Transcripts encoding CYP707A, an ABA-catabolising enzyme, were lowly expressed and only increased in expression from daa87 levelling off at daa135 to daa150 (Figure 4.9). The transcripts expressions do not provide evidence of ABA conjugation

but, though at low levels, show that ABA is modulated endogenously in apple fruit pulp by catabolism.

Transcripts encoding dehydration response protein 22 (RD22) showed an early peak expression at daa35 followed by a decline throughout fruit development (Figure 4.9). In an expression profile analysis of grape berry development, *RD22* was reported to have a large increase with the on set of ripening that continued to increase during berry maturation (Deluc*, et al.*, 2007; Grimplet*, et al.*, 2007). Transcripts encoding calcium-dependent protein kinase (AtCDPK1) and a probable protein phosphatase 2C (AtPP2C), ABA-inducible genes, were shown to also increase with increasing ABA biosynthesis genes. These gene products are involved in storage reserves and/or response to desiccation for the survival of the plant.
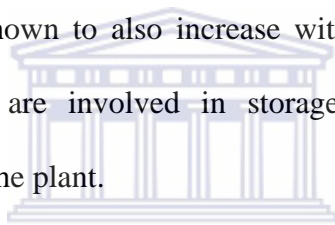
4.2.3.3.2 Ethylene

Apple fruits are climacteric, and ethylene increases pome diameter and ripening processes, such as pigment accumulation and degradation (Alexander and Grierson, 2002; Barry*, et al.*, 2000). The enzymes, 1-aminocyclopropane-1-carboxylic acid (ACC) synthase (ACS) and ACC oxidase (ACO) derive the ethylene biosynthetic process, subject to both positive and negative feedback regulation. ACO activity increase has been shown to precede ACS activity in pre-climacteric fruit in response to ethylene, suggesting its importance in controlling ethylene production (Barry*, et al.*, 2000). Analysis of ACO expression patterns in ripening fruit shows that each gene is highly regulated with transcripts of individual members accumulating to varying degrees at distinct

developmental stages (Barry, *et al.*, 2000). In this study, ACO was highly expressed and increased with pome development (Figure 4.10).

Ethylene receptor (ETR) encoding transcripts declined in abundance from daa25 to daa60 (Figure 4.10), the same trend with ACO encoding transcripts, except for ETR1, which increased after daa60. ETRs were reported by Zegzouti, *et al*. (1999) to be expressed in different temporal and spatial patterns depending on developmental stage and external stimuli (Zegzouti, *et al.*, 1999). ETR are in themselves ethylene inducible and increased levels of receptors reduces ethylene sensitivity, supporting the negative regulation of ethylene receptor model (Ciardi, *et al.*, 2000).

Transcripts encoding multiple ethylene responsive factors RAP2-3 (ERF-RAP2-3) were observed residing on multiple loci. ERF-RAP2-3 encoding transcripts increased in expression gradually throughout development, a trend similar to ACO (Figure 4.10). In grape, however, Deluc, *et al,*. (2007), observed a steady increase in ERF transcripts abundance with maximal expression at ripening (Deluc, *et al.*, 2007). According to the *Arabidopsis* model of ethylene signalling, reduced expression of transcripts and activity of receptors increases the sensitivity to ethylene, whereas increased receptor expression and activity decreases sensitivity (Alexander and Grierson, 2002).
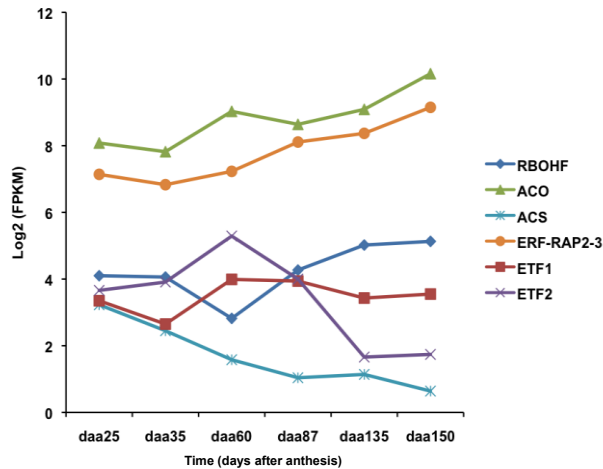
Figure 4. 10 Ethylene biosynthesis and perception encoding transcripts expression profiling. The Figure shows expression trends of more than two-fold up regulated pulp expressed genes encoding ethylene metabolism and perception pathway proteins in developing apple fruit at six harvesting time points. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

The *Respiratoryburstoxidasehomolog* (*Rboh*) gene family encodes the key enzymatic subunit of the plant NADPH oxidase. Rboh proteins are the source of reactive oxygen species (ROS), which play a key signal transduction role in cells. They are involved in the regulation of growth, development, responses to environmental stimuli and cell death (Foreman*, et al.*, 2003; Suzuki*, et al.*, 2011; Torres and Dangl, 2005). Transcripts encoding *RbohF* were highly expressed and declined in expression from daa25 to daa60 followed by a rapid increase from daa60 to daa150 (Figure 4.10), a trend similar to ETR1, possibly to signal the start of the ripening process. The pre-climacteric ethylene pathway activation appears to occur around daa35 in response to ETF2, which showed rapid increase from daa35 peaking up at daa60. The increase in ETF2 might have resulted in the activation and gradual increase in ACO expression from da35 to daa150. Sustenance of the increasing ACO levels and the building up ethylene levels was supported by an increase in ERF-RAP2-3 (Figure 4.10).

ACS encoding transcripts decreased with fruit development, and may be important in initiating ACO synthesis. It is possible the pre-climacteric phase may be started at daa35 rather than after daa60 when starch accumulation reaches its peak. An initial decrease in respiration characterizes the pre-climacteric phase before the ultimate respiratory burst in climacteric fruits (Lelièvre, *et al.*, 1997; Li, *et al.*, 2012). This characteristic is supported by the decrease in the expression of RBOHF encoding transcripts after daa35 to daa60 followed by a sharp increase to daa150. RBOHF expression pattern mimics that of ethylene genes *ACO* and *ERF-RAP2-3* suggesting a possible close relationship, though the mechanism of the interplay among these genes is not clear. What is known, however, is that RBOHs are able to integrate calcium signalling and protein phosphorylation with ROS production, in many different biological processes in cells, placing RBOHs at the centre of the ROS network of cells (Lin, *et al.*, 2009).
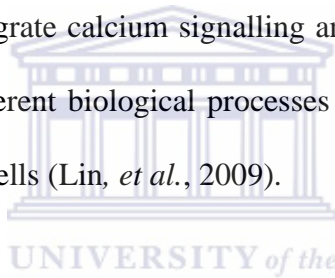
4.2.3.3.3 Brassinosteroids

Brassinosteriods (BR), the polyhydrated steroid hormones of plants are linked to many developmental processes, including photomorphogenesis. BR increase cell size and correlate to castasterone levels in the cell(Ashraf, *et al.*, 2010; Bajguz, 2011; Davies, 2010). BR insensitive 1 (BRI1), a plasma-localized receptor serine/threonine kinase, is essential for BR perception and accounts for most of BR binding activity in *Arabidopsis*(Nemhauser, *et al.*, 2004; Vert, *et al.*, 2008). Transcripts encoding BRI1 were identified and shown to increase in expression with peak expression at daa135 (Figure 4.11). Considering BRI strongly correlates to the levels of BR in plant tissue, the levels of BRI encoding transcripts may therefore be indicators of the levels of BR in the plant at any given time (Wang, *et al.*, 2007). The increase in BRI1 encoding transcript

abundances may suggest increased capacity for binding BR, promoting cell elongation (Senadheera and Maathuis, 2009; Wang, *et al.*, 2007). The period of the final increase in BRI1 coincides with the period of YUCCA2 increase in auxin biosynthesis (Figure 4.12) during the final cell elongation and expansion phase, at daa87 suggesting co-regulation of cell expansion by BR and Auxin.

A shaggy/KSG8 protein kinase (ASK8) and BR insensitive (BIN) are negative regulators of the BR pathway downstream of BRI action (Nemhauser, *et al.*, 2004; Vert, *et al.*, 2008). Brassinosteroid-associated kinases (BAK) are receptor-like protein kinases (RLKs) involved in BR signal transduction and plant resistance. The expression of BAK suggests an ever presence of the plant defence system, especially early in fruit development when other physical barriers such as waxy formation are still being strengthened. The relatively high BAK levels throughout fruit development (Figure 4.11) provides BRI1 with ready receptors for signal transduction, and also implies that BAK may not be the major bottle neck in BR synthesis and signalling.
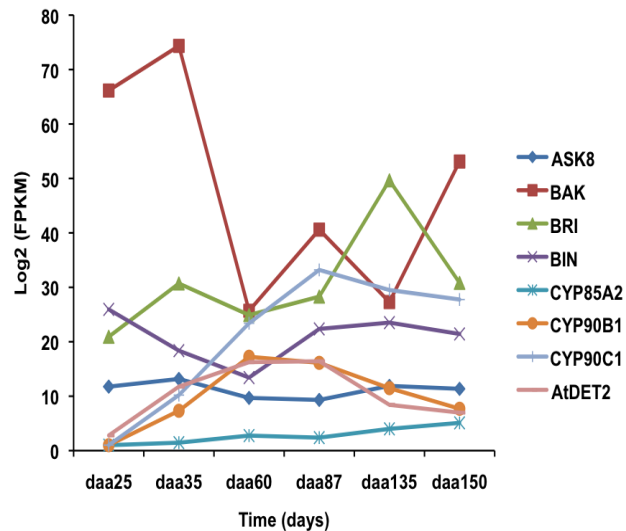
Figure 4. 11 Brassinosteroids (BRs) synthesis and perception encoding transcripts expression profiling. The Figure shows expression trends of more than two-fold up regulated pulp expressed genes encoding BR metabolism and perception pathway proteins in developing apple fruit at six harvesting time points.
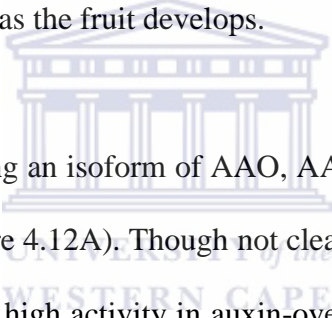
Transcripts encoding two monooxygenases, CYP85A2 and CYP90C1 were detected in 'Golden Delicious' fruit pulp. CYP85A2 encoding transcripts were lowly expressed with a gradual increase throughout fruit development (Figure 4.11). CYP90C1, on the other had a higher level of expression with peak expression at daa87 followed by gradual decrease (Figure 4.11). Steroid 5-alpha-reductase DET2 (AtDET2) acts up-stream of steroid 22-alpha-hydroxylase (CYP90B1) during formation of brassinolides. Transcripts encoding the enzymes, AtDET2 and CYP90B1, had similar levels and pattern of expression (Figure 4.11), suggesting they are both regulated at the same levels during brassinolide formation. Clearly, there are a number of significant changes in transcript abundance that are associated with BR response during pome development.

4.2.3.3.4 Auxins

Auxins have profound plant growth and development modulation effects. Biosynthesis of auxins is thought to be through two pathways, tryptophan (trp)-dependant and trp-independent (Zhao, 2010). The indole-3-acetic acid (IAA) pool is augmented by hydrolytic cleavage of IAA-conjugates (IAA-amides, IAA-sugar, and IAA-methyl-esters). In the trp-dependant pathway, it is suggested that trp is converted to IAA through possibly three sub-pathways: Indole-3-acetaaldehyde (IAOx)-glucosinolate (YUC) and indole-3-pyruvate (IPA) pathways. None of these pathways however has been fully elucidated (Zhao, 2010). Transcripts encoding two Flavin-containing monooxygenase-like *YUCC10* and *YUCCA2* genes were detected in apple pulp transcriptome with significant expression during fruit maturity. Both genes, *YUCC10* and *YUCCA2* had very low levels of expression throughout fruit development (Figure 4.12A). YUCC10 was highly expressed early in fruit development decreasing at daa35 to a constant level for the rest of the period of fruit development. YUCCA2 on the other hand, was lowly expressed during early fruit development but had peak expression at daa87, coinciding with the final increase in fruit volume between daa87 and daa135. It is possible YUCCA2 is responsible for the final cell expansion and fruits attaining their potential size. No transcripts encoding TAA1 and CY7907A flavan monooxygenases were detected suggesting the predominance of the YUC pathway in apple pulp auxin biosynthesis.

In *Arabidopsis*, seven members of the amidase share a conserved glycin and serine rich amidase (AMI) signature, and convert indole-3-acetamide (IAM), into IAA. AMI expression is strongest in places of highest IAA content in the plant (Stepanova*, et al.*, 2011). The transcripts encoding AMI showed peak expression at daa60 (Figure 4.12A) a

165

similar trend to IAA-amino acid hydrolase (ILR) and auxin responsive factors (ARF2 and ARF8, Figure 4.12A, 4.12B). Anthranilate synthase (ASA) is involved in tryptophan biosynthesis, which produces precursors for IAA biosynthesis. The enzyme is negatively controlled by tryptophan accumulation and hence IAA levels in the plant tissues through feedback inhibition (Caligiuri and Bauerle, 1991; Tozawa, *et al.*, 2001)). ASA encoding transcripts showed a progressive increase in abundance from daa25 peaking up at daa135 (Figure 4.12A). The observation supports the notion for the high demand of auxin hormone early in fruit development, for rapid cell proliferation and enlargement as well as for the final cell expansion just before physiological maturity. The levels of auxin are then required in lesser amounts as the fruit develops.

In this study transcripts encoding an isoform of AAO, AAO4, were lowly expressed with peak expression at daa35 (Figure 4.12A). Though not clear, AAO3 isoform is reported by Sekimoto, *et al,* (1998) to have high activity in auxin-overproducing-super root 1 mutant of *Arabidopsis*(Sekimoto, *et al.*, 1998). AAO involvement in IAA biosynthesis may be activated by auxin or auxin-induced transcription factors. The actual role of AAO in auxin biosynthesis however is still undefined.

A number of auxin responsive factors (ARFs 1-19) encoding transcripts were detected, however, none had an expression pattern similar to YUCCs, rather ARF2 had a similar profile to AMI, suggesting regulation of the terminal auxin biosynthesis enzyme, AMI, by ARF2. The synthesis of IAA conjugates, for both permanent inactivation and temporal storage of auxin, controls IAA concentrations during apple fruit development. The

hydrolysis of these conjugates by IAA-amino-acid hydrolase may provide for local concentrations of auxin within the pome fruit (Nemhauser*, et al.*, 2004), promoting mesocarp cell enlargement. The absence of a full compliment of annotated auxin metabolism and perception genes makes it difficult to identify and annotate high throughput output transcripts by similarity, owing to their great numbers. The challenge thus is in enriching the annotation databases with properly annotated and curetted genes for reference.
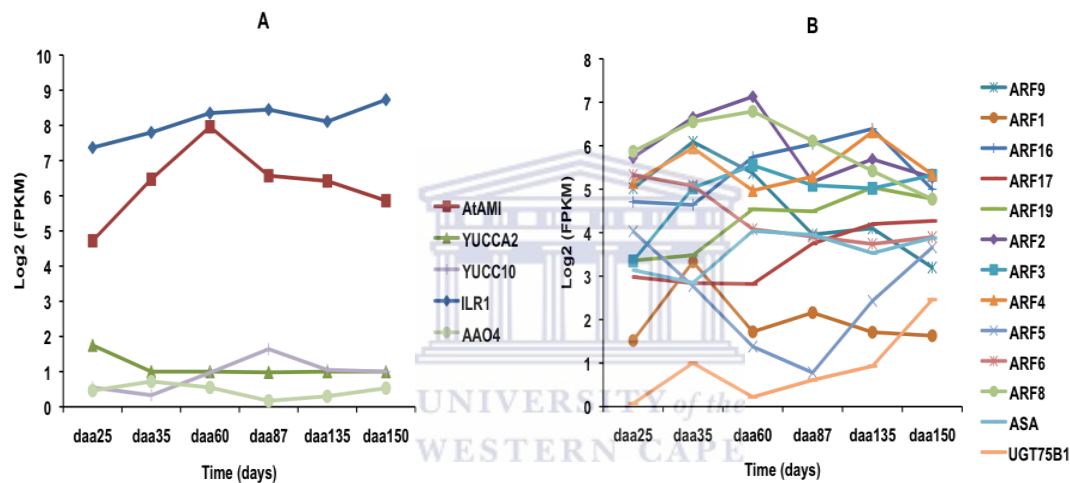
Figure 4. 12 Auxin metabolism and perception encoding transcripts expression profiles. The Figure shows expression trends of more than two-fold up regulated pulp expressed genes encoding BR metabolism and perception pathway proteins in developing apple fruit at six harvesting time points. A: Expression trends of pulp up regulated genes encoding for auxin biosynthesis enzymes, B: Expression profiles of pulp up regulated genes encoding auxin responsive factors (ARF) involved in auxin signal transduction. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

4.3 Conclusion

The pulp constitutes the major portion of the apple fruit that is consumed as well as a reservoir of compounds, such as carbohydrates, that attract seed dispersers. To optimise the properties of this tissue conferring sugar content an in depth understanding of the patterns of expression of genes related to the carbohydrate accumulation and hormonal changes and regulation is required. In the work described in this chapter, the aim was to

167

elucidate the carbohydrate and hormonal developmental changes during fruit development using high throughput sequencing of 'Golden Delicious' apple fruit pulp mRNA transcriptome. The results demonstrated that sugar content in 'Golden Delicious' positively correlates with an increase in fruit fresh weight.

Transcripts reconstructed from pulp tissue were grouped into eleven main groups showing; an increase or decrease throughout fruit development; transitory peaks or decreases at daa35, daa60, daa87, and daa135; as well as dynamic expression pattern with multiple troughs at daa35 and daa135. Most of the transcripts that decreased with fruit development encoded genes in the plant hormone signal transduction, plant pathogen interaction, ribosome, photosynthesis and amino sugar and nucleotide sugar metabolism pathways. Transcripts that increased in expression with fruit development on the other hand, encoded genes in the plant hormone signal transduction, carbon fixation, amino sugar nucleotide sugar metabolism, peroxisome, plant pathogen interaction and oxidative phosphorylation. A huge decrease in the transcripts encoding protein processing in endoplasmic reticulum and oxidative phosphorylation were observed at daa60. The reconstructed transcripts were further annotated using the main MIPS categories, and daa150 had the highest frequency of transcripts in all the functional categories. Functional categories of protein with binding function, metabolism, interaction with environment and cell defence were the most enriched categories.

The carbohydrate metabolism and hormonal biosynthesis and regulation pathways were looked at in detail to show the characteristics of the pulp tissue. Genes encoding Calvin

cycle, sucrose as well as starch biosynthesis had peak expression between daa60 and daa87 a possible period of peak starch accumulation. Transcripts encoding starch and sucrose degrading increased in expression with fruit development.

Genes encoding ABA biosynthesis pathway proteins increased in expression with fruit development suggesting possible increase in fruit stress with fruit development. Ethylene biosynthesis encoding transcripts increased with fruit development except for ACS, which decreased with fruit development. Genes encoding BR biosynthesis increased with fruit development peaking up at daa135 just after YUCCA peak expression at daa87, suggesting possible involvement of both BR and auxin pathway enzymes in the final fruit cell expansion before physiological maturity. This study has demonstrated that fruit development; carbohydrate accumulation and hormonal metabolism and perception are intricately related. The extensive pulp catalogue of gene expression patterns defined here will serve as an invaluable reference for future fruit quality investigations including the exploration of the complex transcriptional regulatory hierarchies that govern apple pulp development and differentiation.

5   CHAPTER 5

Developmental mRNA profiling in *M. x domestica* apple fruit enriched for peel tissue

Abstract

The peel is an invaluable tissue of the apple fruit, accounting for fruit traits such as tensile strength, glossiness, colour, flavour as well as for plant's survival against desiccation, pathogen attack and burst from internal pressure. An RNA-seq analysis of the 'Golden Delicious' fruit peel mRNA transcriptome at six time points (25, 35, 60, 87, 135, and 150 days after anthesis, daa) revealed that cell fate; interaction with environment; protein with binding function; development and metabolism were over-represented across all the time points. Genes encoding cuticle biosynthesis increased with fruit maturity, suggesting thickening of the wax layer with fruit development. Genes encoding the mevalonate (MVA), 2-C-methyl-D-erythritol4-phosphate (MEP) and ester biosynthetic pathway enzymes increased in expression with time of harvesting, suggesting an increase in aroma compounds with fruit maturity. Genes encoding anthocyanin biosynthetic pathway enzymes had high expression during early fruit development, similar to phenol peroxidase (PPO) expression, suggesting PPO might be responsible for the degradation of anthocyanins. The change in fruit colour to golden yellow at physiological maturity was corroborated by an increase in the expression of genes encoding carotenoid biosynthesis enzymes. Genes encoding cell wall loosening related proteins decreased in expression with fruit maturity, except for expansin A8 (EXPA8) and expansin A10 (EXPA10), which had peak expression at daa87 and daa60 respectively. The results of this experiment provide a comprehensive catalogue of peel wax accumulation, flavour, pigment and cell wall related metabolic changes during 'Golden Delicious' fruit development important in defining fruit quality traits.

Key words: Fruit peel, cell wall, pigment accumulation, flavour, fruit development, wax accumulation

5.0 Introduction

The anatomical structure of apple fruit is composed of different tissue types. It consists of the mesocarp, formed from the ovary walls surrounding the placental tissue, the locular tissue, and the seeds, as well as the hypanthium. The hypanthium, formed from the fused bases of petals and calyx, enlarges during fruit development into mesocarp and exocarp (Merzlyak*, et al.*, 2008; Merzlyak*, et al.*, 2005; Solovchenko and Merzlyak, 2003; Solovchenko*, et al.*, 2010). The edible fleshy mesocarp encompasses layers of large, highly vacuolated parenchymatous cells and contains vascular bundles. The exocarp is composed of several layers of collenchymatous cells and a single layer of epidermal cells, which in turn is covered, and in some cases encased, in a waxy cuticle that thickens as the fruit develops (Merzlyak*, et al.*, 2008; Merzlyak*, et al.*, 2005; Solovchenko and Merzlyak, 2003; Solovchenko*, et al.*, 2010). Loosely used, the term 'peel' denotes the outer layers, epidermis, collenchyma, and sometimes parenchyma, depending on how the peel is removed.

The fruit peel is crucial to fruit and plant survival, it serves as the interface between fruit and the environment. In 'Golden Delicious' fruit, the peel tissue has the highest pigmentation concentration (Merzlyak*, et al.*, 2008; Merzlyak*, et al.*, 2005; Solovchenko and Merzlyak, 2003; Solovchenko*, et al.*, 2010). During the early stages of fruit development the fruit peel is reddish, but turns green after 30 to 35 days after anthesis (daa) and then to a golden colour at physiological maturity. The fruit peel colour changes have been shown to relate to changes in chlorophyll, flavonoid and carotene concentrations with fruit development. Each physiological pigment and hence colour change, is tailored to protect the fruit from excessive light effects or for photosynthesis.

Fruit photosynthesis can contribute up to 10% of fruit carbon needs (Solovchenko, *et al.*, 2010). The decreasing fruit photosynthetic capacity during fruit development has been attributed to decreasing chlorophyll concentration; reduced stomatal density as well as the biological clock rendering the fruit a strong sink of photosynthates as it matures. Carotenoids, on the other hand, increase in concentration with maturity giving 'Golden Delicious' its characteristic golden colour at physiological maturity (Solovchenko, *et al.*, 2010).

The cuticle wax encasing the epidermis gives the fruit its glossiness, and protects the fruit from desiccation by minimizing non-stomatal water loss. The shiny surfaces of the cuticle-encased fruit, as well as terpenoids and flavonoids, protects the fruit from the effects of excessive light (Solovchenko, *et al.*, 2010). The mechanical strength provided by cuticle protects the fruit from mechanical damage, thus protecting it from insect bites or growing fungal hyphae, as well as from inside by increasing turgor pressure. Mechanical failure of fruit surfaces leads to substantial losses of production through fruit bruising, splitting or cuticle cracking. As such, the properties of the peel influence the outward appearance of the fruit (in terms of colour, glossiness, texture, and uniformity), efficacy of postharvest treatments, storage, transport, and shelf life (Mintz-Oron, *et al.*, 2008). Hence, knowledge regarding fruit peel properties is fundamental for the improvement of fruit quality traits.

Despite the availability of molecular tools and the importance of the peel tissue, both in terms of fruit biology and in relation to fruit quality traits, only a limited number of

studies have investigated the fruit peel biology at the molecular level to date (Lin-Wang, *et al.*, 2011). This study aims therefore to obtain a better understanding of the mechanisms and processes defining fruit peel properties in 'Golden Delicious'. This goal is better achieved by dividing the fruit into separate tissues in order to obtain transcripts enriched for peel, which would otherwise be missed in pooled sampling. Here global profiles of transcripts changing in expression during fruit development as well as detailed profiles of transcripts encoding enzymes for cuticle, terpenoid, flavonoid, carotenoid, chlorophyll as well as cell wall metabolism were provided.

5.1 Functional Annotation

5.1.1   Clustering of apple fruit peel tissue significantly expressed genes

A reference transcriptome of 186498 transcripts was reconstructed from combined apple fruit peel and pulp tissue extracted mRNA as outlined in section 2.2.9.1 and 2.2.9.2. Differential transcript expressions were then calculated using RSEM (section 2.2.15) expressed as FPKM. A set of transcripts significantly expressed in fruit peel tissue was obtained by performing a two-way ANOVA with two tissue types and six time points as factors at a p-value of 0.05. Out of 186498 total transcripts, 27255 transcripts were significantly expressed in 'Golden Delicious' apple fruit peel tissue and across the six time points.

The transcripts significantly expressed in the fruit peel tissue were then clustered into closely correlated groups of transcripts with statistically significant ($p > 0.05$, and False Discovery Rate (FDR) $= 0.05$) similar expression using the *k-means* method in the MeV package (Saeed, *et al.*, 2003) to determine cluster membership (Figure 5.1). The choice of

the number of clusters was optimised to 20 clusters using Figure of Merit (FOM) (Yeung, *et al.*, 2001) as shown in section 2.2.17. Within the 20 closely correlated clusters (Figure 5.1), eleven general expression trends were noted and clusters with similar expression trends were regrouped into eleven groups of distinct general expression trends (Figure 5.1). Groups one and two constituted 59% of the total peel tissue significantly expressed transcripts with the majority clustered in group two (32%). The majority of the 8636 transcripts in group two were clustered in cluster 10 with transcripts increasing from daa60 to a plateau at daa87 to daa150 (Figure 5.1).

Each cluster group was analysed for pathway enrichment using KOBAS (Xie*, et al.*, 2011) to reveal the trend of physiological processes affected by enriched pathway genes encoded by the reconstructed transcripts (section 2.2.17) (Table 5.1). Transcripts encoding proteins in the RNA transport pathway were the most enriched in group one (57) and 54 in group two. RNA species, tRNA, rRNAs, snRNAs and mRNAs, are produced in the nucleus and are exported through the nuclear pore complexes (NPCs) by specific export receptors. This process is fundamental for gene expression (Merkle, 2011; Vazquez-Pianzola and Suter, 2012). The level of the RNA transport enrichment could be an indicator of the amount of RNA load in the nucleus, which could also reflect the rate of transcription in a cell. Across the groups RNA transport pathway encoding transcripts are expressed (Table 5.1), but were most enriched in groups one and two. This observation correlates to the number of transcripts in a group, supporting the notion of a close relationship between transcript number and transport. The enrichment of ribosome and RNA-transport pathways early in fruit is in agreement with the intense cell division

and cell enlargement processes characteristic of early fruit development, which requires a significant amount of protein and nucleic acid for replication.

Photosynthesis and carbon fixation pathways were most enriched in group two, which showed an increase in expression with fruit development.

Flavonoid biosynthesis pathway encoding transcripts were most enriched in group one. 'Golden Delicious', unlike most red skinned apple fruits, does not accumulate substantial amount of the colourful flavonoids (Fleancu, 2007). Nonetheless during early fruit development in 'Golden Delicious' the fruits are reddish in colour, turning green and then golden yellow at physiological maturity. The decreasing enrichment of the transcripts encoding the flavonoid biosynthesis pathway is thus in agreement with the colour changes with fruit development.
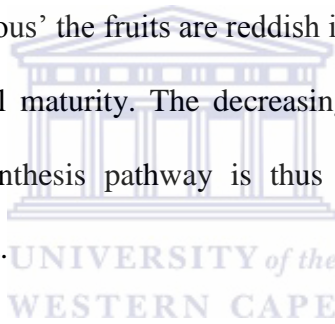
Plant hormone signal transduction and pathogen interaction pathway encoding transcripts were also enriched in groups one and two (Table 5.1). The prolonged exposure of the developing fruit to the harsh environment results in a constant interaction of the fruit with various pathogens. Each encounter elicits a plethora of new and acquired defence mechanisms. As such the enrichment of the plant pathogen interaction pathway with fruit development is in agreement with the continued developing plant pathogen response.

The endoplasmic reticulum (ER) is the centre for protein folding into various primary, secondary, tertiary and quaternary structures suited to a physiological function. Correctly

175

folded proteins are shuttled to the Golgi body, while miss-folded proteins are retained and marked for ER associated degradation (Deshpande*, et al.*, 2008; Lu*, et al.*, 2012).
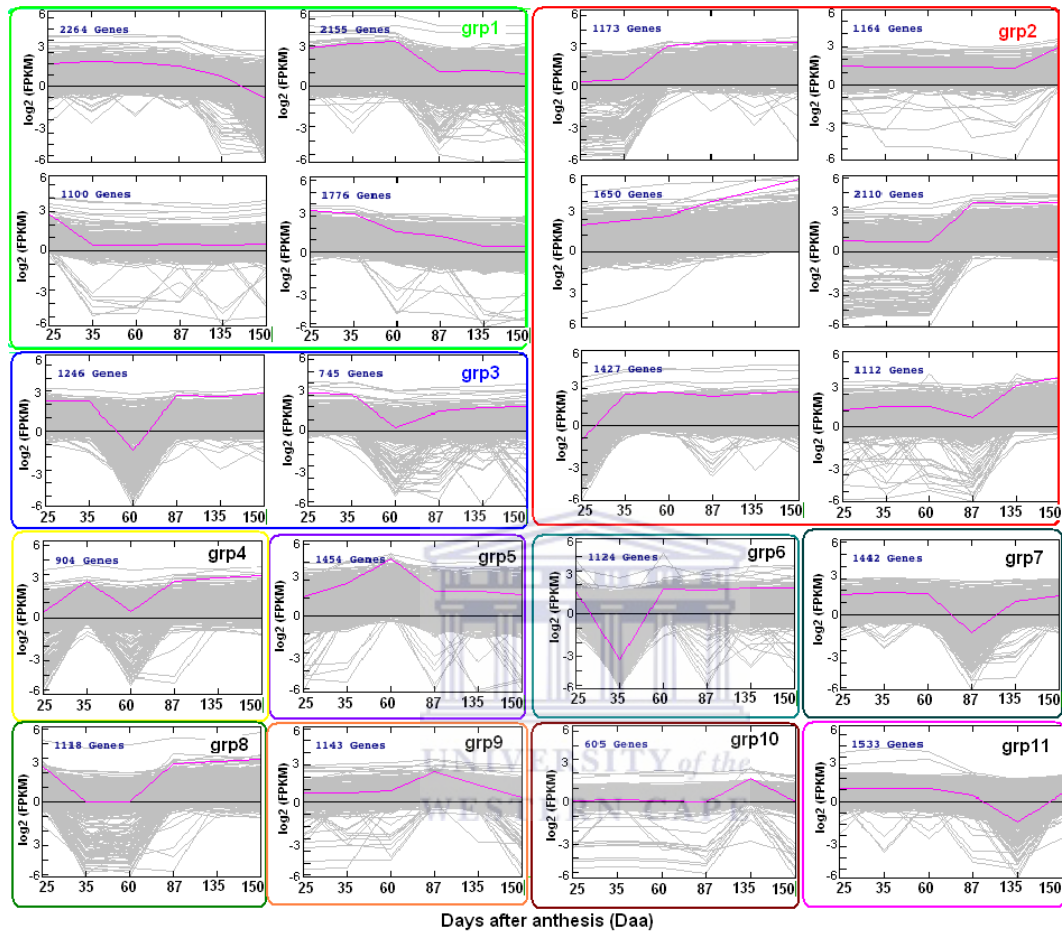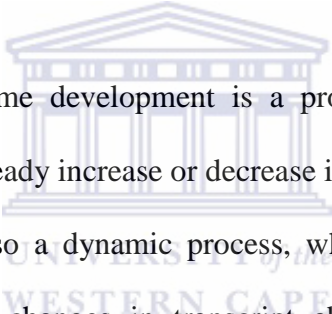
Figure 5. 1 Cluster analysis of statistically significantly peel tissue expressed genes using *k-means*. Clustering of transcripts with respect to expression levels was done using *k-means* method and Pearson Correlation at 5% critical value available from MeV 2.0.2 (Saeed*, et al.*, 2003). Differentially expressed transcripts obtained using RSEM were subjected to a two-way ANOVA to identify peel tissue significantly expressed transcripts, which were then separated into closely correlated clusters of expression trends. The Figure of Merit (FOM) was used to optimise cluster number to twenty. Within the twenty clusters, eleven general expression trends were observed and the twenty clusters were re-grouped into the eleven general trend groups. The general trend groups depicted: a decrease with fruit development (group one); an increase with fruit development (group 2,); a decrease at daa60 (group 3); transient peak at daa35 and an increase from daa60 (group 4); peak at daa60 (group 5); a trough at daa35 (group 6); trough at daa87 (group 7); trough at daa35 to daa60 (group 8); peak at daa87 (group 9); peak at daa87 (group 10); trough at daa135 (group 9). Grp=group, Group numbers are written in the right upper corner of each colour-coded cluster group. The number of genes per cluster is shown in the left upper corner. The horizontal axis is the harvesting time points (daa) and the vertical axis is the transcripts. Expression values (FPKM) were $\log_2$ transformed to normalize the data (section 2.2.17).

However sometimes the miss-folded proteins accumulate and do not trigger the unfolded protein response and the whole cell dies due to apoptosis. The increase with fruit development (group two) of transcripts encoding enzymes in the protein processing in endoplasmic reticulum pathway reflects intense protein processing, which correlates with the enrichment of the ubiquitin mediated proteolysis (Table 5.1). The huge volume of the proteins passing through the folding machinery may also result in increased rates of miss-folding, hence triggering their degradation as well. This close relationship between protein folding and ubiquitin mediated proteolysis assists the cell in keeping check of the levels of miss-folded proteins, in order to avoid cell apoptosis.

This analysis showed that pome development is a progressive process, wherein the majority of the genes show a steady increase or decrease in expression across all stages of development (profiles), but also a dynamic process, wherein a substantial number of genes exhibit large, transient changes in transcript abundance at specific times of development.

Table 5. 1 KEGG pathway enrichment analysis of the peel expressed reconstructed transcripts

KEGG pathways	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11
Amino sugar and nucleotide sugar metabolism	43	54	9	8	3	12	7	6	12	10	0
Aminoacyl-tRNA biosynthesis	21	28	4	4	2	3	10	4	3	5	0
Arginine and proline metabolism	10	22	6	7	2	2	1	2	2	1	0
Carbon fixation in photosynthetic organisms	17	63	13	3	4	8	8	6	1	3	2
Endocytosis	20	46	10	4	10	4	5	5	6	2	2
Flavonoid biosynthesis	19	10	1	1	0	0	1	1	0	0	0
Oxidative phosphorylation	27	51	20	6	15	10	5	8	4	4	1
Peroxisome	18	48	13	7	6	7	4	5	3	6	4
Photosynthesis	14	20	9	1	0	2	3	9	2	1	0
Plant hormone signal transduction	54	58	13	2	7	9	12	10	16	10	3
Plant-pathogen interaction	41	62	13	5	10	23	10	7	9	8	4
Protein processing in endoplasmic reticulum	18	74	18	7	12	6	9	6	6	6	1
Ribosome	20	12	16	1	2	2	1	8	2	0	0
Ribosome biogenesis in eukaryotes	15	16	3	3	1	5	4	1	1	1	3
RNA degradation	11	35	13	3	6	7	6	7	2	2	0
RNA polymerase	1	6	2	0	0	1	2	1	0	0	0
RNA transport	57	54	14	4	7	14	13	6	9	11	4
Spliceosome	42	36	13	5	5	8	14	7	8	9	2
Ubiquitin mediated proteolysis	12	27	7	6	4	1	4	3	2	3	0

The most enriched pathways per cluster pooled into groups with similar expression pattern and changes thereof across the eleven expression patterns. Group = a collection of clusters of similar expression profiles, Pathway = KEGG accession of each pathway, KEGG pathways = a description of the KEGG pathway.

5.1.2 Functional profiling of up and down regulated transcripts in apple fruit peel tissue

At daa25 the fruit's total cell number has already been determined, and the processes of rapid cell expansion and enlargement have started. The fruit cells are still differentiating and will continue to do so till physiological maturity, at about daa135. These cell changes are accompanied by a plethora of physiological processes essential in the provision of energy, proteins, and membrane and cell wall building blocks. Each physiological change is important in defining the ultimate quality of the mature apple fruit. The MIPS functional category has divided the major changes in plant physiological processes into 22 main functional categories.

The relative distributions of the reconstructed genes within each of the twenty-two main functional MIPS categories were evaluated (Figure 5.2). Transcripts from each time point were profiled and compared across other time points throughout fruit development. Statistically significant differences in the distribution of genes within functional categories amongst these developmental stages were observed (Figure 5.2).
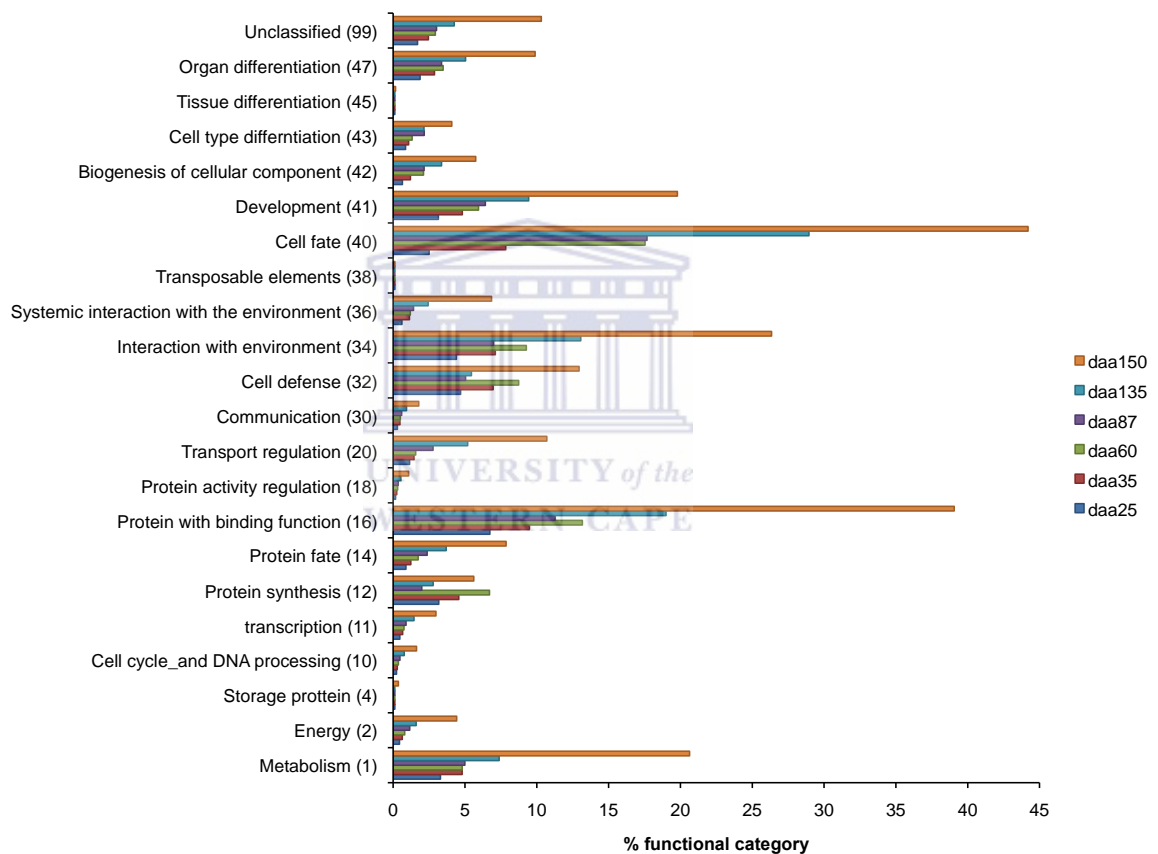
Figure 5. 2 Comparison of peel gene transcripts by function. The functions of the transcripts were assigned by similarity to *Arabidopsis* top hits MIPS functional categories (MIPS 2.0). Percentages are based on the number of Unigenes in each set. Number in parenthesis following category names indicates the MIPS number for each category.

In 21 of the 22 functional categories, daa150 had the highest percentage of expressed transcripts (Figure 5.2). Across 18 categories the percent enrichment of transcripts within a category increased with an increase in harvesting time from daa25 to daa150 (Figure

5.2), except for protein synthesis, protein with binding function, cell defence, and interaction with environment. The categories cell fate, protein with binding function, interaction with environment, development and metabolism in decreasing order were the most enriched from daa25 to daa150, with cell fate being the most over represented category. Daa25 and daa35 time points were mostly enriched in transcripts encoding proteins with binding function. Tissue differentiation, transposable elements, protein activity regulation and storage proteins categories were under-represented across all time points.

Low expression of genes encoding storage proteins (Figure 5.2) is expected considering apple has less than 0.26 g protein per 100 g serving according to the United States 'Department of Agriculture (http://www.ndb.nal.usda.gov, cited 2012.09.25). The pome fruit is not known to accumulate or store protein, but starch and other carbohydrates in plastids.

Transposable elements, 'jumping genes', are activated and expressed in various apple tissues resulting in the regulation of several genes. Peel colour in pome fruits may be regulated by transposable elements in addition to other transcription factors and structural genes (Telias, Bradeen*, et al.*, 2011; Telias, Lin-Wang*, et al.*, 2011). Though discovered in pome fruits, the evidence associating colour variation and transposable elements is yet to be found (Telias, Lin-Wang*, et al.*, 2011). It was assumed that pome fruit gene regulation is not principally driven by transposable elements, considering there are only 42 transposable elements discovered to date in apple (Velasco*, et al.*, 2010). This

therefore could mean that most of the gene expression regulation is controlled most likely by transcription factors and structural genes in response to environmental factors.

5.1.3   Genes associated with important molecular events of apple fruit peel development

To identify genes with potentially important roles in peel development at during fruit maturity, transcripts were identified from within the *k-means* algorithm-defined profile groups. Transcripts were annotated for KEGG pathways as outlined in section 2.2.17, and genes encoding cuticle, flavour formation (mevalonate, MEP and ester metabolism), colour formation (anthocyanin and carotenoid metabolism), carbon fixation, cell wall metabolism (pectin and cellulose metabolism) pathways, were analysed.

5.1.3.1 Cuticle metabolism and regulation

The cuticle is a spatially continuous lipophilic layer sealing epidermal cells. It is involved in restricting non-stomatal transpiration, taking part in plant-pathogen interactions, protecting plants against excess irradiation as well as protecting fruit cells against uncontrolled water uptake (Albert, *et al.*, 2013). The epicuticular waxes may contain hydrocarbons, alcohols, aldehydes, fatty acids, diols, esters, ß-diketones, terpenoids and phenolic compounds and their various compositions determine the wax layer structure. Cuticle components are synthesised in several cellular compartments (Aharoni, *et al.*, 2004; Albert, *et al.*, 2013). C16 or C18 chain length fatty acids are synthesized in the plastids, followed by further extension in elongation complexes residing in the endoplasmic reticulum. The resulting very long chain fatty acids are modified further to different wax compounds that are finally loaded onto the outer surface of the epidermal

cells (Curry, 2009; Kunst and Samuels, 2009; Solovchenko and Merzlyak, 2003). Because of the agronomic importance of the cuticle, as well as absence of detailed elucidation of the genes involved in fruit cuticle production, this section attempts to shed light on the changes in cuticle metabolism in peel tissue during apple fruit development.

Initiation of fatty acid synthesis

Acetyl-Co-enzyme (CoA) produced from photosynthate is required for the *de novo* biosynthesis of fatty acids in the chloroplasts. Activation of acetyl-CoA to malonyl-CoA in the plastids is catalysed by acetyl-CoA carboxylase (ACC, EC 6.4.1.2) and is the first committed step in fatty acid biosynthesis (Nobusawa*, et al.*, 2013; Sayanova*, et al.*, 2012). ACC is a bi-functional enzyme with a biotin and trans carboxylase activity both of which are required in converting acetyl-CoA to malonyl-CoA. Genes encoding ACC, ACC carboxyl transferase subunit alpha (ACCA) and ACC biotin carboxyl carrier protein (ACCB) were significantly expressed in peel tissue (Figure 5.3) and increased in expression with fruit development (Figure 5.4A). The increase in the expression of the ACC fatty acid *de novo* synthesis rate-limiting enzyme suggests a progressive increase in the supply of malonyl-CoA residues for fatty acid synthesis. The increase in ACC expression was also corroborated by an increase in genes encoding malonyl-CoA:ACP-acyltransferase (FabD, EC 2.3.1.39) (Figure 5.3A). FabD is involved in the transfer of malonyl-CoA to acyl carrier protein (ACP), a component of the fatty acid synthase (FAS) multiprotein complex (Nobusawa*, et al.*, 2013; Sayanova*, et al.*, 2012). An increase in the expression of FabD suggests an accumulation of malonyl-ACP residues in the plastids,

which are used as donors of two-carbon units throughout fatty acid chain elongation, a reaction catalysed by FAS.

Fatty acid synthesis

Plant FAS is a type II FAS enzyme, which exist as a multiprotein complex involved in *de novo* synthesis of fatty acids (Nobusawa*, et al.*, 2013; Sayanova*, et al.*, 2012). The overall goal in fatty acid biosynthesis is to attach a two-carbon acetate unit from malonyl-ACP to, initially acetyl-CoA, then to the acyl growing chain. Four enzyme reactions are needed for the successive addition of the two-carbon units and these are catalysed by a condensing enzyme (ß-ketoacyl-acyl carrier protein (ACP) synthase, (KASIII, EC 2.3.1.180, FABH; KASII, EC 2.3.1.179, FABF)) first reductase (ß-ketoacyl-ACP reductase, FABG, EC 1.1.1.100), dehydrase (ß-hydroxyacyl-ACP dehydrase, FABZ, EC 4.2.1.59) and a second reductase (enoyl-ACP reductase, FABI, EC 1.3.1.9) (Nobusawa*, et al.*, 2013; Sayanova*, et al.*, 2012). Three different condensing enzymes are found in plants, called KASI, II, and III. The initial condensation of acetyl-CoA and malonyl-CoA is catalysed by KAS III, while the subsequent condensation reactions, which elongate 4- to 14- carbon chains is catalysed by KAS I and finally KAS II to produce a fatty acid stearate (Nobusawa*, et al.*, 2013; Sayanova*, et al.*, 2012; Walley*, et al.*, 2013). Genes encoding FAS complex enzymes involved in condensation (ß-ketoacyl-ACP synthase II and III (FabF and FabH)), ß-keto reduction (ß-ketoacyl-ACP reductase (FabG, EC 1.1.1.100), dehydration (hydroxylacyl-ACP dehydratase/isomerase (FabZ, EC 4.2.1.59) and enoyl reduction (trans-2-enoyl-ACP reductase I (FabI, EC 1.3.19) were enriched in the peel tissue and increased in expression with fruit development (Figure 5.3, 5.4A). The

progressive increase in the expression of the FAS enzymes suggests an increase in fatty acid synthesis with fruit development.

Release of fatty acid biosynthesis products from the FAS complex

De novo fatty acid synthesis can be halted by direct use of the 16- to 18-carbon fatty acid products to form phosphatidate through the Kornberg-Pricer pathway. The fatty acid products may also be released from the FAS enzyme complex by hydrolysis, a process catalysed by thioestarases. The free un-esterified fatty acids may be exported outside the plastids to undergo modifications on the endoplasmic reticulum (ER) or used for complex lipid biosynthesis in the extra-plastidic compartment (Nobusawa*, et al.*, 2013; Sayanova*, et al.*, 2012; Walley*, et al.*, 2013). Fatty acly-ACP thioesterases (FATA, FATB) are constitutively expressed and are not limiting, their activities reflect the combined needs of glycolipids, wax and cutin biosynthesis (Curry, 2008; Curry, 2009; Rhee*, et al.*, 1998).

FATA preferentially hydrolysis oleoyl-ACP whereas FATB has highest activity with saturated 14C- to 18C acyl-ACPs. Both FATA and FATB acyl-ACP thioesterases were differentially regulated to provide and increase the pool of saturated fatty acids. The increase in the genes encoding the acyl-ACP thioesterases (FATA, EC 3.1.2.14 and FATB, EC 3.1.2.-) with fruit maturity suggested an increase in the accumulation of un-esterified fatty acid synthesis products in the plastids (Figure 5.3, 5.4A).

In chloroplast (4<n<16)

[acyl-carrier-protein] S-malonyltransferase
acetyl-CoA carboxylase (acc)

malonyl-ACP — EC 2.3.1.39 — malonyl-CoA — EC 6.4.1.2 — acetyl-CoA
FABD

acetyl-CoA

acyl-ACP (4<n<16)

2.3.1.180 KASIII (FABH)
2.3.1.179 KASII (FABF)

In endoplasmic reticulum (16<n)

Malonyl-CoA

A long-chain acyl-CoA (16<n)

CUT1
6.2.1.3 2.3.1.- KCS
LACS2

3-Oxo-acyl-ACP
1.1.1.100 FABG

A long-chain 3-oxoacyl-CoA
KAR 1.1.1.211

3-Hydroxy-acyl-ACP
4.2.1.59 FABZ

A long-chain 3-hydroxyacyl-CoA
4.2.1 PHS1

2,3-Dehydroacyl-ACP
1.3.1.9 FABI

A long-chain 2,3-dehydroacyl-CoA
1.3.1.- TER

Acyl-ACP (n+2)
3.1.2.14 FATA

A long-chain acyl-CoA (n+2)
3.1.2.- FATB

Fatty acid (Hexadecanoate)

Long chain fatty acid

C16
Palmitic acid — 16-Hydroxypalmitate — 16-Oxo-palmitate — Hexadecanedioate
CYP86A4 — 1.1.- ACE, HTH — 1.2.-.-
CYP77A6 — 10,16-Dihydroxypalmitate

Phenylpropanoid biosynthesis — Feruloyl-CoA — 2.3.1.88 HHT1 — 16-Feruloyloxypalmitic acid

C18
Oleic acid — 18-Hydroxyoleate — 18-Oxooleate
CYP86A1 — 1.1.- ACE, HTH — 1.2.-.- — Octadec-9-ene-1,18-dioic-acid
1.11.2.3   1.11.2.3

9,10-Epoxy-stearate — CYP94A1 — 3.3.2.- — 9,10,18-Trihydroxystearate
9,10-epoxy-18-Hydroxystearate
3.3.2.- — CYP94A1
9,10-Dihydroxystearate

C22
Docosanoic acid — 22-hydroxydocosanoate — 22-Oxo-docosanoate
CYP86B1 — 1.1.- ACE, HTH — 1.2.-.- — Docosanedioate

Cutin and suberin biosynthesis

DESA1 acyl-[acyl-carrier protein desaturase (DESA1)
DESA1

ω-Hydroxy fatty acid — ω-Oxo fatty acid
CYP86 — 1.1.- ACE, HTH — 1.2.-.- — α,ω-Dicarboxylic acid
1.11.2.3 — 3.3.2.- — ω-Hydroxy epoxy-fatty acid / Polyhydroxy α,ω-dicarboxylic acid

1.14.-.- CYP86A4S — Polyhydroxy-fatty acid

Unsaturated fatty acids

Wax biosynthesis

A long-chain acyl-CoA — A long-chain aldehyde — A long-chain alkane — A long-chain secondary alcohol — A long-chain keton
1.2.1.50 — CER1 — MAH1 — MAH1

3.1.1.50
WSD
FAR — 2.3.1.75 WSD — A long-chain wax ester
A long-chain pimary alcohol — 2.3.1.20 WSD

A long-chain diketone

Figure 5. 3 Localisation of cuticle biosynthesis enzymes enriched in apple fruit peel tissue. The Figure shows a modified combined fatty acid synthesis, fatty acid elongation, cutin and wax biosynthesis, unsaturated fatty acid biosynthesis KEGG pathway reference map. Genes encoding fatty acid metabolism and cutin/wax biosynthesis were enriched in the peel tissue and are indicated in green coloured boxes. Names of tissue-enriched enzymes are written close to the EC number of the enzyme. Pathway enzymes not highlighted in colour were not significantly enriched in apple fruit peel tissue.

Fatty acid elongation – Very long chain fatty acids and unsaturated fatty acids

Fatty acid elongation and desaturation enzymes are membrane-bound protein complexes on the cytosolic surfaces of the ER. Un-esterified fatty acids exported to the cytosol from the plastids bind to the fatty acid elongation or desaturation enzyme complexes (Nobusawa, *et al.*, 2013; Sayanova, *et al.*, 2012; Walley, *et al.*, 2013). Unlike ACP in FAS, CoA acts as the acyl carrier molecule in fatty acid elongation. Elongation initiates in a condensation reaction between acyl– and malonyl-CoA, the latter being also required as a chain-extending unit in the subsequent cycles. The subsequent condensation, keto group reduction, dehydration and, finally, enoyl reduction reactions are similar to the sequence of reactions in FAS (Nobusawa, *et al.*, 2013; Sayanova, *et al.*, 2012; Walley, *et al.*, 2013). Genes encoding fatty acid elongase (LACS, EC 6.2.1.3) enzyme were highly expressed with peak expression at daa60 and a gradual decrease thereafter (Figure 5.3, 5.4A), suggesting peak fatty acid elongation to be at daa60.

Fatty acids are desaturated by introducing of double bonds into the fatty acid chains a process catalysed by desaturases (Sayanova, *et al.*, 2012). Starting from C16 and C18, elongation specific desaturases generate cis-monounsaturated palmitoleic acid ($\Delta 9$, C16:1) and oleic acid ($\Delta 9$, C18:1). Both palmitoleic and oleic acid are then used for the synthesis of a variety of other long-chain unsaturated fatty acids. In plants, fatty acid elongation precursors, linoleate and linolenate, are synthesized *de novo,* unlike in humans were they must be provided in diet (Tapiero, *et al.*, 2002). These essential fatty acids act as precursors of arachidonic acid, which in turn, is required for the synthesis of eicosanoids, including prostaglandins, leukotrienes and thromboxanes. Genes encoding

an acyl [acyl-carrier protein] desaturase (DESA1) gradually increased in expression throughout fruit development with a transitory peak in expression at daa60 (Figure 5.3, 5.4A). An increase in the expression of DESA1 throughout fruit development suggests a progressive increase in the accumulation of unsaturated fatty acids. The progressive increase in ACC, FAS, thioesterases, elongases, desaturases suggests the increase in the pool of saturated and unsaturated fatty acids precursors for cuticle/wax biosynthesis.

In as much as the synthesized fatty acids are invaluable to the plant, their accumulation is checked to avoid unnecessary expenditure of energy and carbon resources. Thus, depending on the developmental stage, and environmental conditions the plant may initiate fatty acid degradation. The degraded fatty acids may be used as carbon sources or to generate energy through ß-oxidation in the mitochondria. In this study fatty acid metabolism enzymes, alcohol dehydrogenase (ADH, 1.1.1.1), acyl-CoA oxidase (ACX), acetyl-CoA acyltransferase (ACAA, 2.3.1.16), enoyl-CoA hydratase (MFP2, 1.1.1.211) and aldehyde dehydrogenase 7A1 (ALDH7A1, EC 1.2.1.3), were encoded for by genes, which increased in expression with increase in fruit maturity (Figure 5.3, 5.4A). The increase in fatty acid metabolism enzymes with fruit maturity suggests increased control of fatty acid reserves as well as possible increase in the demand for energy or carbon skeletons. The increase in the demand for energy in plausible, considering the rate of energy consumption by the fruit increases with fruit maturity, more so during the climacteric phase. The degraded fatty acids also result in aldehydes, alcohols, and ketones, which contribute to fruit flavour.

Fatty acids are destined for a number of processes besides cutin, glycolipids, and waxes. They may act as reservoirs of carbon skeletons for steroid, terpenoid biosynthesis, as well as for conversion back into the carbon pool (Kosma, *et al.*, 2010).

Wax formation

The wax forming genes *wax-ester synthase (WSD1, EC 2.3.1.75), fatty acid omega-hydroxylase (CYP86A4S, EC 1.14.-.-),* and *omega-hydroxypalmitate O-feruloyl transferase (HHT1, EC 23.1.1.88)* were encoded by transcripts, which increased in expression with fruit development (Figure 5.3, 5.4B). The increasing levels of expression of transcripts encoding cuticle synthase 1 (CUT1, KCS, EC 2.3.1.-) with fruit development (Figure 5.3, 5.4B), suggests the accumulation and hence thickening of cuticle waxes as the fruit develops. This physiological development is important considering that the stomatal plasticity decreases with fruit development. As the fruit develops the stomata become suberized loosing their ability to control water and gaseous exchange (Kosma, *et al.*, 2010). Coating the fruit epidermal layer with cuticle might therefore be a mechanism by the plant to surviving desiccation.

In dicots grown in the tropics/subtropics, the young leaves/fruits are exposed to arid environments soon after they emerge from the buds, and therefore a heavier or more complex mixture of wax components may be required to prevent the young tissues from desiccation. Cuticle waxes thus are important components of the apple peel function to shed water; prevent non-stomatal water loss; as the first line of defence against biotic and abiotic stress, such as UV damage; and in plant-insect communication, by attracting or

deterring insects (Rhee, *et al.*, 1998). The differential expression of the wax metabolism genes, suggest that waxes are a complex heterogeneous mixture of saturated/unsaturated very-long-chain (C_{20}-C_{34}) fatty acids and their derivatives, and that wax content and composition, are affected by the developmental process (Curry, 2008; Rhee, *et al.*, 1998).
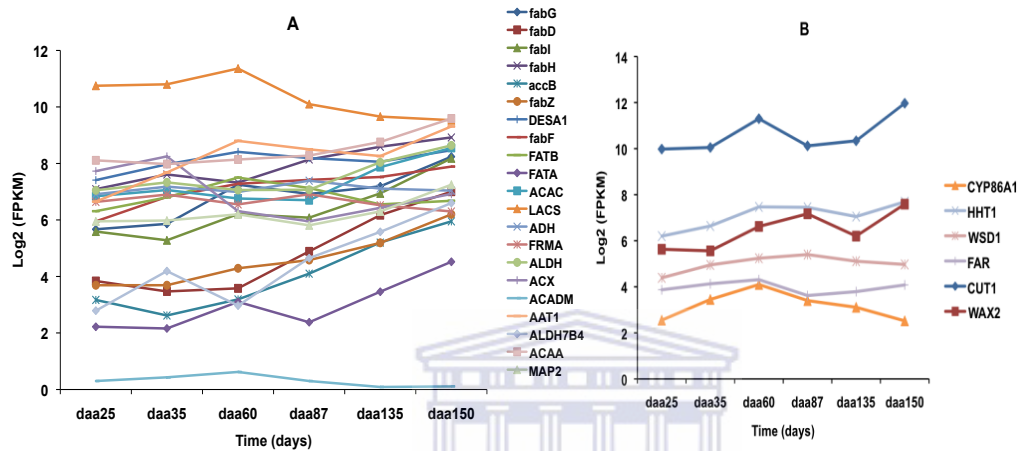
Figure 5. 4 Expression profiles of epicuticular-related genes encoding transcripts. The Figure shows the expression trends of apple peel tissue-expressed genes encoding epicuticular related pathway enzymes across six fruit harvesting time points. A: Expression profiles of genes encoding fatty acid metabolism, fatty acid metabolism, long chain fatty acid elongation and fatty acid degradation pathway enzymes, B: Expression trends of peel expressed genes encoding enzymes in the wax formation pathway. Expression values (FPKM) were log_2 transformed to normalize the data (section 2.2.17).

5.1.3.2 Flavour – terpene and terpenoid metabolism and regulation

The aroma produced by apple fruit at physiological maturity is invaluable to the flavour of the marketed fruit. Flavour has been attributed to the accumulation, retention and or release of volatile substances, such as terpenes and terpenoids. Terpenes are hydrocarbons resulting from the combination of several isoprene units, while terpenoids are modified terpenes, wherein methyl groups have been moved or removed, or oxygen atoms added (Green, *et al.*, 2007; Tsantili, *et al.*, 2007). Fruit flavour volatiles are mainly products of esters, branched chain esters, mevalonate (MVA), 2-C-methyl-D-erythritol 4-

189

phosphate/1-deoxy-D-xylulose 5-phosphate (*MEP*/DOXP) and the phenylpropanoid biosynthetic pathways. Transcripts encoding five MVA and five MEP biosynthetic pathway genes were observed (Figure 5.5).

Flavour compound synthesis through the mevalonate (MVA) pathway

The MVA pathway enzymes are localized in the cytosol on the endoplasmic reticulum (ER), and produces isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). The IPP and DMAP are hemiterpenes (5-carbon) isomers, end-products in MVA and MEP pathways, and are the precursors of isoprene, monoterpenoids (10-carbon), diterpenoids (20-carbon), carotenoids (40-carbon), chlorophylls, and plastoquinone-9 (45-carbon) (Muranaka and Ohyama, 2013). Thus IPP and DMAP serve as the basis for the biosynthesis of molecules used in processes as diverse as terpenoid synthesis, protein prenylation, cell membrane maintenance, hormones, protein anchoring, and N-glycosylation, as well as steroid biosynthesis(Muranaka and Ohyama, 2013; Schaffer, *et al.*, 2007). Transcripts encoding MVA pathway enzymes, acetoacetyl-CoA thiolase (AAT1, EC 2.3.1.9), 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR, EC 1.1.1.34), 3-hydroxy-3-methyl-glutaryl-CoA synthase (HMGS, EC 2.3.3.10), mevalonate kinase (MK, EC 2.7.1.36), and phosphomevalonate kinase (PMK, EC 2.7.4.2), increased in expression with fruit development (Figure 5.5, 5.6A).

Figure 5. 5 Localisation of flavour forming pathway enzymes enriched in peel apple fruit tissue. The Figure shows a modified combined terpenoid, sesquiterpenoid and ester metabolism KEGG pathway reference map. The 'Golden Delicious' fruit peel tissue shows enrichment of triterpenoids, sesquiterpenoids and esters pathway proteins indicated by green coloured boxes. Names of tissue-enriched enzymes are written close to the EC number of the enzyme.

HMGR encoding transcripts had the highest expression among the MVA pathway genes

(Figure 5.6A). HMGR is a rate–limiting enzyme of the MVA pathway (Olofsson, *et al.*,

2011), its steady increase as the fruit developed might have ensured an increasing supply of IPP and DMADP precursors of flavour compounds.
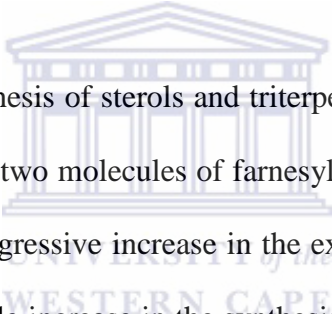
Flavour compound synthesis through the MEP pathway

In contrast to the classical MVA terpenoid biosynthetic pathway, plants have an additional/alternative non-MVA pathway, MEP in plastids, for generating IPP and DMADP pool (Muranaka and Ohyama, 2013). MEP genes were encoded for by genes, which showed and increased in expression throughout fruit development (Figure 5.5, 5.6B). The increasing expression of the transcripts encoding MEP pathway enzymes with fruit development suggests a significant contribution of the MEP pathway to the IPP and DMADP pool. DXS (1-deoxy-D-xylulose-5-phosphate synthase, EC 2.2.1.7) plays a major role in the overall regulation of the MEP pathway. Higher expression of DXS leads to enhanced production of IPP and DMADP, influencing the synthesis of terpenoids in the plastids. It is possible that DXR (1-deoxy-D-xylose-5-phosphate reductoisomerase, EC 1.1.1.267)and HDR may be rate limiting too (Tholl and Lee, 2011).

Sesquiterpenoid biosynthesis

Sesquiterpenes (15-carbon) are polymers of farnesyl-pyrophosphate (FPP) a class of terpenes. Sesquiterpenes are found naturally in plants, as semiochemicals important as defensive agents or as pheromones (Reis, *et al.*, 2009). Farnesyl diphosphate synthase (FDPS, FPS, EC 2.5.1.1) is a housekeeping gene, found in both MVA and MEP pathways, which converts IPP and DMADP to various terpenoids such as sterols and sesquiterpenoids (Reis, *et al.*, 2009). Transcripts encoding FPS increased in expression as

the fruit developed (Figure 5.5, 5.6B), possibly a strategy of copping with the increasing IPP and DMADP products from MVA and MEP. An increase in FPS suggests an increase in the accumulation of sesquiterpenoids and triterpenoids, e.g. alpha-fernesene. Fernesense volatile compounds have been reported to be associated with the apple peel and give the fruit its aroma (Nieuwenhuizen, *et al*., 2013). Transcripts encoding sesquiterpenoid and triterpenoid biosynthesis enzymes, squalene monooxygenase (SMO, EC 1.14.13.132), sesqualene synthase (SQS, EC 2.5.1.21) and ß-amyrin synthase (BAS, EC 5.499.39) increased with fruit development (Figure 5.5, 5.6C), suggesting accumulation of the respective products with fruit development.

The key enzyme in the biosynthesis of sterols and triterpenes is squalene synthase (SQS, EC 2.5.1.21), which condenses two molecules of farnesyl diphosphate (FDP) to squalene (Tholl and Lee, 2011). The progressive increase in the expression of genes encoding the SQS enzymes suggested possible increase in the synthesis of the triterpene squalene with fruit development. This increase in SQS (Figure 5.4C) as well as FPS (Figure 5.4A) expression with fruit development is corroborated by an increase in fruit aroma with maturity of the fruit. However increase in squalene and alpha farnesene volatile compounds may be good for aroma development but under long apple fruit storage conditions apple fruit scald disorders have been shown to increase. The reason for the increase in the scald disorders has been attributed to the accumulation of conjugated trienes, non-volatile alpha-farnesene oxidation products, which are toxic to fruit cells (Leisso*, et al.*, 2013). Several apple scald control measures have been instituted but the use of farnesene and squalene apple fruit dips produced the best results (Whitaker, 2012).

Squalene accumulation in the apple fruit could be invaluable for humans in the synthesis of cholesterol, steroid hormones, and vitamin D (Whitaker, 2012).

High levels of transcripts encoding ß-amyrin synthase (BAS, EC 5.499.39), suggests that BAS could be a major sesquiterpenoid in apple. A number of other genes down stream of IPP and DMAP were encoded for by transcripts, which increased in expression with increase in fruit harvesting time (Figure 5.5, 5.6C). These findings suggested that, though BAS might be the dominant sesquiterpenoid, there are various other volatiles that may contribute to the final fruit flavour.

Ester biosynthesis

Lipooxygenases are a class of iron-containing proteins, which catalyse the dioxygenation of polyunsaturated fatty acids in lipids containing a cis, cis-1,4-pentadiene structures to hydroperoxides (Schaffer, *et al.*, 2007). The major metabolic products from these hydroperoxides are divinyl ethers and epoxy alcohols, obtained via the activities of divinyl ether biosynthesis I (DES) and epoxy alcohol synthase (EAS), respectively. Divinyl ethers and epoxy alcohols are volatile defence and aroma compounds (Schaffer, *et al.*, 2007). The expression levels of transcripts encoding linoleate 9S-lipooxygenase 1 (LOX, EC 1.13.11.58) enzyme increased with fruit maturity, with a slight decrease at daa87 (Figure 5.5, 5.6C).
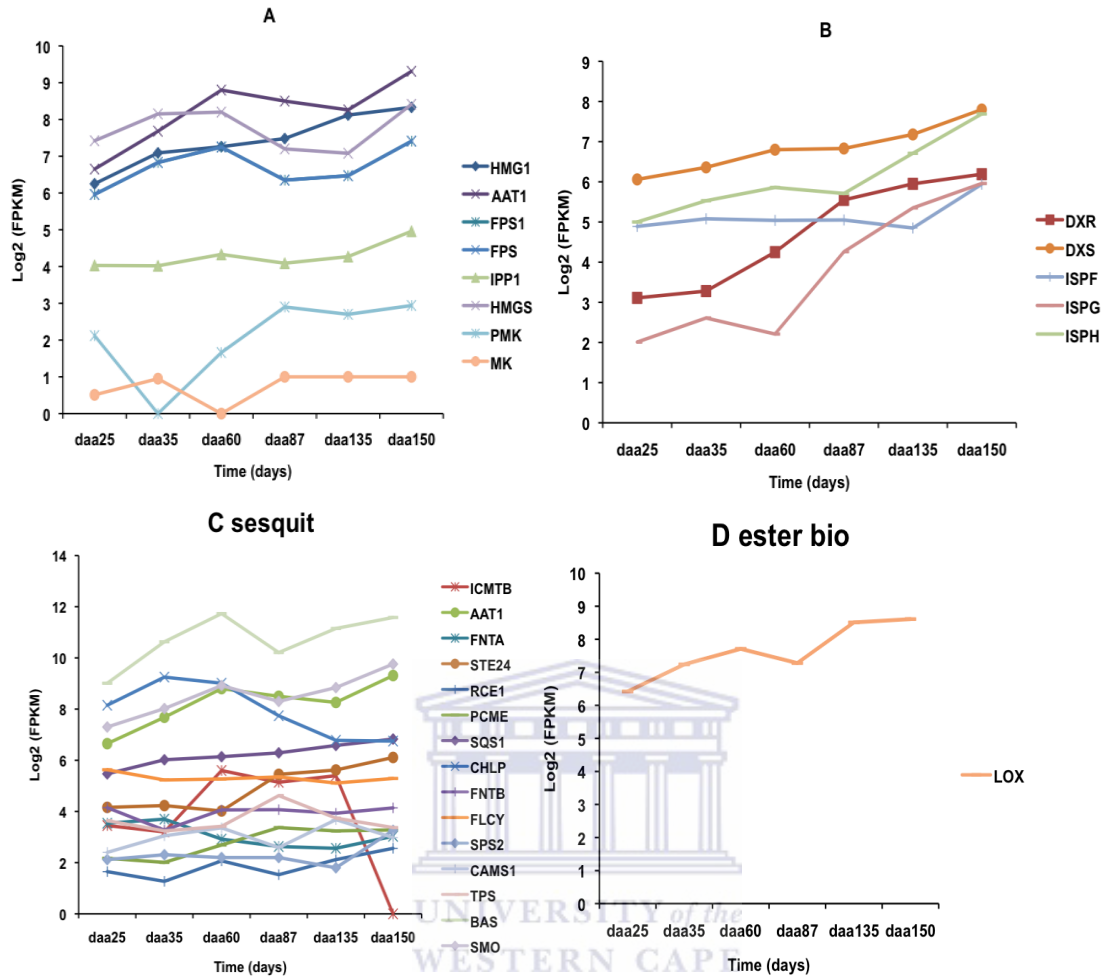
Figure 5. 6 Expression profiles of transcripts encoding MVA and MEP biosynthetic pathways responsible for fruit flavour. A: expression pattern of peel genes encoding MVA pathway enzymes, B: Expression pattern of peel genes encoding MEP pathway enzymes, C: Expression pattern of peel expressed genes encoding sesquiterpenoid, protein post-translation modification, D: Expression trend of peel expressed genes encoding ester biosynthesis pathway enzymes. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

The increasing levels of transcripts encoding LOX, suggests the active involvement of fatty acids in the development of flavour in developing 'Golden Delicious' fruit peel. The synergistic roles of the MVA, MEP and ester biosynthetic pathways in fruit peel flavour development has been demonstrated, supporting the importance and active participation of the apple fruit peel towards fruit aroma (Schaffer, *et al.*, 2007).

5.1.3.3 Fruit pigmentation –Anthocyanin and carotenoid metabolism
and regulation

Fruit colour at the point of physiological maturity and consumption is crucial for marketing. Colour development in apple is as a result of differential gene expression in the flavonoid, carotenoid and phenylpropanoid biosynthetic pathways, which are regulated by internal and external factors (temperature, light and clock) (Steyn*, et al.*, 2009).

During early 'Golden Delicious' fruit development (from flower bud formation to daa35) fruit skin is usually colourful (red) (Figure 5.7). This reddish colour however disappears as the fruit matures, or blushes are left as in daa87 (Figure 5.7). It has been reported that pigment content changes as the fruit develops bringing out the characteristic golden yellow colour of 'Golden Delicious' (Fleancu, 2007). In the same report, Fleancu, (2007) points to the negative correlation between fruit diameter and photosynthetic rate, respiration rate, fruit chlorophyll a and b content and positive correlation between fruit diameter and carotenoid content.
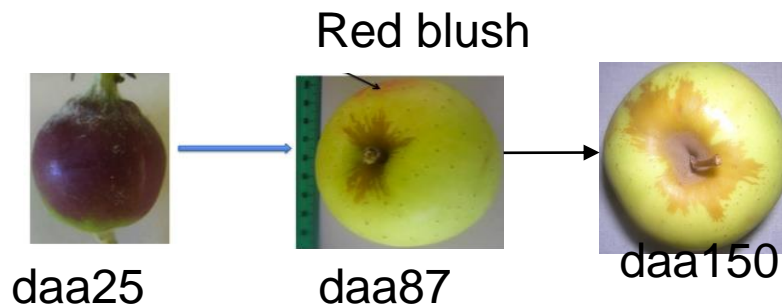
Figure 5. 7 Change in fruit colour during 'Golden delicious' apple fruit development at daa25 and daa87. The Figure shows change in fruit colour with fruit development, from day 25 to day 150 after anthesis. Three stages were selected to depict major changes in colour with fruit development.

However, Fleancu, (2007) gives no report of anthocyanin related pigment accumulation as the fruit grows. Figure 5.7 shows extremes of fruit colour changes in developing 'Golden Delicious' fruit sampled at daa25, daa87 and daa150. Gene expression changes, responsible for the colour differences and change thereof from daa25 to daa150 were elucidated.

Anthocyanidin biosynthesis

Transcripts encoding leucoanthocyanidin oxidase (LDOX, EC 1.14.11.9), dihydroflavonol reductase (DFR, EC 1.1.1.219), chalcone synthase (CHS, EC 2.3.1.74), anthocyanin reductase (ANR, 1.3.1.77), flavonol synthase (FLS, EC 1.14.11.23) and flavonoid 3´-hydroxylase (F3´H, EC 1.14.11.9) were highly expressed early in fruit development, decreasing at daa35 to daa150 (Figure 5.8, 5.9A). LDOX and DRF produce leucoanthocyanidins, precursors for anthocyanins, or condensed proanthocyanidins (condensed tannins) through leucoanthocyanidin reductase (LAR, EC 1.17.1.3), or after LDOX, through ANR (EC 1.3.1.77). Thus, the high levels of LODX and DFR encoding transcripts early in fruit development, reflects the demand for leucoanthocyanidin or proanthocyanidin synthesis. It is important to note that not all of the LODX and DFR encoded transcripts will produce leucoanthocyanin only, for anthocyanin accumulation. CHS encoding genes were the highest expressed across all time points followed by LDOX and F3´H encoded genes respectively (Figure 5.9A). The profile shown by these transcripts suggests that during daa25 the rate of proanthocyanidins formation is still increasing up to daa35, after which the synthesis of condensed tannins declines. LAR and ANR enzymes are involved in condensed tannin synthesis early in fruit development for

microbial and UV light protection, as the fruit builds a protective waxy exocarp. The fate of leucoanthocyanidins towards anthocyanin formation is determined by the levels of F3′H encoding transcripts, which showed a similar pattern of expression to LDOX, DRF, CHS and to some extend ANR across all the six time points (Figure 5.9A). In this study LAR encoding transcripts were not significantly expressed in peel and across the time points, as such were excluded from analysis, suggesting proanthocyanidin formation through catechin is not a differentially regulated process in apple peel.

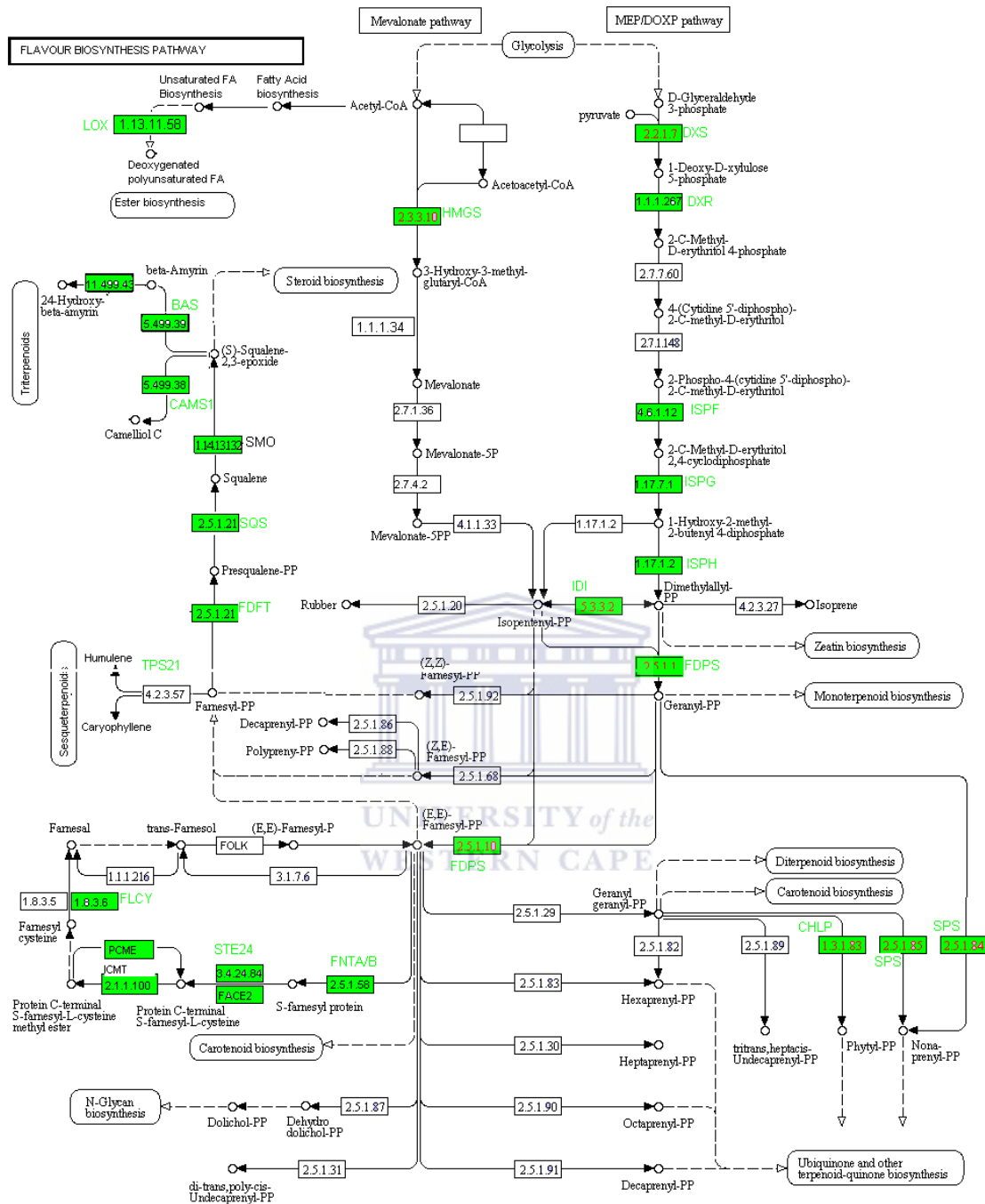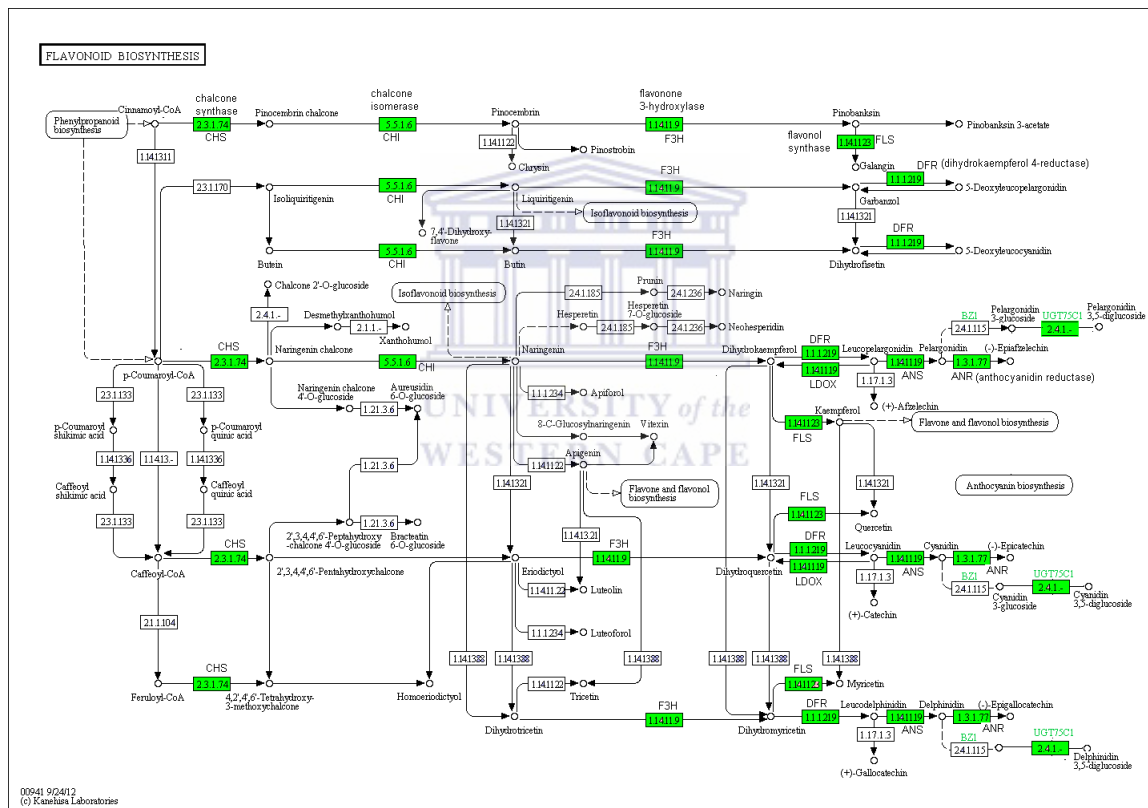**Figure 5. 8 Localisation of flavonoid and anthocyanin pathway enzymes enriched in peel apple fruit tissue.** The Figure shows a modified combined flavonoid and anthocyanin biosynthesis KEGG pathway reference map. The Figure shows the apple fruit peel tissue to be enriched in flavonoid and anthocyanin pathway enzymes. Boxes coloured green represent sub pathway genes enriched in 'Golden Delicious' fruit peel tissue. Names of tissue-enriched enzymes are written close to the EC number of the enzyme. Pathway enzymes not highlighted in colour were not significantly expressed in the peel tissue.

It is also possible that the LAR route for condensed tannin synthesis is not preferred to the ANR route in apple, as in *A. thaliana* were LAR is not expressed at all (Bogs*, et al.*, 2007).

Anthocyanin biosynthesis

Glycosylation of the anthocyanins forms an array of colourful anthocyanins, which colour the fruit peel early in fruit development. Of the four glucosyltransferases, UGT73B2, UGT75C1, UGT78D2 and CYP706A5, UGT78D2 encoding genes showed an expression pattern in agreement with fruit peel colour transitions (Figure 5.9A). UGT78D2 could be the preferred glucosyltransferase in apple peel, responsible for the reddish colourations early in fruit development. Differential transcription of anthocyanidin biosynthesis genes in distinct branches of the pathway is likely to play a major role in deciding the branch of the pathway to be used, which would then define fruit peel colour (Koes*, et al.*, 2005). It is also possible that some of these anthocyanidin biosynthesis genes might be subjected to post transcriptional control, further modifying fruit peel colour (Koes*, et al.*, 2005).

Vacuole transportation of flavonoids and signalling

Transparent testa (TT) protein 12 (TT12) encoding transcripts had high expression early in fruit development, but decreased significantly at daa35 only to increase after daa87 to daa150 (Figure 5.9B). Transparent testa (TT) proteins are a group of MATE efflux proteins involved in vacuolar transportation of flavonoids (Chen*, et al.*, 2012). High levels of TT12 early in fruit development correlate with levels of UGT78D2, which

199

produces anthocyanins. The increase in expression of TT12 after daa87 may be involved in some physiological function other than anthocyanin transportation. Genes encoding WRKY, a superfamily of proteins with various physiological functions in plants, were observed (Rushton*, et al.*, 2011). WRKY proteins are integral part of signal webs controlling various physiological processes in the plant. Genes encoding WRKY, were highly expressed and increased in expression from daa25 to daa35 followed by a sharp declining pattern similar to anthocyanin forming enzymes (Figure 5.9B). This trend similarity between WRKY and anthocyanin forming genes suggests their close relationship, possibly WRKY activates the expression of anthocyanin genes. The actual mechanism of their relationship is however not clear.

Anthocyanin and anthocyanidin regulation

MYB, basic helix-loop-helix (bHLH) and WD40 transcription factor proteins have been shown to be common denominators in the regulation of structural anthocyanin and proanthocyanidin genes (Allan*, et al.*, 2008). These proteins interact and activate transcription of structural genes directly. Abundances for genes encoding bHLH and MYB showed no clear uniform expression pattern across the time points (Figure 5.9 C). MYB44 was highly expressed early in fruit development, but sharply decreased from daa25 to daa60 followed by a step-wise increase to daa150. MYB30 on the other hand, increased in expression peaking up at daa35 followed by a gradual decrease for the rest of the fruit development period (Figure 5.9 C). The increase in MYB30 encoding transcripts correlate to the expression pattern of anthocyanin forming proteins, suggesting possible regulation of these structural genes by MYB30. bHLH encoding genes had low

expression throughout fruit development, while bHLH137 encoding transcripts increased with harvesting time, and WD40 peaked up at daa60 followed by a slight increase to daa150 (Figure 5.9 C).
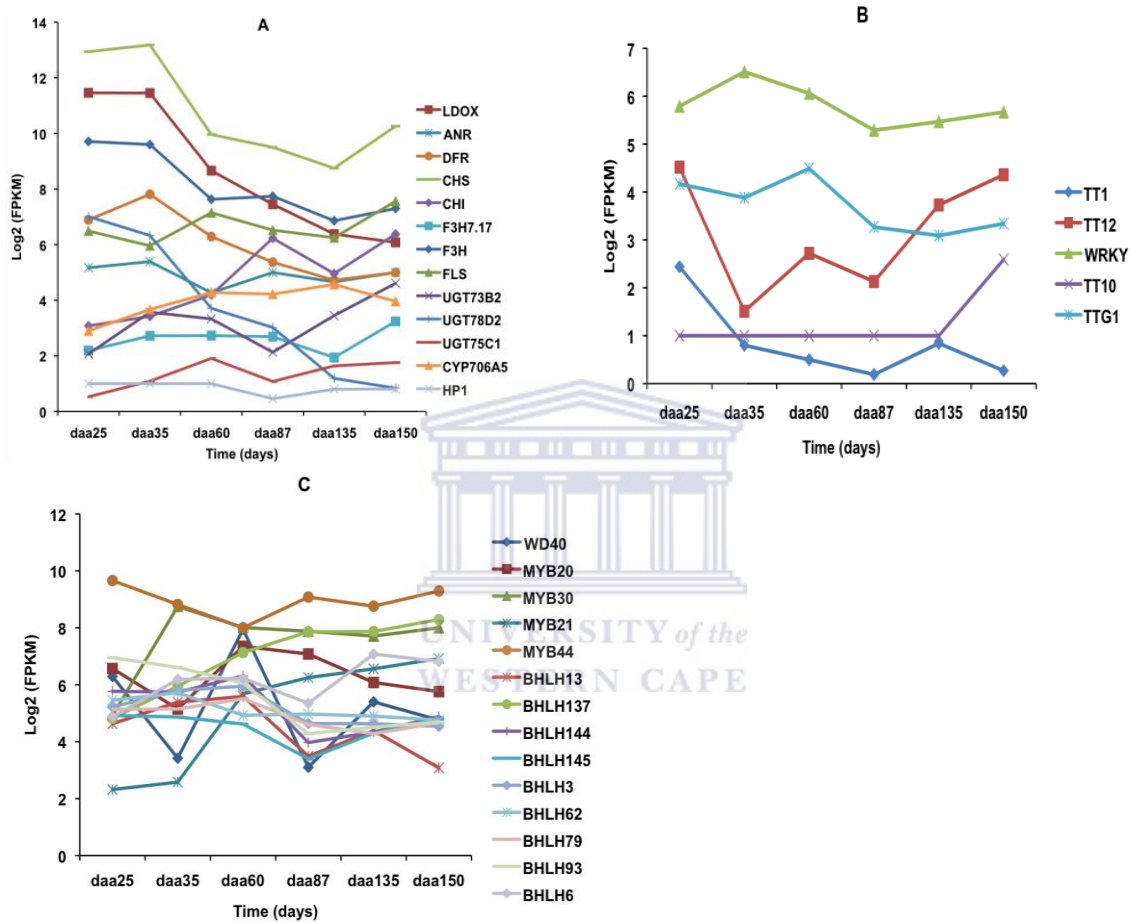
Figure 5. 9 Expression profiles of peel expressed genes encoding phenylpropanoid metabolic pathway proteins. A: Expression profiles of anthocyanidin and anthocyanin biosynthesis enzymes, B: Expression trends of peel expressed genes encoding proteins involved in vacuole transportation and signalling of flavonoids, C: Expression patterns of peel expressed genes encoding transcription factors involved in the regulation of flavonoid biosynthesis. The levels of gene expressions are estimated as fragments per kilo base of exon per million reads mapped (FPKM). Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

Considering MYB, bHLH and WD40 interact in a 1:1:1 ratio (Allan, *et al.*, 2008), their expression profile was expected to be similar. Anomalies are indeed expected considering

that bHLHs and MYBs of different clades may show regulatory redundancies, and /or functional specificity (Allan*, et al.*, 2008). Even though some of the MYB, bHLH, and WD40 encoding genes levels were high, not all of these members will control pigmentation in fruit. Differential expression and regulation by these factors is also expected in different fruit tissues and phases of development.

Flavonoid degradation

Little is known about the stability and catabolism of anthocyanins in plants. In some plants, anthocyanins accumulate transiently, appearing and disappearing during plant development or with changes in environmental conditions (Oren-Shamir, 2009). The accumulation of anthocyanins in young apple fruit peel (Figure 5.8, 5.9A), might act in protecting the tender tissues in particular their photosynthetic apparatus, from damaging UV light as well as providing antioxidant activity, protecting the cells from oxidative damage.

With the accumulation of protective waxes that reflect sunlight, thereby providing photo-protective function, 'Golden Delicious' fruit peel changes colour from reddish to green. Loss of red pigmentation could be due to increased chlorophyll accumulation during fruit expansion and growth, termination of anthocyanin biosynthesis and dilution by growth and/or increased activity induced anthocyanin degradation when they are no longer required as photo-protectants (Oren-Shamir, 2009).

Anthocyanin degradation is apparent in greening of *Chrysobalanus liaco* (*Cocoplum*) and *Photiniax frasei* c.v 'Red Robin' leaves as they mature (Oren-Shamir, 2009). In rose, foliage colour change from young red to mature green leaves is due to increased chlorophyll synthesis, termination of anthocyanin synthesis and dilution of the anthocyanins in the expanding leaves (Oren-Shamir, 2009). In *Capsicum spp.*, the anthocyanins in the vacuoles are degraded while chlorophyll is degraded in the plastids and the synthesis of carotenoids result in the red, orange and yellow colours of the mature pepper fruit (Oren-Shamir, 2009).

Polyphenol oxidase (PPO, EC 1.14.18.1), peroxidase (POD, EC 1.11.1.7) as well as ß-glucosidase (EC 3.2.1.21) are thought to be responsible for the anthocyanin degradation in fruit vacuoles or upon membrane degradation (Grimplet*, et al.*, 2007; Oren-Shamir, 2009). The expression pattern of ß-glucosidase encoding transcripts was highest early in fruit development, but decreased with increase in harvesting time (Figure 5.10). PPO encoding transcripts on the other hand had low expression early in fruit development, but increased as the fruit matures (Figure 5.10). Here it is evident that the inter-play between ß-glucosidase and PPO proteins brings about the colour transitions observed during 'Golden Delicious' fruit development.
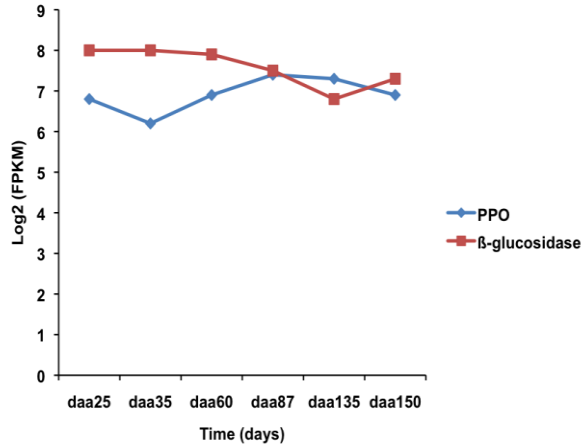
Figure 5. 10 Expression profiles of peel expressed genes encoding anthocyanin degradation pathway enzymes during 'Golden Delicious' fruit development. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

Carotenoid metabolism

The transition from green to golden colour atypical of ripening 'Golden Delicious' fruit is due to the developmental transition of chloroplasts to chromoplasts. As photosynthetic membranes degrade, chlorophyll is degraded and carotenoids including ß-carotene and lycopene, accumulate in response to ripening related and ethylene-inducible gene expression (Giovannoni, 2001).
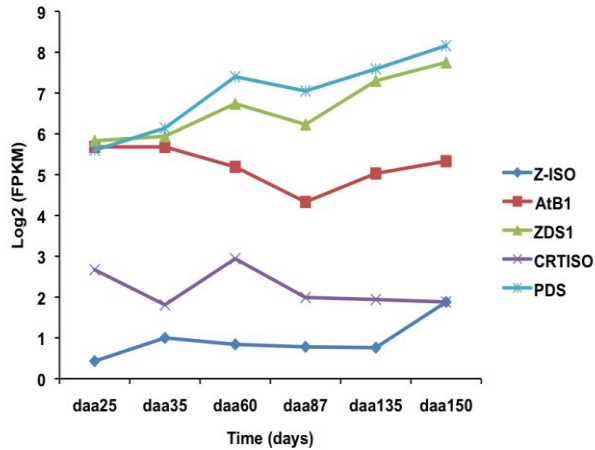
Figure 5. 11 Expression profiles of peel expressed genes encoding carotenoid pathway enzymes during 'Golden Delicious' fruit development. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

Transcripts encoding Z-ISO and ZDS1 increased in expression as the fruit matures, a trend correlating to the increase in the golden colouration (Figure 5.11). The expression of genes encoding AtB1, CRTISO and PDS enzymes showed low expression and decreased in expression with fruit development (Figure 5.11). The levels of CRTISO, AtB1 and PDS are too low to be responsible for the change in colour as the fruit matures. As such, it is possible that Z-ISO and ZDS1 are responsible for the relative levels of golden colouration in maturing 'Golden Delicious'. In tomato, however, lycopene-ε-cyclase is responsible for the relative levels of ß-carotene and lycopene in tomato fruit (Giovannoni, 2001).

5.1.3.4 Cell wall metabolism – Cellulose and pectin metabolism and regulation

Ripening-related softening of the fruit is generally associated with the dissociation of middle lamella and primary cell walls (Atkinson, *et al.*, 2012; Harb, *et al.*, 2012; Wei, *et al.*, 2010). Plant cell walls are composed of rigid cellulose microfibrils held in concert by

networks of matrix lycans (hemicellulose) and pectin, with varying levels of structural proteins and phenolics, which when solubilised, depolymerised and rearranged affect cell wall strength leading to softening (Ortiz*, et al.*, 2011a, 2011b; Wei*, et al.*, 2010).

Pectin metabolism

Pectin consists of a complex set of polysaccharides and is a major component of the middle lamella, where it helps to bind cells together, but is also found in primary cell walls where it allows primary cell wall extension and fruit growth (Gou*, et al.*, 2007).

Transcripts encoding pectin metabolism were highly expressed early in fruit development when the rate of pectin formation and degradation was high (Figure 5.12A). During early fruit development cell walls are rapidly formed and destroyed to allow for cell division and expansion. Once full cell size and count are reached maximum pectin incorporation into cell walls is also reached, and pectin-degrading enzymes are no longer required for cell softening to allow cell expansion and division, and their concentration decreases. This observation was corroborated by the progressive decrease in the expression of genes encoding pectin degradation. The pectin degrading enzyme concentration however, will be expected to increase in response to ethylene induction as the fruit starts to ripen or upon pathogen infection.

Galacturonosyltransferase (GAUT, EC 2.4.1.-) is an enzyme involved in pectin biosynthesis. GAUT catalysis the transfer of galacturonic acid from uridine 5´-diphosphogalacturic acid onto the growing pectic polysaccharide homogalacturonan chain (Harb*, et al.*, 2012). Unlike the other pectin metabolism enzymes (Figure 5.12A),

which decreased in expression with fruit development, transcripts encoding GAUT had peak expression between daa35 and daa87, suggesting peak pectin biosynthesis to be between daa35 and daa87 (Figure 5.12C). These findings are in agreement with the known decrease in pectin concentration as apple fruit matures (Zude, *et al.*, 2006) and shows that the highest pectin content in 'Golden Delicious' fruit peel was at daa35. The progressive decrease in genes encoding pectin-degrading enzymes is desirable in keeping the cell paste intact hence keeping the firmness of the apple fruit for longer.

Cell wall loosening

Expansins encoding transcripts can be grouped into two, high early expression and low but increasing expression with increase in fruit maturity (Trujilio, *et al*., 2012; Zhang, *et al*., 2014). Expansins, A4, A6 and B3 had high expression early in fruit development, but decreased with fruit maturity (Figure 5.12B). Transcripts encoding expansins A10, and A8 increased in expression peaking at daa60 and daa87 respectively (Figure 5.12B). The first group of expansins could possibly be responsible for the early cell wall softening allowing for cell division and expansion between daa25 and daa35. The second group might be responsible for the final cell wall softening and cell division and expansion, characteristic of daa60 to daa135 (Figure 4.1).

Cell well metabolism

The oriented deposition of load-bearing cellulose microfibrils in the cell wall determines the growth and organ formation in plants. Cellulose is synthesized by plasma membrane–bound complexes containing cellulose synthase (CESA, EC 2.4.1.12) proteins (Crowell,

et al., 2009). Transcripts encoding CESAs had peak expression between daa35 and daa60, followed by gradual decrease throughout fruit development (Figure 5.12D). Peak expression of these CESAs at daa35 and daa60 suggests that most of the cell wall formation was done in the early phases of fruit development when new cell walls were required for new cells. Once cells stopped dividing, after daa87, what was left was just repairing of cell walls, hence the decrease in the genes encoding these enzymes. It might be possible that the potential cell number in the fruit might be ultimately set between daa35 and daa60. If so this period might be agronomical invaluable in setting the potential fruit size, and hence fruit yield.
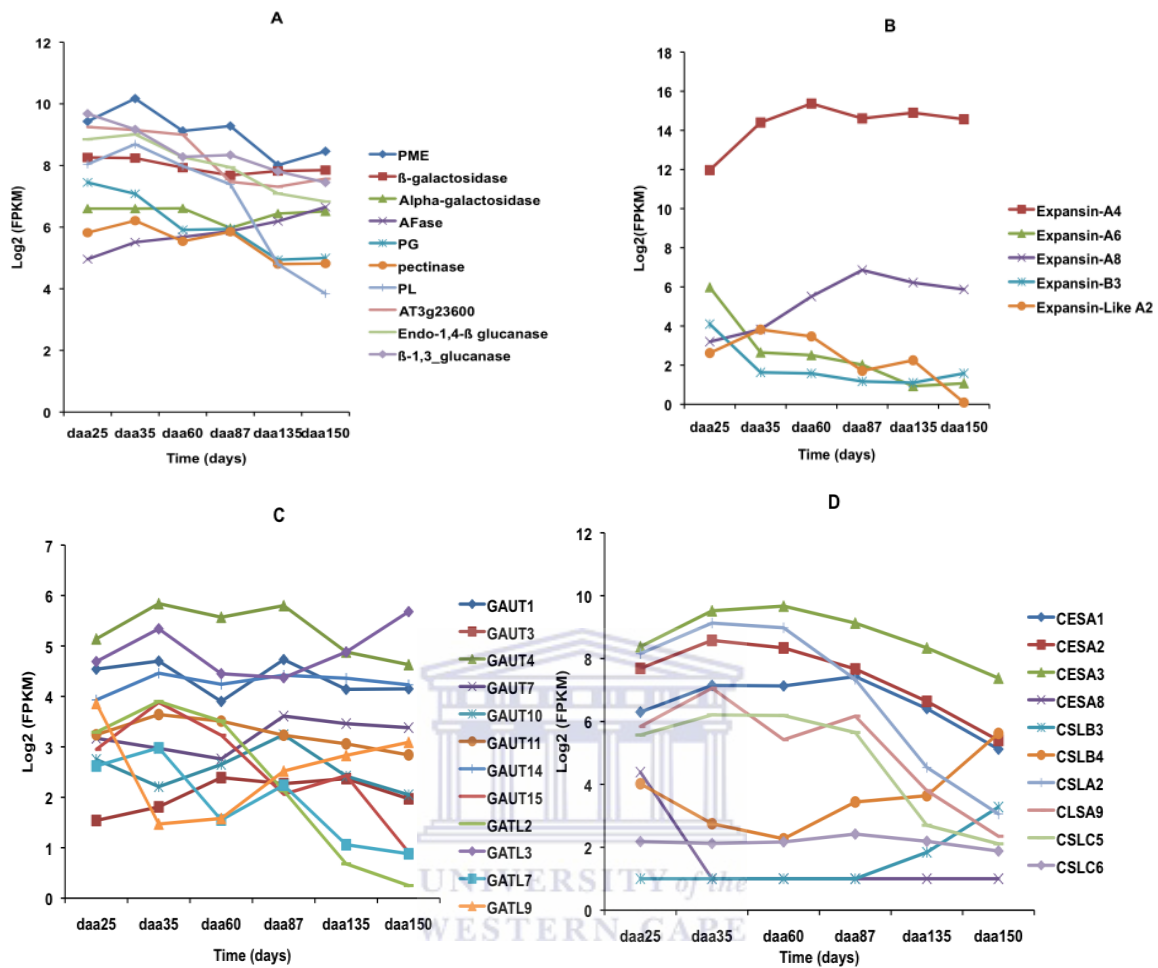
Figure 5. 12 Expression profiles of peel-enriched genes encoding cell wall related enzymes. A: Expression trends of peel expressed genes encoding pectin metabolism pathway enzymes, pectin methyltransferase (PME), ß-galactosidase (ß-GS), Alpha-galactosidase (œ-GS), arabionofructosidase (AFase), polyglucoronidase (PG), pectate lyase (PL), endo-1,4-ß-glucanase (EC 3.2.1.4), and a ß-1,3-glucanse (EC 3.2.1.39); B: Expression trends of peel expressed genes encoding Cell wall loosening enzymes, C: Expression trends of peel expressed genes encoding pectin biosynthesis pathway enzymes, Galacturonosyltransferase (GAUT), D: Expression trends of peel expressed genes encoding cell wall metabolism pathway enzymes, cellulose synthase-like (CSLC, CLSA), cellulose synthase (CESA), during 'Golden Delicious' fruit development from daa25 to daa150.. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

Low levels of cell wall loosening and degrading enzymes are desirable, since high levels of these enzymes are inversely related to fruit firmness. In this study the levels of these enzymes decreased with fruit maturity, and were lowest at daa150 (Figure 5.12A),

suggesting the fruits were still firm and marketable at daa150 with possibly long shelf life.

5.2 Conclusion

The peel is a crucial component of the apple fruit, invaluable for fruit and plant's survival against the internal and external environment. This chapter aimed at providing an overview of the genes encoding cuticle, flavour, pigmentation and cell wall pathway enzymes and the changes thereof with 'Golden Delicious' fruit peel tissue development. Functional classification of the transcripts significantly expressed in apple fruit peel using the MIPS categories revealed that cell fate; interaction with environment; protein with binding function; development and metabolism were the predominant functional categories in apple peel. As the fruit developed 59% of the peel significantly expressed transcripts increased or decreased in expression. Transcripts that showed an increase in expression with fruit development were mostly enriched for protein processing in endoplasmic reticulum, plant pathogen interaction, carbon fixation, RNA transport and oxidative phosphorylation KEGG pathway enzymes. Fatty acid, very-long fatty acid and cuticle biosynthesis encoding genes increased in expression with fruit development, suggesting an increase in epicuticular wax thickness with fruit development.

Fruit flavour changes with fruit development and the mevalonate (MVA), 2-C-methyl-D-erythritol4-phosphate (MEP) and ester biosynthesis pathways contribute to fruit flavour. Genes encoding MVA, MEP and ester metabolism pathway enzymes increased in expression with time of harvesting, suggesting an increase in aroma compounds with fruit

maturity. The terpenoids produced by the MVA and MEP pathways may be further processed into sesquiterpenoids or triterpenoids. The high expression levels of ß-amyrin synthase (BAS) suggest that a triterpenoid ß-amyrin might be the predominant accounting for most of the fruit's flavour.

Colour development in 'Golden Delicious' has been shown to change with fruit maturity from reddish to golden yellow at physiological maturity. These fruit colour transitions have been corroborated by high expression of anthocyanin biosynthesis pathway genes during early fruit development. The colour transitions have also been attributed to anthocyanin degradation by phenol peroxidase (PPO) enzymes whose expression had a similar pattern to anthocyanin genes. Genes encoding carotenoid biosynthesis enzymes increased with fruit maturity supporting the observed change to golden colour with fruit maturity.
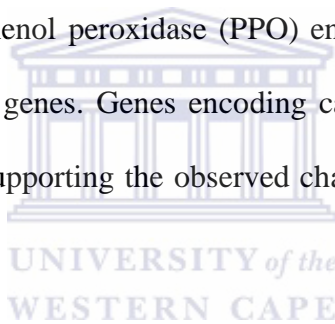
The integrity of the cell wall is compromised by cell loosening enzymes. However, cell wall loosen is a prerequisite for cell division and expansion hence fruit development. So, depending on when, during fruit development, cell wall loosen may be a desirable process. Genes encoding cell wall loosening enzymes were shown to decrease in expression with maturity. Two expansin proteins (A8 and A10) however had peak expression at daa87 and daa60 respectively, suggesting they could be responsible for the final push in cell expansion between daa60 and daa87 when fruit cells attain their final sizes. Cellulases encoding genes had peak expression at daa35 to daa60, suggesting cell

wall formation for new cells stops at daa60 after which less cellulose might be needed for cell repair.

We thus have elucidated the most significant pathways and gene changes in the definition of peel colour, flavour, cuticle formation and cell wall integrity. The observations in this study provide an invaluable catalogue of important determinants of peel properties and their changes during fruit development. The catalogue will allow targeted pathway and gene studies towards apple fruit improvement. The major limitation in comprehensive annotation of the reconstructed transcripts remains the depth of the publicly available non-redundant databases for interrogating.

6   CHAPTER 6

6.0 *In silico* mRNA transcriptional profile comparisons of *M. x domestica f*ruit peel and pulp tissues

Abstract:

The growth and development of 'Golden Delicious' fruit occurs over a period of about 135 to 150 days after anthesis (daa) to fully ripeness. During this period morphological and physiological changes occur, which define quality attributes. These changes are a result of spatial and temporal differential gene expressions in 'Golden Delicious' fruit. Peel and pulp tissue-up-regulated genes were spatially and temporally enriched for pathways defining tissue specific properties. Genes encoding proteins involved in photosynthesis, carbon fixation, fatty acid metabolism, terpenoid, carotenoid and flavonoid pathways were mostly enriched in the peel tissue. On the other hand, genes encoding proteins involved in oxidative phosphorylation, plant hormone signal transduction and C_4-carbon fixation pathways were enriched in the pulp tissue. This study has thus provided an in depth catalogue and elucidation of the pathways enriched in peel and pulp 'Golden Delicious' tissues, which might function in defining the properties of the tissues, and their contribution thereof to total fruit development. An in depth understanding of the molecular mechanisms of fruit development and fruit tissue contribution to development allows for refined and focused cultivar breeding and manipulation approaches.

6.1 Background

Differences in fruit tissue properties is well documented for antioxidant properties of peel and pulp tissues in pear (Muchuweti and Chikwambi, 2008) and apple (Manzoor, *et al.*, 2012) among other plants. Variations in colour, wax and starch accumulation as well as

the build up in aroma compounds has also been noted in apple peel and pulp tissues (Geigenberger, 2011; Kosma*, et al.*, 2010; Kühn and Grof, 2010; Kunst and Samuels, 2009; Lata and Tomala, 2007; Lister*, et al.*, 2006). These phenotypes are manifests of temporal and spatial gene expression as elicited by the changes in the fruit growing conditions. In studies elucidating molecular mechanisms responsible for the variations in phenotypes between tissues and across developmental phases, a comparison of the different tissues provide better insight into tissue specific and temporal gene expression. Such a study better elucidates physiological processes responsible for defining the properties of the tissue and possible function as part of the whole fruit.

This study comparatively identified greater than two-fold spatially and temporally up regulated transcripts in peel and pulp 'Golden Delicious' fruit tissues. These transcripts were enriched for enzyme pathways defining peel and pulp tissue properties.

6.2 Results and discussion

6.2.1   Apple fruit peel and pulp pathway enrichment analysis

A KEGG pathway enrichment analysis of the transcripts up-regulated in peel, by at least two-fold change, revealed that the most enriched pathways were plant pathogen interaction, terpenoid metabolism, photosynthesis, carbon fixation, fatty acid metabolism, plant hormone signal transduction, carotenoid and flavonoid pathways (Figure 6.1). In the fruit pulp tissue, transcripts with a two-fold up-regulation were mostly enriched for oxidative phosphorylation; amino sugar nucleotide sugar metabolism; protein processing in endoplasmic reticulum; plant hormone signal transduction; plant pathogen interaction and carbon fixation pathway enzymes (Figure 6.1). Though expressed in pulp the plant

214

pathogen interaction, terpenoid, photosynthesis, carbon fixation, fatty acid metabolism, carotenoid and flavonoid metabolic pathways were more than twice enriched in the fruit peel. On the other hand protein processing in endoplasmic reticulum, oxidative phosphorylation, and plant hormone signal transduction were more enriched in fruit pulp than in fruit peel (Figure 6.1).
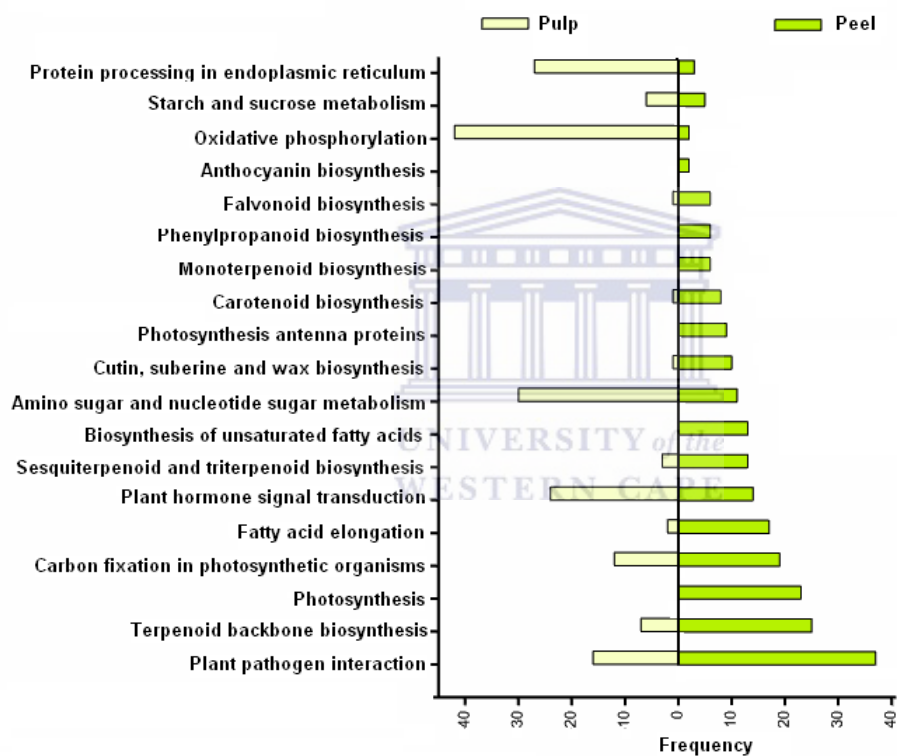
Figure 6. 1 Peel vs pulp tissue KEGG pathway enrichment. The top twenty most enriched pathways in either the peel or the pulp tissues of 'Golden Delicious' fruit. Transcripts assembled from the peel and pulp tissues annotated with KEGG Orthology terms (KO) and their respective pathways identified by submitting the KO terms to KEGG database. The count of the KO terms in a pathway indicated the level of enrichment of the pathway. Only transcripts up regulated or down regulated by more than two-fold in the peel or pulp tissues were used in the comparison. The y-axis =KEGG pathway description of the accession, Frequency=number of unique genes enriched in a pathway.

6.2.2   Photosynthesis

The capturing of light energy to drive the synthesis of organic compounds is fundamental to the survival of plants (Foyer and Shigeoka, 2011). These physico-chemical processes takes place in the chloroplast and depend on a set of complex protein molecules that are located in and around a thylakoid membrane (Pinheiro and Chaves, 2011). Through a series of energy transducing reactions, the photosynthetic machinery transforms light energy into stable forms of ATP and NADPH, essential for energy and as sources of reducing equivalents during CO_2 fixation (Kuzyakov and Gavrichkova, 2011).

Genes with a two-fold up-regulation in peel encoding photosynthesis and photosynthesis antenna complex proteins were enriched in the apple fruit peel (Figure 6.1, 6.2 and 6.3) compared to pulp (Figure 6.1, 6.2). The enrichment of the photosynthesis related pathway genes in the peel were in agreement with the high chlorophyll content and photosynthetic capacity of the peel tissue (Fleancu, 2007). Non-enrichment of the photosynthesis pathway genes in the pulp tissue suggested that the pulp tissue was not actively involved in photosynthesis. The observations were in agreement with the findings that chlorophyll a and b content were higher in apple peel than pulp tissues and decreased during fruit development (Fleancu, 2007). These results suggested that the peel tissue dominates the light dependant reactions, which produces energy rich ATP and NADPH compounds mostly because pulp does not receive light for activation of the photosynthetic machinery. As such, though pulp parenchyma cells contain chloroplasts, their activity or contribution to photosynthesis might be very low. The pulp however, is not completely independent of the light reaction process since it uses ATP and NADPH products of the light reaction

(Tada, *et al.*, 2006). A time series analysis of the genes encoding photosynthesis pathway enzymes revealed peak expression of all the genes at daa35 to daa60.

Figure 6. 2 Localisation of apple fruit peel and pulp tissue enriched genes encoding the light reaction process of photosynthesis. The Figure s shows the passage of energy absorbed from light photons from photosystem II to photosystem I. Boxes coloured green represent sub pathway genes enriched in 'Golden Delicious' fruit peel.

Our observation suggested that fruit photosynthetic capacity reached its maximum between daa35 and daa60 then gradually decreased with fruit maturity. PSBO, PSBS, PSAH, PSAN, LHCA2, LHCB4, ATPG and ATPF were the most highly expressed, but slightly increased in expression after daa135 (Figure 6.3). The increase after daa135 was unexpected, considering 'Golden Delicious' fruit reaches physiological maturity at daa135 (Janssen, *et al.*, 2008). However, the result suggested that a basic minimum ATP

and/or NADPH supply could be required to augment the ATP produced by oxidative phosphorylation during the climacteric phase.
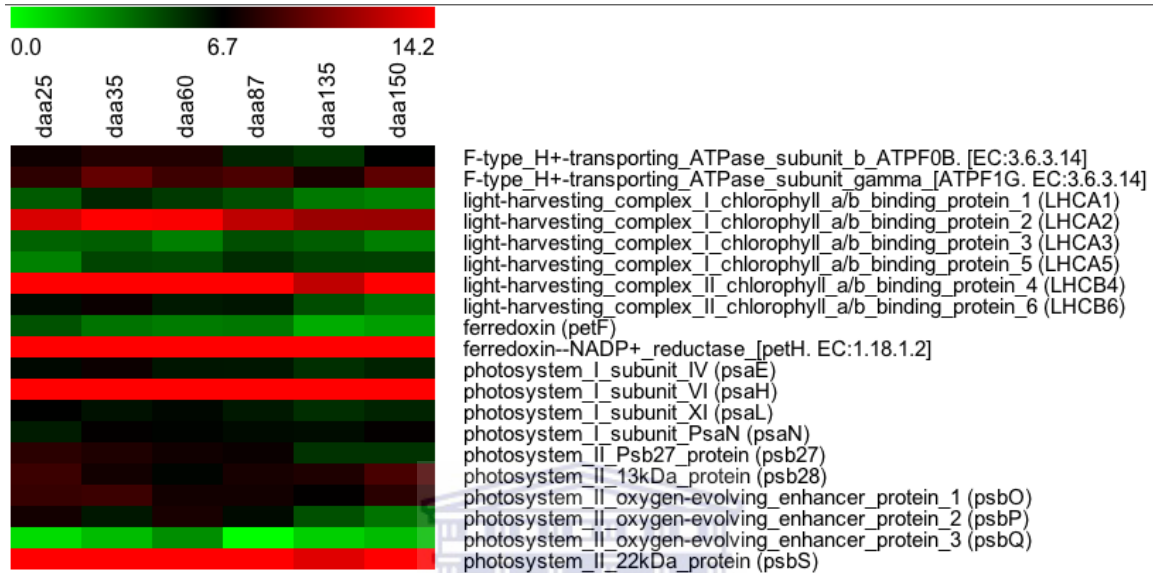
Figure 6. 3Heat map of expression profiles of peel tissue expressed genes encoding photosynthesis pathway proteins. The Figure shows a heat map of expression trends of two-fold up regulated apple fruit peel tissue expressed genes encoding the light harvesting complex and light reaction complex across six harvesting time points (35, 60, 87, 135, 150 in days after anthesis, daa). Expression values (FPKM) were log_2 transformed to normalize the data (section 2.2.17). The heat map key shows expression intensity in colour from lowest to highest as green to red respectively.

6.2.3   Oxidative phosphorylation

The mitochondria in plants can use its structure, enzymes, and energy released by the oxidation of nutrients to reform ATP in a process called oxidative phosphorylation. During oxidative phosphorylation, electrons are transferred from electron donors through a series of redox reactions carried out by protein complexes within mitochondria, which release energy used in the formation of ATP in response to a potential gradient known as chemiosmosis (Janicka-Russak, 2011).

Twenty-two ATP producing oxidative phosphorylation pathway genes were encoded for by genes enriched in the pulp (Figure 6.4) compared to two genes expressed in the peel tissue (Figure 6.5A, B, and Figure 6.1). Pulp enriched genes had lowest expression at daa60 (Figure 6.5A, B) while peel transcripts had peak expression at daa60 with a slight increase from daa87 to daa150 (Figure 6.5C).

Figure 6. 4 Oxidative phosphorylation KEGG pathway enzymes enriched in peel and pulp tissues of developing apple fruit. The Figure shows a schematic diagram of the localization of the electron transport chain complex proteins along the inter mitochondrial membrane. Boxes coloured pink, green and yellow represent sub pathway enzymes enriched in 'Golden Delicious' fruit pulp, peel and in both tissues (peel and pulp) respectively. Enriched oxidative phosphorylation pathway proteins: NADH dehydrogenase (ubiquinone) Fe-S protein 1 (NDUFS1), ubiquinol-cytochrome c reductase subunit 9 (QCR9), V-type H+-transporting ATPase subunit I (ATPeVI), inorganic pyrophosphatase (ppa), succinate dehydrogenase (ubiquinone) flavoprotein subunit (SDHA), V-type H+-transporting ATPase 16kDa proteolipid subunit (ATPeVPL), V-type H+-transporting ATPase subunit A (ATPeVA), V-type H+-transporting ATPase subunit B (ATPeVB), V-type H+-transporting ATPase subunit D (ATPeVD), V-type H+-transporting ATPase subunit AC39 (ATPeVAC39), V-type H+-transporting ATPase subunit E (ATPeVE), V-type H+-transporting ATPase subunit G (ATPeVG), and V-type H+-transporting ATPase 16kDa proteolipid subunit (ATPeVPL).

Figure 6.5 B shows that pulp transcripts were mostly enriched in EC 3.6.3.14, a multisubunit non-phosphorylated ATPases involved in the transportation of ions. EC 3.6.3.14 consists of V-type, F-type and A-type ATPases. These non-phosphorylated ATPases are composed of a membrane sector (Fo, Vo, Ao) and a cytoplasmic-compartment sector (F1, V1, A1) (Janicka-Russak, 2011). F-type ATPases are enzymes found in the inner membrane of mitochondria and chloroplast, which act as ATP synthases producing ATP in response to a H+ electrochemical potential gradient (Beyenbach and Wieczorek, 2006).

Vacuolar (H$^+$)-ATPases on the other hand, are ATP-dependent proton pumps that acidify intracellular compartments and, in some cases, transport protons across the plasma membrane of eukaryotic cells. V-ATPases are mostly found associated with vacuoles and clathrin coated vesicles. Unlike the F-type, V-type ATPases pump H+ rather than synthesise ATP (Toei, *et al.*, 2010).

In this regard, V-type ATPases are similar in mode of function to AHA2 (Janicka-Russak, 2011). AHA2 is a P-type ATPase that undergoes covalent phosphorylation during the transport cycle (Janicka-Russak, 2011). Though AHA2 enrichment was observed in the fruit peel in this study, in *A. thaliana*, AHA2 expression was observed in all tissues, but predominantly in roots. AHA2 enrichment in fruit peel might have functioned to control stomatal movements. That function, however, might have been lost with suberisation of the stomata as the fruit developed, possibly explaining the observed peak AHA2 expression at daa60 and a decrease thereafter (Figure 6.5C).
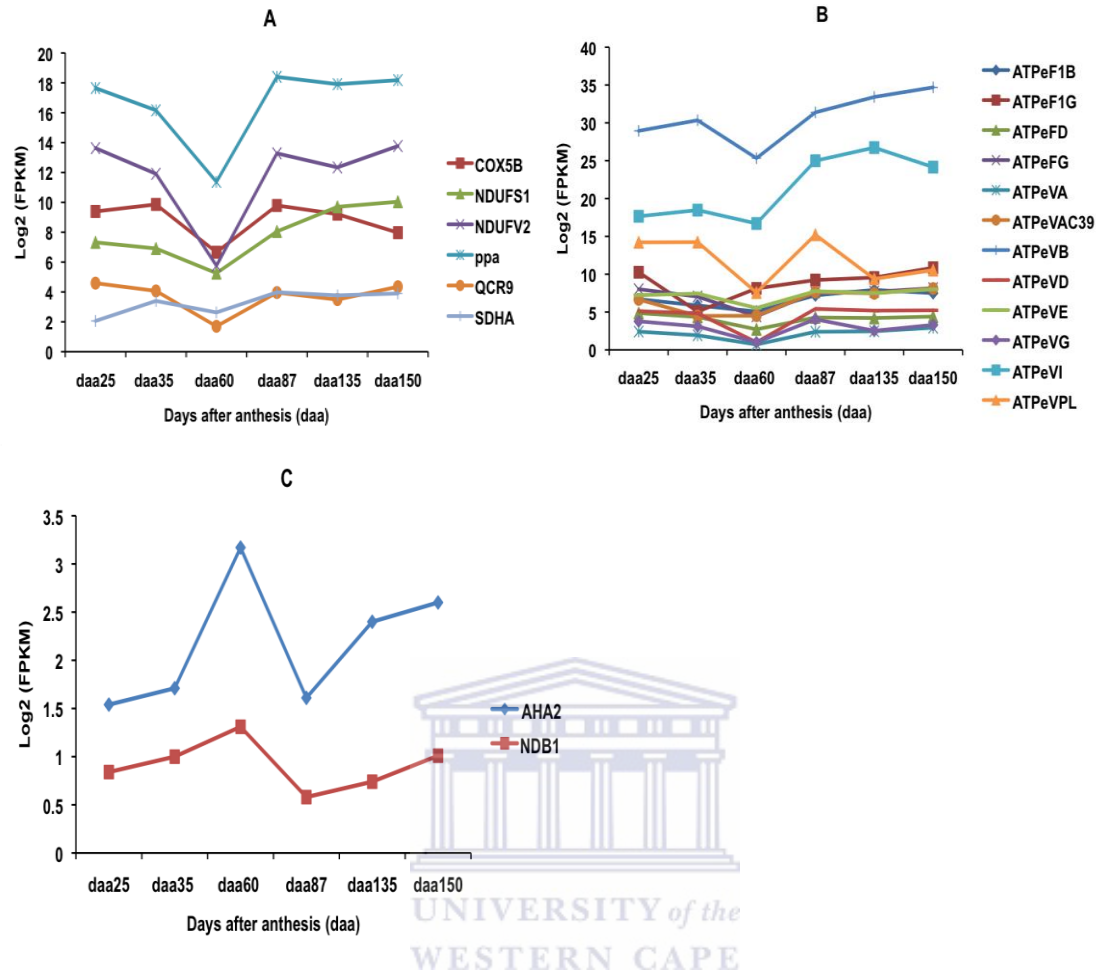
Figure 6. 5 Expression trend of transcripts encoding oxidative phosphorylation pathway enzymes during apple fruit development. Transcripts with more than two-fold up-regulation in pulp or peel ,Golden Delicious' fruit tissues were compared. A: Pulp expressed transcripts encoding components of the electron transport chain; B: pulp expressed transcripts encoding ATPases of the F-type and V-type; C: Peel expressed transcripts encoding an ATPase and an electron transport chain component I protein. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

High levels of V-type ATPases in the pulp might have been required for regulating cellular solute content; hence turgidity and optimum pH balance for optimum physiological functions (Saroussi and Nelson, 2009). The proton gradient established by V-type ATPases might be involved in the membrane energization used for fruit phloem unloading of photosynthates and xylem unloading of nutrients (Janicka-Russak, 2011). The "acid growth" theory suggests that protons extruded by an activated H+-ATPase

221

decrease the apoplastic pH and activate enzymes involved in cell wall loosening, a process activated by auxin, resulting in loosening of the cell wall (Hager, 2003). Thus most of the 'Golden Delicious' fruit cell growth might be accounted for by pulp cell growth changes during fruit development.

The enrichment of the F-type ATPases in the pulp suggests that physiological processes occurring in the pulp were high ATP (energy) demanding. The expression pattern of the F-type ATPases in pulp showed an expression trend similar to the reported decrease in respiration rates towards pre-climacteric and an increasing respiration rate as the fruit enters the climacteric phase (Janicka-Russak, 2011). It is however important to note that the ATP requirements for the physiological processes in the peel might have been augmented by ATP produced from the light reaction process of peel photosynthesis, hence reducing the need for high activation of the F-type genes. Since V-type ATPases and AHA2 significantly redundant in function, in peel tissue the function of the V-type ATPases might have been carried out by AHA2. This observation suggests differential tissue specific H+-ATPase enrichment.

6.2.3 Cuticle metabolism in apple fruit peel tissue

Cuticleand epicuticular waxes accumulation in apple fruit is an important traitin the agronomy of apple fruit. Components, accumulation and importance of cuticle in apple fruit are as described in section 5.2.3.1.Because of the agronomic importance of the cuticle, as well as absence of detailed elucidation of the genes involved in fruit cuticle

production, this section sheds light on the changes in cuticle metabolism and contribution of peel and pulp tissues to cuticle formation during apple fruit development.

Fatty acid hydroxylases, CYP86A4S (cytochrome P450, family 86, subfamily A, polypeptide 2/4/7/8 (fatty acid omega-hydroxylase, (EC:1.14.-.-)) and CYP86A1 (cytochrome P450, family 86, subfamily A, polypeptide 1 (fatty acid omega-hydroxylase, EC:1.14.-.-)), catalyse the conversion of 16-palmitic acid to 16-hydroxypalmitate and oleic acid to 18-hydroxyoleate respectively (Beisson*, et al.*, 2012). HHT1 (omega-hydroxypalmitate O-feruloyl transferase, EC 2.3.1.188) is involved in the conversion of hydroxypalmitate and feruloyl-CoA to 16-feruloyloxypalmitate (Cajuste*, et al.*, 2010).

The fatty acid hydroxylases as well as HHT1 enzymes are involved in unsaturated fatty acid biosynthesis, precursors for very long chain fatty acid biosynthesis (Beisson*, et al.*, 2012). Transcripts encoding these unsaturated fatty acid biosynthesis enzymes had peak expression at daa60 (Figure 6.7A), suggesting at daa60 the formation of precursors for cutin, suberin and wax formation had reached their maximum.
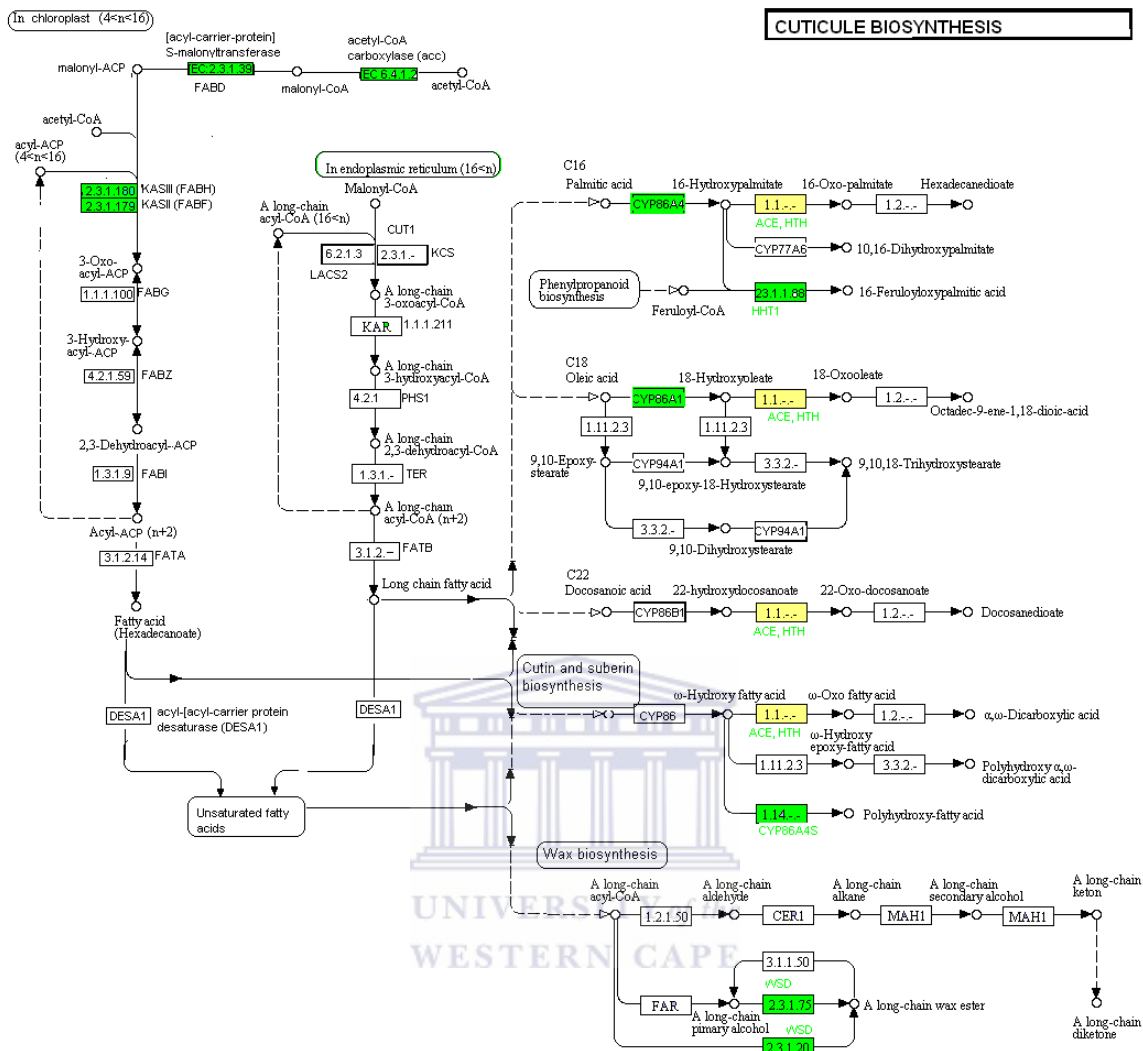
Figure 6. 6 Localisation of cuticle biosynthesis enzymes enriched in apple fruit peel and pulp tissues. The Figure shows a modified combined fatty acid synthesis, fatty acid elongation, cutin, wax, unsaturated fatty acid biosynthesis KEGG pathway reference map. Genes encoding fatty acid and cutin/wax biosynthesis were enriched in the peel tissue, while pulp tissue was enriched for a cutin-forming enzyme. Boxes coloured green and yellow represent sub pathway genes enriched in 'Golden Delicious' fruit peel and in both peel and pulp tissues. Names of tissue-enriched enzymes are written close to the EC number of the enzyme. Pathway enzymes not highlighted in colour were not significantly enriched in apple fruit peel tissue.

The very long chain fatty acids produced by the fatty acid elongation and formation of unsaturated fatty acids processes are again hydroxylated by either CYP86A1 or CYP86A4S to ω-hydroxy fatty acid, and subsequently to ω-oxo fatty acids by fatty acid

omega-hydroxy dehydrogenase (ACE, HTH, EC 1.1.-.-) towards formation of cutin (Beisson, *et al.*, 2012).
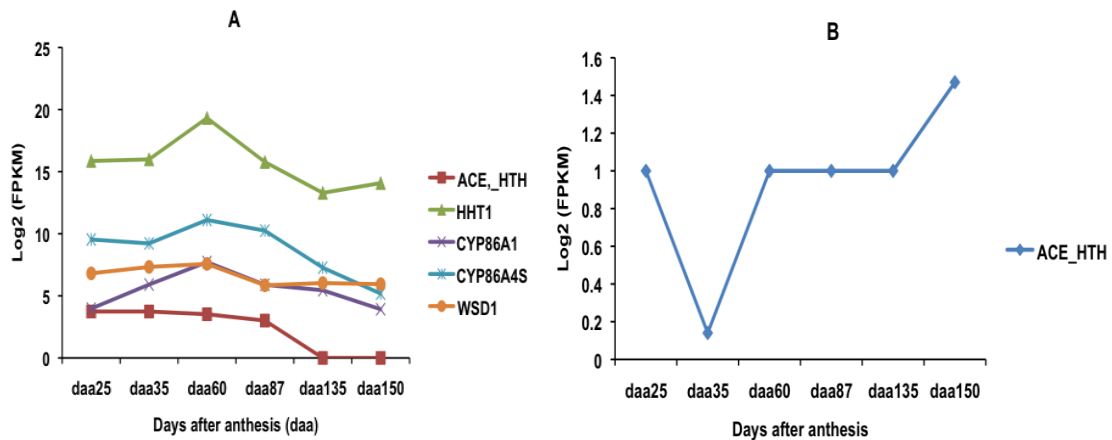
Figure 6. 7 Expression profiles of cutin, suberin and wax metabolism encoding transcripts. A: Expression trends of peel up regulated transcripts enriched for fatty acid and wax biosynthesis enzymes; B: Expression trend of pulp up regulated transcript enriched for fatty acid biosynthesis. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17).

Genes encoding ACE_HTH had high expression early during fruit peel development but decreased at daa135 (Figure 6.7A). The progressive decrease in ACE_HTH coincided with a decrease in long chain fatty acid synthesizing enzymes, suggesting cutin formation might have stopped at daa135, when the fruit reached physiological maturity. Pulp expressed transcripts encoding ACE_HTH, showed low enrichment of the gene (Figure 6.7B). Though lowly expressed, transcripts encoding ACE_HTH in pulp decreased early during fruit pulp development to the lowest level at daa35, after which they increased to daa150. The expression of ACE_HTH in the pulp, though low, suggested the possible contribution of pulp tissue towards cutin and suberin accumulation on apple fruit surfaces. Nonetheless, the result suggested that peel tissue was the major contributor towards cutin and suberin accumulation.

The wax-ester synthase / diacylglycerol O-acyltransferase (WSD1, EC 2.3.1.75, 2.3.1.20) catalyses the conversion of long chain acyl-CoAs as well as long chain primary alcohols to long chain wax esters (Dong*, et al.*, 2012). Peak WSD1 expression at daa60, suggested that wax accumulation reached its maximum at daa60 after which formation of new wax decreased (Figure 6.7A). Accumulation of wax on apple fruit surfaces have been reported by Atkinson, *et al.* (2012) and by Dong, *et al.* (2012), but no detail is provided on the progression of wax formation indicating periods of peak formation (Atkinson*, et al.*, 2012). The accumulation of wax on the peel surface regulates water loss, thus acting as a protective layer against desiccation, turgor pressure from inside and biotic and abiotic factors from outside, as well as determining the fruit's glossiness, hence its marketability.
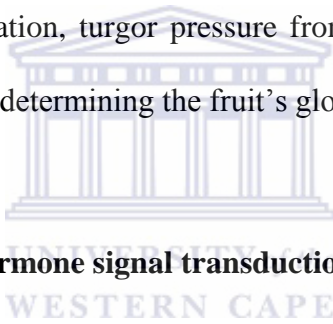
6.2.4   Plant hormone signal transduction

6.2.4.1 Auxin

The whole repertoire of auxin signal transduction pathway genes were represented by the combined data sets of peel and pulp 'Golden Delicious' fruit tissue transcripts. A comparison of peel and pulp tissue enriched genes shows that TIR1 (transport inhibitor response 1), auxin response factor (K14486, ARF), and auxin responsive GH3 gene family (GH3) were exclusively enriched in the pulp (Figure 6.8A), while auxin influx carrier (AUX1 LAX family) (AUX1, LAX) was exclusively enriched in the peel (Figure 6.8B). The auxin influx carrier proteins consist of four family members, AUX1, LIKE AUX 1, 2, and 3 (Kleine-Vehn*, et al.*, 2006).
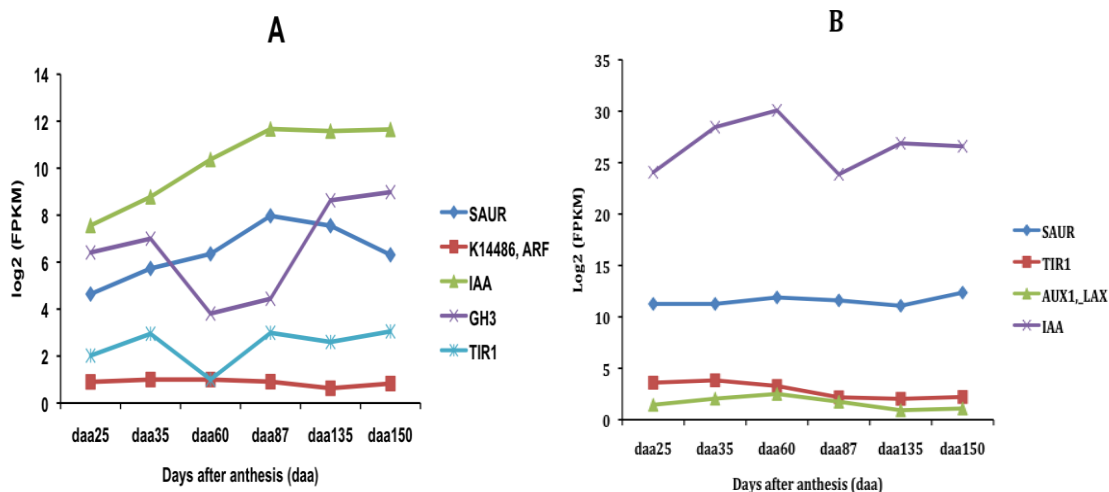
Figure 6. 8 Expression profiles of peel and pulp transcripts encoding enriched proteins in the auxin hormone signal transduction pathway. A: Expression profiles of pulp two-fold up regulated transcripts encoding auxin signal transduction pathway proteins. B: Expression profiles of peel two-fold up regulated transcripts encoding auxin signal transduction pathway proteins. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

These proteins together with auxin diffusion are responsible for the influx of auxin into cells. The distribution of auxins to other cells is through the action of auxin efflux carrier proteins, such as PIN3 (Kleine-Vehn*, et al.*, 2006; Vandenbussche*, et al.*, 2010). Auxins are synthesised in actively dividing plant organs, such as root tips, shoot and leaf primordial (Cova*, et al.*, 2012). In apple fruit the actively dividing and enlarging parenchyma cells are located in the mesocarp, which constitute the pulp tissue (Cova*, et al.*, 2012). Considering the peel tissue has no actively dividing cells, auxin-induced physiological processes in the peel might therefore relay on auxin synthesised in the mesocarp, hence the need for auxin influx carrier proteins to import and maintain an auxin gradient in the target cells. The auxin gradient and concentrations might then trigger a myriad of physiological processes, ultimately defining peel properties. The pulp tissue on the other hand might be relaying on local auxin synthesis and efflux carrier

227

proteins for distribution to outer parenchyma cells, possibly explaining the non-enrichment of the influx proteins in the pulp.

The Aux/IAAs (auxin-responsive protein IAA) are negative regulators of auxin signalling. They are composed of four domains. The first domain contains an ERF-associated amphiphilic repressor (EAR) motif essential for transcriptional repression recruitment of the transcriptional co-repressor TOPLESS (TPL) (Chapman and Estelle, 2009). Domain II of Aux/IAAs is essential for auxin-stimulated Aux/IAA proteolysis. The last two domains are also essential as transcriptional repressors through homo- and hetero-dimerization and interaction with the other transcription regulators, ARF (Chapman and Estelle, 2009). ARFs mediate auxin-dependent transcriptional regulation by binding to the auxin responsive element, a consensus sequence found in promoters of auxin-inducible genes. Thus the amino acid sequence of the auxin responsive element determines whether the ARF activate or repress transcription (Chapman and Estelle, 2009). The activating ARFs are however negatively regulated by interaction with Aux/IAAs, which recruit transcription repressor TPL.

Increasing auxin concentrations also increase Aux/IAAs turnover, releasing activating ARFs from TPL-mediated repression, hence promoting transcription of auxin-regulated genes (Hagen, *et al.*, 2010). The existence of numerous Aux/IAAs as well as the ability of ARFs to form homo- and heterodimers results in different combinations that can drive or inhibit transcription of a particular set of auxin-responsive genes depending on the context (Hagen, *et al.*, 2010). This hypothesis is corroborated by the fact that expression

of different stabilised Aux/IAAs in the same cell types does not always result in the same phenotypes (Hagen*, et al.*, 2010). The presence of other factors that can interact with ARF, suggests that there might be an even greater combinatorial possibilities in auxin-mediated transcriptional regulation. The Skp1-cullin-F box protein (SCF) E3 ubiquitin ligase, F box component, TIR1 (SCF^{TIR1}) recruits Aux/IAA, which interacts with TIR1 through domain II in an auxin-dependent manner (Hagen*, et al.*, 2010). The binding of the Aux/IAA-TIR1 to SCF^{TIR1}, results in the ubiquitination of the complex marking it for degradation, thus regulating the stability of the Aux/IAAs. TIR1 binds auxins at physiologically relevant concentrations (K_d ~20-80 nM) stimulating the interaction of TIR1 and Aux/IAA (Hagen*, et al.*, 2010). At low auxin concentrations Aux/IAAs are more stable and can recruit co-repressor, TPL and interact with ARFs inactivating them. At high auxin concentrations the TIR1-Aux/IAA complexes are stabilised and in association with SCF^{TIR1} are ubiquinated and targeted for proteolysis, resulting in ARF de-repression and modulation of transcription (Hagen*, et al.*, 2010; Vanneste and Friml, 2009).

The SMALL AUXIN UP RNA (SAUR) family protein, GH3 and Aux/IAAs are the three classes of auxin responsive gene families. Members of each of these families are up regulated rapidly in response to auxin. GH3 encode auxin conjugating enzymes, which act to reduce free auxin levels, while Aux/IAAs encode the rapidly turned over transcriptional repressors of auxin-inducible genes (Leyser, 2006). Transcripts encoding *SAUR* were lowly expressed in both peel and pulp tissues but gradually increased and decreased in pulp and peel respectively (Figure 6.8 A, B). SAUR form the largest group

of auxin-responsive genes, are largely redundant and have no elucidated function (Spartz*, et al.*, 2012). However gain and loss-of-function studies suggested SAUR might function as effectors of cell expansion through modulation of auxin transport (Spartz*, et al.*, 2012). The expression of these genes in pulp and peel suggested auxin auto-regulation in these tissues. Unlike in peel, the pulp tissue had enriched levels of GH3, suggesting the existence of multiple auxin auto-regulation mechanisms in pulp tissue, possibly to cope with the high auxin levels in the mesocarp.

6.2.4.2 Cytokinin

All four components of the cytokinin signal transduction pathway genes were encoded for by a combined set of transcripts up regulation in peel and pulp tissues of developing apple fruit.
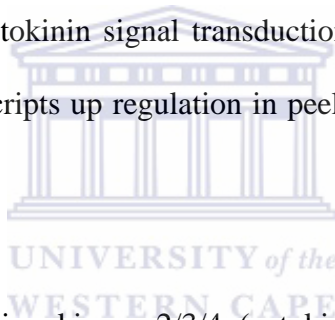
AHK2/3/4 (Arabidopsis histidine kinase 2/3/4 (cytokinin receptor), EC 2.7.13.3),are histidine kinases, which initiate intracellular phosphotransfer (Jones*, et al.*, 2010; Schaller*, et al.*, 2011) and were enriched in the peel tissue with peak expression at daa60 (Figure 6.9 B). Two-component response regulator ARR-B family (ARR), type-A and type B are cytokinin response regulators, which act as repressors of cytokinin-activated transcription and DNA-binding transcriptional activators, respectively (Kakimoto, 2003). Activation of cytokinin-mediated transcription has been linked to increased cell division and shoot initiation (Piwien-Pilipuk*, et al.*, 2011). The histidine-containing phosphotransfer protein (AHP), gene products are positive elements in cytokinin signal transduction, mediating signalling between the plasma membrane-bound AHK2/3/4

receptors and ARRs within the nucleus (Kakimoto, 2003). In fruit pulp, increased expression of AHP might have increased interaction with ARR-B hence increased mesocarp meristematic cell division, maintenance and differentiation in the pulp (Yamashino and Mizuno, 2009).
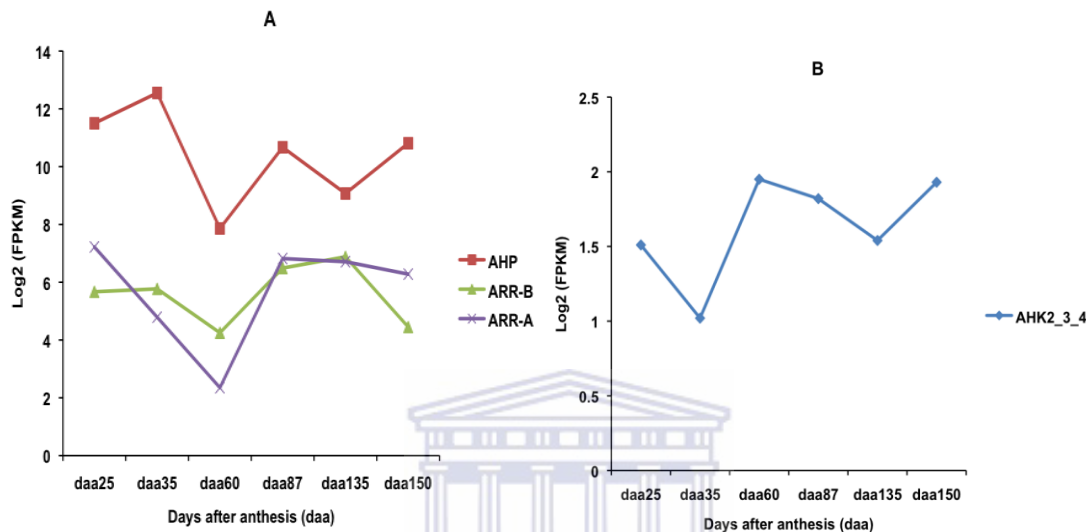
Figure 6. 9 Expression profiles of peel and pulp transcripts encoding cytokinin signal transduction pathway proteins. A: Expression profiles of pulp greater than two-fold up regulated transcripts encoding cytokinin signal transduction pathway proteins. B: Expression profiles of peel greater than two-fold up regulated transcripts encoding cytokinin signal transduction pathway proteins. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.4.3 Brassinosteroids (BRs)

Transcripts encoding the growth-promoting steroid hormone, BR signal transduction pathway genes, *brassinosteroid resistant 1-2 (BZR1 /2),* showed more than two-fold up-regulation in apple pulp tissue. The apple peel tissue transcripts did not show enrichment of any of the BR signal transduction pathway genes (Figure 6.10). The *BZR1 /2* gene encoding pulp transcripts had transitory peak at daa35 and daa87 (Figure 6.10). The BR signal transduction cascade, includes BR perception by the BRI1 receptor kinase at the cell surface, activation of BRI1/BAK1 kinase complex by trans phosphorylation,

subsequent phosphorylation of the BSK kinases, activation of the BSU1 phosphatase, dephosphorylation and inactivation of the BIN2 kinase, and accumulation of unphosphorylated BZR transcription factors in the nucleus (Clouse, 2011; Kim and Wang, 2010; Tang*, et al.*, 2008). The accumulation of BZR1 /2 in the nucleus might have functioned to regulate nuclear gene expression and plant development by promoting cell division and cell enlargement. Genome wide analyses of protein-DNA interactions identified BZR1 target genes linking BR signalling to various cellular, metabolic, and developmental processes, as well as other signalling pathways (Tang*, et al.*, 2008; Wang, Bai*, et al.*, 2012). By controlling photomorphogenesis, BR signalling is highly integrated with the light, gibberellin, and auxin pathways through both direct interactions between signalling proteins and transcriptional regulation of key components of these pathways. BR signalling has been reported to cross talk with other receptor kinase pathways as well as modulate stomata development and innate immunity (Tang*, et al.*, 2008; Wang, Bai*, et al.*, 2012).
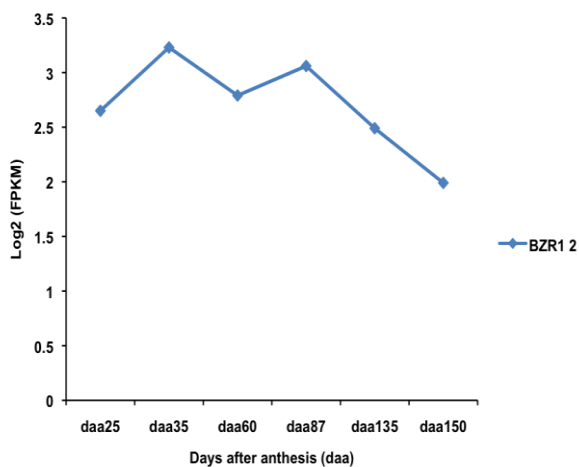
Figure 6. 10 Expression profile of pulp transcripts encoding BR signal transduction pathway proteins. The Figure s shows expression trend of pulp greater than two-fold up regulated transcript encoding brassinosteroid signal transduction pathway protein, brassinosteroid resistant 1-2 (BZR1

2).Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.4.4 Abscisic acid (ABA)

Protein phosphatase 2C (PP2C, EC 3.1.3.16) regulates the phosphorylation status of serine/threonine-protein kinase SRK2 (SNRK2, EC 2.7.11.1), hence regulating their activation (Hirayama and Shinozaki, 2007). However the physiological function of SNRK2 is dependent upon which calcium-binding regulator is attached, indicating to the complex network of SNRK2, calcium-binding regulatory components, and PP2C in ABA response (Hirayama and Shinozaki, 2007). Ca^{2+}-dependent protein kinase (CDPK) functions to regulate ABA responsive element binding factor (ABF). SNRK2 interacts directly with ABFs and/or phosphorylate them. Multiple phosphorylation sites on ABFs, suggest that ABFs might function as an integration point for multiple signals (Hirayama and Shinozaki, 2007). Expression of ABF in the peel (Figure 6.11B) might have functioned to control stomatal responses through transcriptional gene regulation. High expression of ABF in the peel early during fruit development was in concert with the changes in stomatal anatomy with fruit development (White, 2002). From the result it could speculated that peak stomatal plasticity might at daa35 after which the stomata is suberized and looses its plasticity. A loss in stomatal plasticity might have resulted in down regulation of ABFs in peel as the fruit matured.

The progressive increase in SNRK2 with fruit development might have been in response to internal stresses (Figure 6.11A). Because of high oxidative phosphorylation in the pulp (Figure 6.1) the levels of reactive oxygen species (ROS) were expected to increase

233

resulting in PP2C inactivation (Hirayama and Shinozaki, 2007), as such the levels of SNRK2 increased. Transcripts encoding PP2C decreased with fruit development (Figure 6.11B), further corroborating the hypothesis of PP2C reduction as the fruit enters the climacteric phase and the deactivating effects of ROS produced during the climacteric, reducing PP2C activity. In tomato, ABA concentration in tissues close to the seed increased with fruit development. The gradient of ABA concentration decreased the further away from seeds, a phenomenon that has been attributed to a decrease in viviparous seed germination (Gillaspy, *et al.*, 1993). Thus PP2C and SNRK2 might be the links to multiple regulatory mechanisms during apple fruit development.
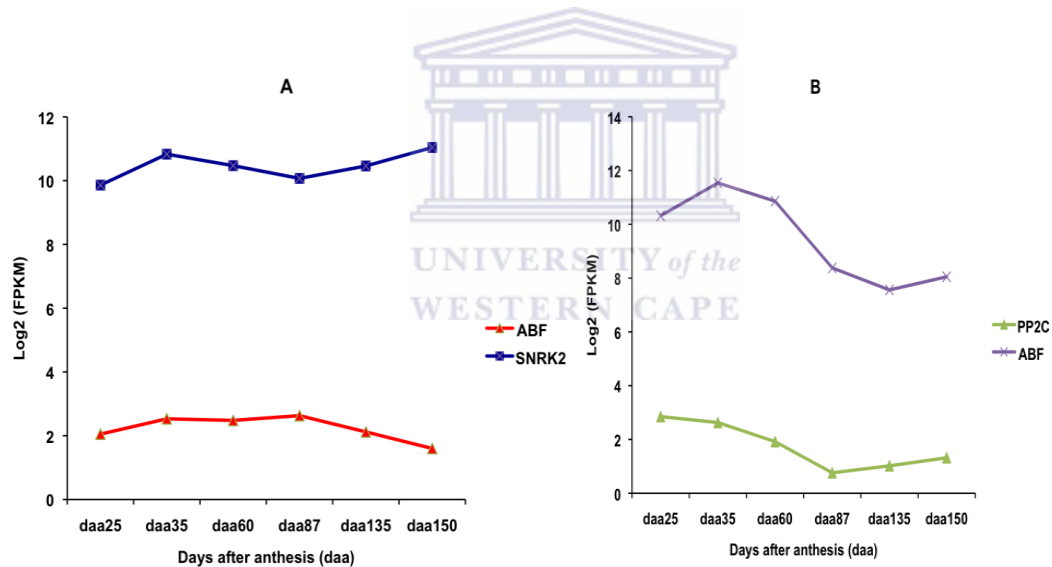
Figure 6. 11 Expression profiles of peel and pulp transcripts encoding Abscisic Acid (ABA) signal transduction pathway proteins. A: Expression profiles of pulp greater than two-fold up regulated transcripts encoding abscisic acid, signal transduction pathway proteins. B: Expression trends of peel greater than two-fold up regulated transcripts encoding abscisic acid signal transduction pathway proteins. Expression values (FPKM) were $\log_2$ transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.4.5 Ethylene signal transduction

Transcripts encoding ethylene signal transduction pathway proteins were only enriched in

the apple pulp. Transcripts that showed more than two-fold up-regulation in the pulp encoded for a mitogen-activated protein kinase 6 (MPK6, EC 2.7.11.24)protein (Figure 6.12), whereas peel transcripts were not enriched for any ethylene signal transduction pathway proteins. MPK6 proteins are ethylene-inducible MAP-kinase proteins, together with the stress-induced MAP-kinase kinase (MAPKK) (SIMKK), they constitute a MAP-kinase cascade that operate downstream of the CONSTITUTIVE TRIPLE RESPONSE 1 (CTR1) protein (Adams-Phillips, *et al.*, 2004). The *MPK6* genes, though they have homologous in *Arabidopsis* and tomato, they have remained poorly defined. Mining of the Tomato Expression Database (http://ted.bti.cornell.edu/) revealed that MPK6 was up-regulated during tomato fruit ripening, a result consistent with other genes encoding ethylene signalling components in maturing tomato fruit (Adams-Phillips, *et al.*, 2004). However, unlike in tomato, the expression of the *MPK6* in 'Golden Delicious' fruit pulp tissue did not show a clear progressive increase with development. MAP-kinases have been implicated in a number of signal transduction pathways. This multi-signal pathway involvement by the MAP-kinases made it difficult to assign changes in MAP-kinase transcript abundance to a single hormone signal transduction pathway. As such the observed MPK6 expression profile might be in response to multiple signals. Downstream of the MPK6 are membrane-bound receptor ethylene-insensitive-2 (EIN-2) and then two sets of transcription factors, including the ethylene-insensitive like (EILs) family and the ethylene response factors (ERFs). ERFs are believed to target genes involved in flavour biosynthesis and texture modification as the fruit ripens (Adams-Phillips, *et al.*, 2004).
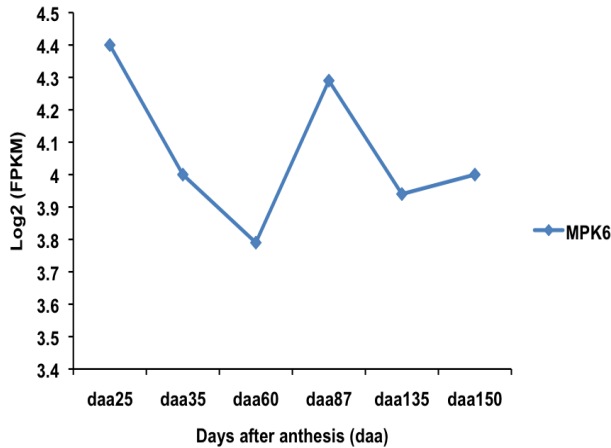
Figure 6. 12 Expression profiles of pulp transcripts encoding genes in the plant hormone signal transduction pathway. Expression trend of pulp greater than two-fold up regulated transcripts encoding ethylene signal transduction pathway mitogen-activated protein kinase 6 (MPK6, EC:2.7.11.24) protein. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.4.6 Gibberellin

Gibberellin signal transduction pathway genes, *DELLA protein (DELLA),* and *F-box protein GID2 (GID2, SLY1)* were encoded for by pulp and peel 'Golden Delicious' fruit expressed genes (Figure 6.13 A, B). DELLA encoding genes were enriched in the peel, while GID2/SLY1 encoding gene was enriched in the pulp (Figure 6.13 A, B). DELLA encoding genes were lowly expressed but decreased in expression with fruit development. GID2/SLY1 encoding genes on the other hand were more expressed and showed an early decrease in expression to daa60 and slight increase there after (Figure 6.13B).

DELLA proteins are considered 'molecular switches' for GA signalling (Hauvermale, *et al.*, 2012; Hedden, 2012). The processing of the GA signal in the nucleus has been reported to depend directly on the presence or absence of DELLA proteins. In the

236

presence of GA the GID2/SLY1 complex is activated triggering GID2/SLY1-mediated DELLA protein degradation, which result in the transduction of the GA signal (Gomi and Matsuoka, 2003; Hauvermale, *et al.*, 2012). The enrichment of DELLA in the peel tissue suggests that the peel tissue might have contained low GA, and that DELLA inhibited GA signal transduction. Enrichment of the GID2/SLY1 complex in the pulp tissue suggested the existence of GID2/SLY1-mediated DELLA protein degradation and hence GA signal transduction. In a developmental analysis of tomato fruit, GA only accumulated during periods coinciding with tomato fruit cell division and expansion (Gillaspy, *et al.*, 1993). As such inactivation of the GA signal by DELLA in the peel and activation of the GA signal transduction in pulp, corroborated pulp as an actively dividing and expanding tissue. Thus most of the fruit's increase in volume might be accounted for by changes in pulp tissue cells.
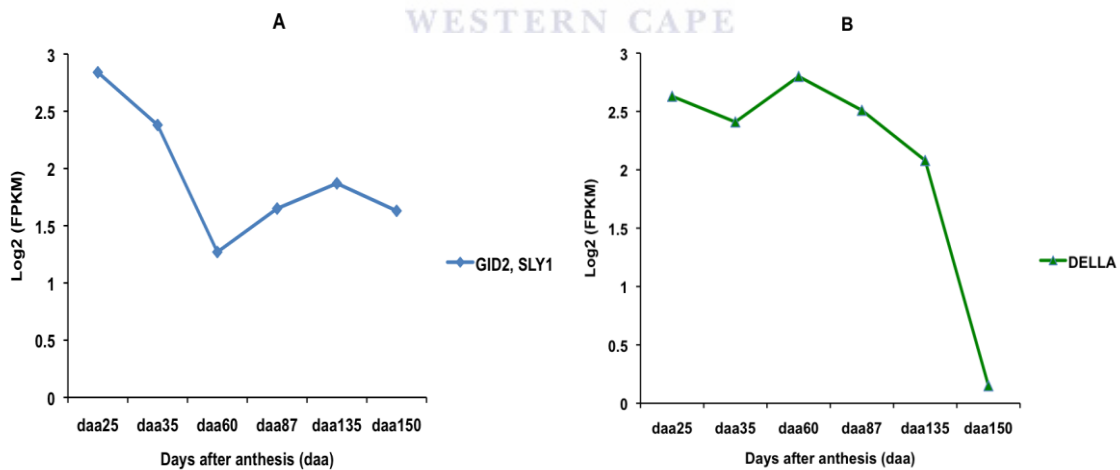
Figure 6. 13 Expression profiles of pulp and peel transcripts encoding gibberellin (GA) signal transduction pathway proteins. A: Expression trend of pulp greater than two-fold up regulated transcript encoding GA signal transduction pathway protein. B: Expression trend of peel greater than two-fold up regulated transcript encoding GA signal transduction pathway protein. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.5   Pigment accumulation

6.2.5.1 Flavonoids

Flavonoids, including anthocyanins and flavonols are a class of secondary metabolites, derived from amino acid phenylalanine, that impart invaluable health attributes most probably through their antioxidant activity (Wolfe, *et al.*, 2003; Wolfe, *et al.*, 2008). Pulps up regulated transcripts were enriched for *chalcone isomerase (CHI, EC 5.5.1.6)*, while peel up regulated transcripts were enriched for *anthocyanidin reductase (ANR, EC:1.3.1.77),* and *flavonol synthase (FLS, EC:1.14.11.23)* (Figure s 6.14, 6.15). No anthocyanin pathway genes encoding transcripts were enriched in the pulp tissue. In the peel however, *anthocyanin 5-O-glucosyltransferase (UGT75C1, EC:2.4.1.-)* was enriched (Figure s 6.15A). 'Golden Delicious' fruit is one of the very unlikely apple fruits were one would look for red colour pigmentation formation and changes thereof. However, the observed colour changes from reddish colouration during early fruit development through deep green to golden yellow fruit colouration at maturity necessitated the study on pigment changes. Glucosyltransferases (UGT75C1), glycosylates anthocyanidins into colourful anthocyanins (Wrolstad, *et al.*, 2006). Though observed in this study, glucosyltransferases are well documented for red pigment accumulating cultivars to increase with fruit maturity, for example Royal gala (Janicka-Russak, 2011).

Genes encoding UGT75C1 had a transitory peak expression at daa60 and increased in expression from daa87 (Figure 6.15A). The enrichment of UGT75C1 does not conclusively indicate as to which of the pelargonidin, cyanidin or delphinidin glycosides

accumulated in 'Golden Delicious' peel fruit tissue (Figure 6.14). Delphinidin based anthocyanidins are known to give blue hues to flowers like dephiniums, and blue-red colour to cranberries. Pelargonidin based anthocyanidins on the other hand produce characteristic orange colour, and are found in large amounts in red kidney beans, strawberries and plums among others (Wrolstad, *et al.*, 2006). Cyanidins give a characteristic reddish-orange colour, which can change with pH from red, to violate then blue at acidic, neutral, and alkaline conditions respectively. The hue of all the anthocyanidins is pH dependent (Cevallos-Casals and Cisneros-Zevallos, 2004; Wrolstad, *et al.*, 2006). Cyanidins are the major anthocyanins in apple, with highest concentration in the peel (Lister, *et al.*, 2006).
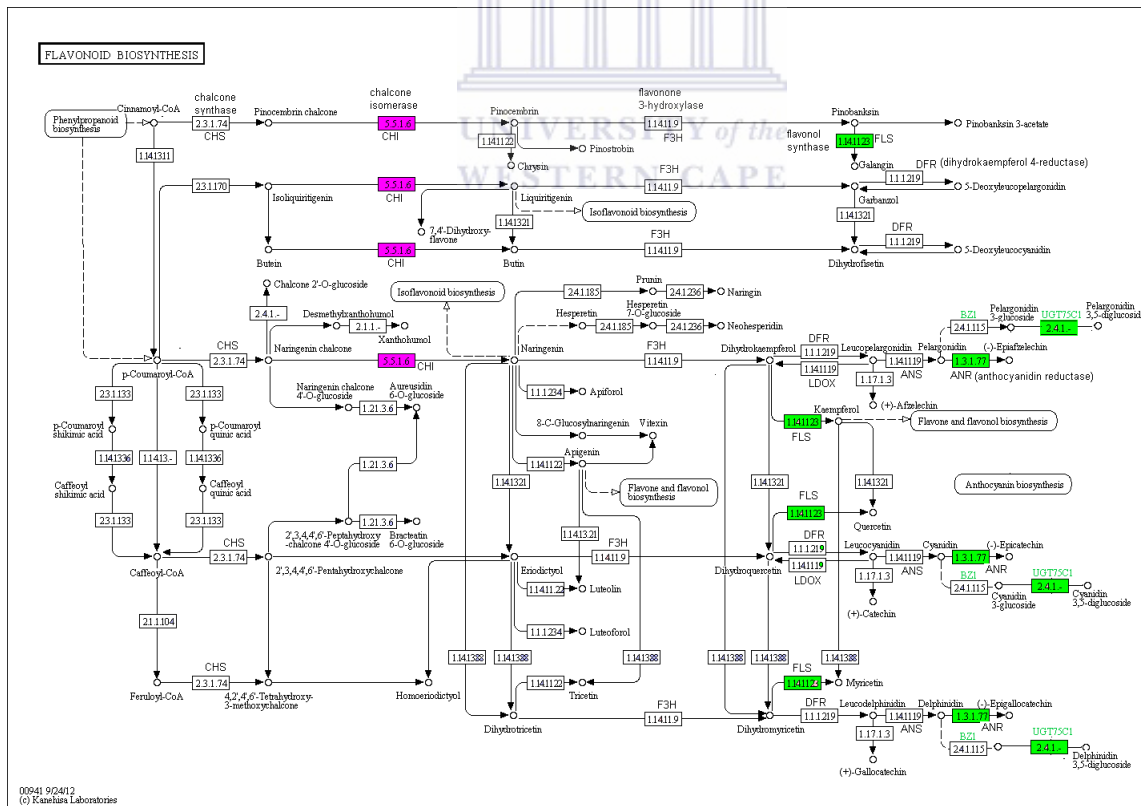
Figure 6. 14 Localisation of flavonoid and anthocyanin pathway enzymes enriched in peel and pulp apple fruit tissues. The Figure shows a modified combined flavonoid and anthocyanin biosynthesis KEGG pathway reference map. Boxes coloured pink and green represent sub pathway genes enriched in 'Golden

Delicious' fruit pulp and peel tissues respectively. Names of tissue-enriched enzymes are written close to the EC number of the enzyme.

FLS converts dihydrokaempferol, pinobenksin, dihydroquercetin, and dihydromyricetin to kaempferol, galangin, quercetin and myricetin respectively (Figure 6.14). The enrichment of FLS in apple peel tissue might suggest accumulation of anyone of the flavanols mentioned above. However in apple, quercetin has been documented as the major flavanol in apple, more so in the peel tissue (Lister*, et al.*, 2006). Genes encoding FLS had high expression early in fruit development, which decreased at daa35 followed by peak expression at daa60 and a final increase from daa135 to daa150 (Figure 6.15A). The observations did not agree with the findings by Lister *et al*, (2006), who observed highest quercetin glycosides content in very young fruits (Lister*, et al.*, 2006). The observations however, suggested that flavonol accumulation decreased as the fruit entered the pre-climacteric but increased with the burst in respiration, possibly to function as antioxidants. The reason for FLS transitory peak at daa60 was not known. The free radical scavenging capacity of quercetin has been reported, and together with other flavonoids are a big area of study under functional foods (Lister*, et al.*, 2006).

The enrichment of ANR in the peel tissue (Figure s 6.14, 6.15A), suggested that proanthocyanidin synthesis in 'Golden Delicious' fruit peel might not relay on the conversion of leucoanthocyanidin by leucoanthocyanidin reductase (LAR), rather leucoanthocyanidin synthesis branches off from anthocyanin pathway after LDOX step via ANR. A similar mechanism of proanthocyanidin (PA) synthesis has been reported in *Arabidopsis*, which lacked leucoanthocyanidin reductase (LAR) (Janicka-Russak, 2011). The resultant flavan-3-ols, catechin, epichatichin, and epiafzelechin and epigallocatechin

are transported into vacuoles were they are polymerized into condensed tannins (Janicka-Russak, 2011). The progressive increase in ANR suggested a possible gradual accumulation of the flavan-3-ols as the fruit developed. In human health, polymeric flavan-3-ols are regarded as xenobiotics, with limited bioavailability depending on the degree of polymerization (Monagas, *et al.*, 2010). However, the metabolites of these condensed tannins might be responsible for the health effects derived from flavan-3-ol consumption (Monagas, *et al.*, 2010).
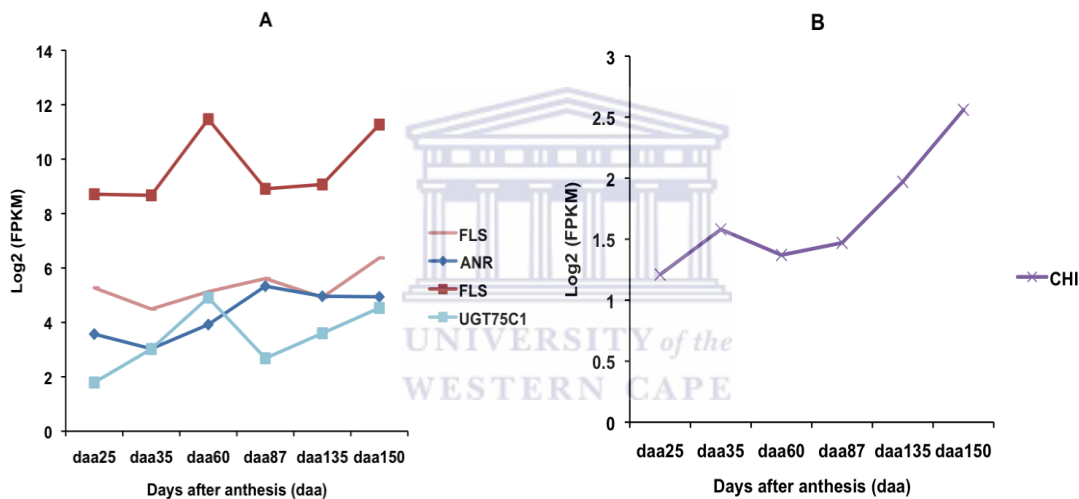
Figure 6. 15 Expression profiles of peel and pulp transcripts encoding flavonoid biosynthesis proteins. Expression patterns of peel (A) and pulp (B) greater than two-fold up regulated transcripts encoding flavonoid and anthocyanin biosynthesis proteins. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.5.2 Carotenoid biosynthesis

Apple fruit chloroplasts have been reported to change to chromoplasts with fruit development. These chromoplasts act as reservoirs of colour compounds that act as attractants of seed dispersers. The characteristic golden colour of 'Golden Delicious' has

been attributed to the accumulation of ß-carotene (Gonkiewicz, 2011). Pulp and peel two-fold up regulated transcripts showed enrichment of carotenoid biosynthesis enzymes shown in colour in Figure 6.16, green, pink, and yellow for peel, pulp and expression in both peel and pulp tissues respectively. However, the majority of the carotenoid biosynthesis enzymes were enriched in the peel, while the pulp tissue was only enriched for *beta-carotene 3-hydroxylase (crtZ, EC 1.14.13.129)* (Figure 6.16).
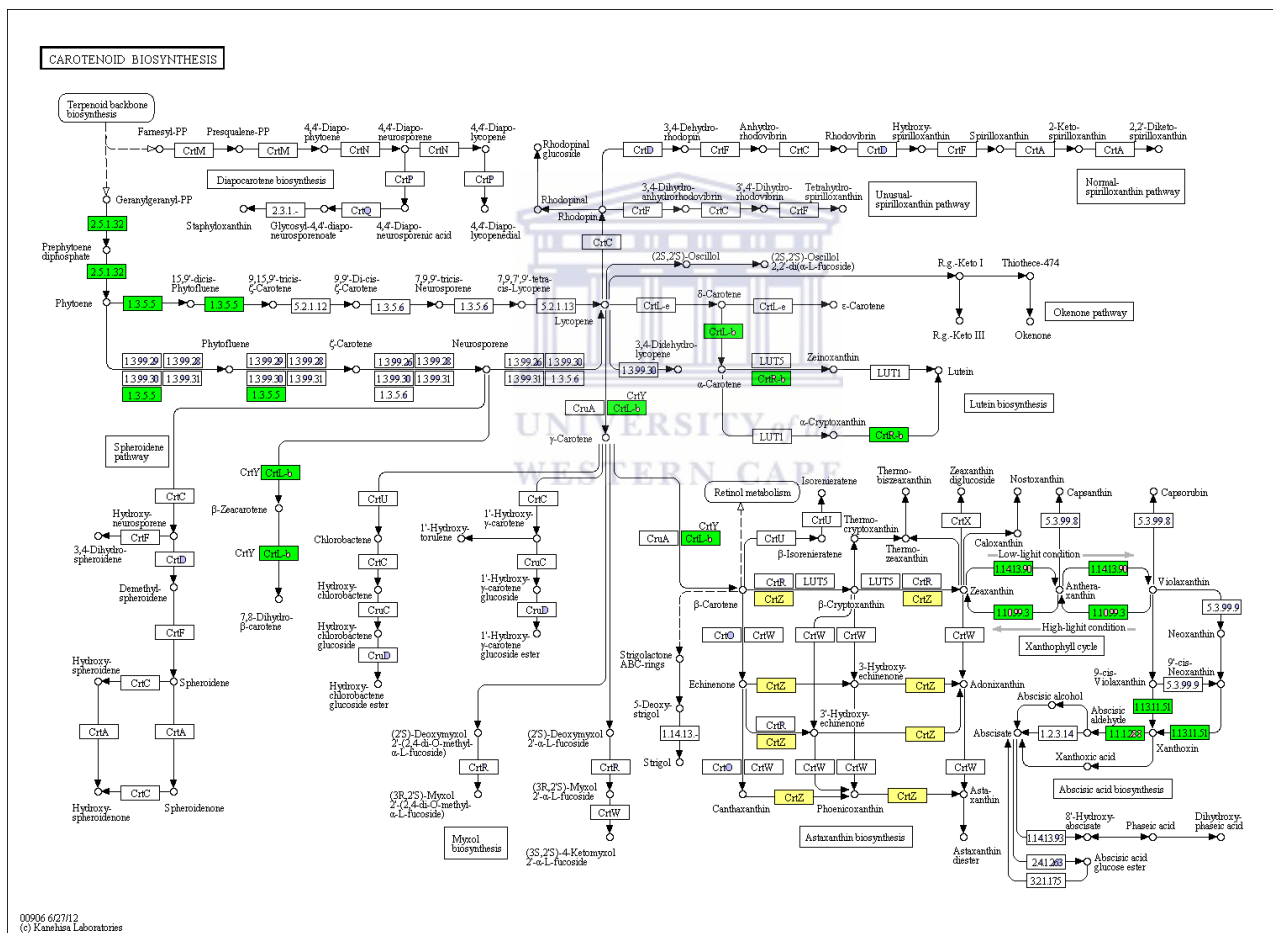
Figure 6. 16 Localisation of carotenoid biosynthesis pathway enzymes enriched in peel and pulp apple fruit tissues. The Figure shows a carotenoid KEGG pathway reference map. The Figure shows pulp and peel tissues to be enriched carotenoid pathway enzymes. The peel tissue also is more enriched in carotenoid biosynthesis enzymes. Boxes coloured green and yellow represent sub pathway genes enriched in 'Golden Delicious' fruit peel and in both peel and pulp tissues respectively. Names of tissue-enriched enzymes are written close to the EC number of the enzyme.

The expression of crtZ in the peel suggested possible accumulation of astaxanthin, adonixanthin, zeaxanthin, zeinoxanthin and lutin (Figure 6.16). In a xanthophyll cycle reaction, zeaxanthin is converted to violaxanthin through antheraxanthin by zeaxanthin epoxidase (ZEP, ABA1, EC 1.14.13.90) under low light conditions. However, under high light conditions violaxanthin is converted back to zeaxanthin through antheraxanthin, a reaction catalysed by violaxanthin de-epoxidase (VDE, NPQ1, EC 1.10.99.3)(Jahns and Holzwarth, 2012). In peel, *ZEP* and *VDE* had similar expression profile, but *ZEP* was more expressed than *VDE*, suggesting net accumulation of violaxanthin throughout fruit development, except at daa87 to daa135. At daa87 to daa135 *VDE* was more expressed than ZEP, suggesting net accumulation of zeaxanthin (Figure 6.17A, B). The net accumulation of zeaxanthin and lutin were in concert with the non-photochemical quenching (NPQ) of excess light energy in the antenna of photosystem II (PSII) (Jahns and Holzwarth, 2012). Lutin function in the deactivation of $^{3}Chl^{*}$, while zeaxanthin deactivates excited singlet Chl ($^{1}Chl^{*}$). Additionally, zeaxanthin has been reported to function as an antioxidant in the lipid phase of the membrane and as a key component in the memory of the chloroplast with respect to preceding photo-oxidative stress (Jahns and Holzwarth, 2012). It was logical then that genes encoding lutin and zeaxanthin formation were enriched in the peel, where they interphase with the external environment.

CRTZ expression levels in the peel decreased with fruit development with lowest expression at daa87 (Figure 6.17A, B). The drop in crtZ expression coincided with a decrease in ZEP expression levels, suggesting the dependence of ZEP on crtZ product formation rates. Pulp crtZ was however lowly expressed compared to peel expressed crtZ

(Figure 6.17B). Nonetheless, unlike peel crtZ, pulp crtZ increased in expression with fruit development suggesting possible accumulation of zeaxanthin, astaxanthin, zeinoxanthin, adonixanthin and lutin.

NCED (9-cis-epoxycarotenoid dioxygenase [EC 1.13.11.51]) catalyses the rate limiting step in ABA biosynthesis, and was enriched in the peel tissue. Peel enriched genes encoding NCED had similar expression pattern to ZEP and VDE, suggesting NCED activity might relay on the combined activities of ZEP and VDE.

The enrichment of lycopene beta-cyclase (lcyB, crtL1, crtY, EC 5.5.1.19) in the peel suggested the accumulation of ß-carotene, α-carotene, and 7,8 dihydro-ß-carotene (Figure 6.17A). However, the downstream enrichment of xanthophyll cycle, astaxanthin biosynthesis and ABA biosynthesis genes might suggest a progressive accumulation of ß-carotene with fruit development.

Genes encoding phytoene synthase (crtB, EC 2.5.1.32) had peak expression at daa60, while those encoding 15-cis-phytoene desaturase (PDS, crtP, EC 1.3.5.5) were lowly expressed and showed a slight increase from daa135 to daa150 (Figure 6.17A). The CRT-B enzyme catalyses the first and second committed reaction steps in the carotenoid biosynthesis pathway (Sharpe and DiCosimo, 2012) suggesting peak accumulation of phytoene at about daa60. However the low expression levels of PDS, suggested that PDS catalysed reaction might be the rate-limiting step in carotenoid biosynthesis. It should be noted that some enzymes compensate for low expression levels by increasing reaction

activities. As such the low expression level of PDS was not enough to conclusively characterize the PDS catalysed reaction step as the rate-limiting step.
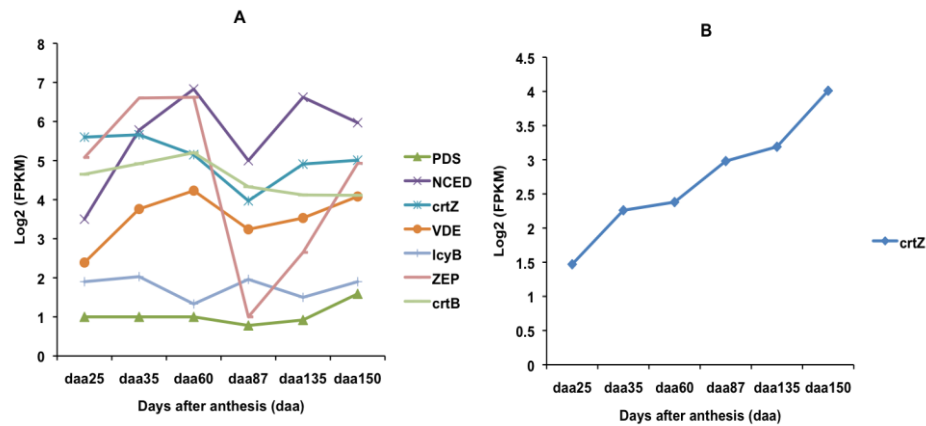
Figure 6. 17 Expression patterns of peel and pulp up regulated transcripts encoding carotenoid biosynthesis pathway enzymes. The Figure shows expression profiles of peel (A) and pulp (B) greater than two-fold up regulated transcripts encoding carotenoid biosynthesis enzymes. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.6 Carbon assimilation

6.2.6.1 Carbon fixation

The C_4 and C_3 pathways have been extensively elucidated in the leaf, and not in fruit (Aubry, *et al.*, 2011; Covshoff and Hibberd, 2012). Nonetheless the knowledge gained from leaf carbon fixation could be extrapolated to fruit. The initial CO_2 fixation to oxaloacetate by phosphoenolpyruvate carboxylase (PPC, EC 4.1.1.31), circumvents the direct use of the less efficient Rubisco CO_2-fixing enzyme (Shay and Kubien, 2012). Enrichment of PPC, aspartate aminotransferase, mitochondrial (GOT2, EC 2.6.1.1), malate dehydrogenase (MDH1, EC 1.1.1.37), malate dehydrogenase (decarboxylating) (EC 1.1.1.39) and alanine transaminase (GPT, ALT, EC 2.6.1.2) in the pulp suggested that CO_2-fixation in apple fruit pulp might take place through the C_4 pathway. This

245

observation was reasonable considering pulp and not peel tissue is permeated by vascular bundles, which even at physiological maturity still looked green. The green appearance suggested that the cells surrounding the vascular bundles (bundle sheath cells) contained chloroplasts. In the pulp mesophyll cells, CO_2 is fixed by PPC forming oxaloacetate, which is converted to aspartate and subsequently malate by the action of GOT2, MDH1, and EC 1.1.1.39 enzymes.

Malate is transported to the bundles sheath cells, where it is decarboxylated to produce CO_2, which enters the Calvin Cycle and pyruvate is transported back to mesophyll cells by glutamate-glyoxylate aminotransferase (GGAT, EC 2.6.1.4 2.6.1.2 2.6.1.44 2.6.1.-) to reforming phosphoenolpyruvate (Shay and Kubien, 2012). The oxaloacetate-aspartate root enrichment was plausible considering that genes in this sub-pathway were enriched in pulp, and also that pulp mesophyll cells are limited in chloroplast content.

The enrichment of only the Calvin cycle genes in the peel, however, suggested that in the peel tissue CO_2 is directly fixed to glycerate-3-phosphate by Rubisco, encoded for by ribulose-bisphosphate carboxylase small chain (rbcS, EC 4.1.1.39). The Calvin cycle enzymes were encoded for by peel genes, which had peak expression between daa35 to daa87, a period correlating to high fruit photosynthetic capacity (Fleancu, 2007). A more accurate period of CO_2-fixation in the peel might be provided by the expression profile of rbcS, which had peak expression at daa35 and a lower expression thereafter. Peak rbcS expression at daa35 just before peak starch accumulation, was logical, and suggested that

rbcS peak expression precedes starch accumulation. Its expression therefore might be an indicator of the level of CO_2-fixation.
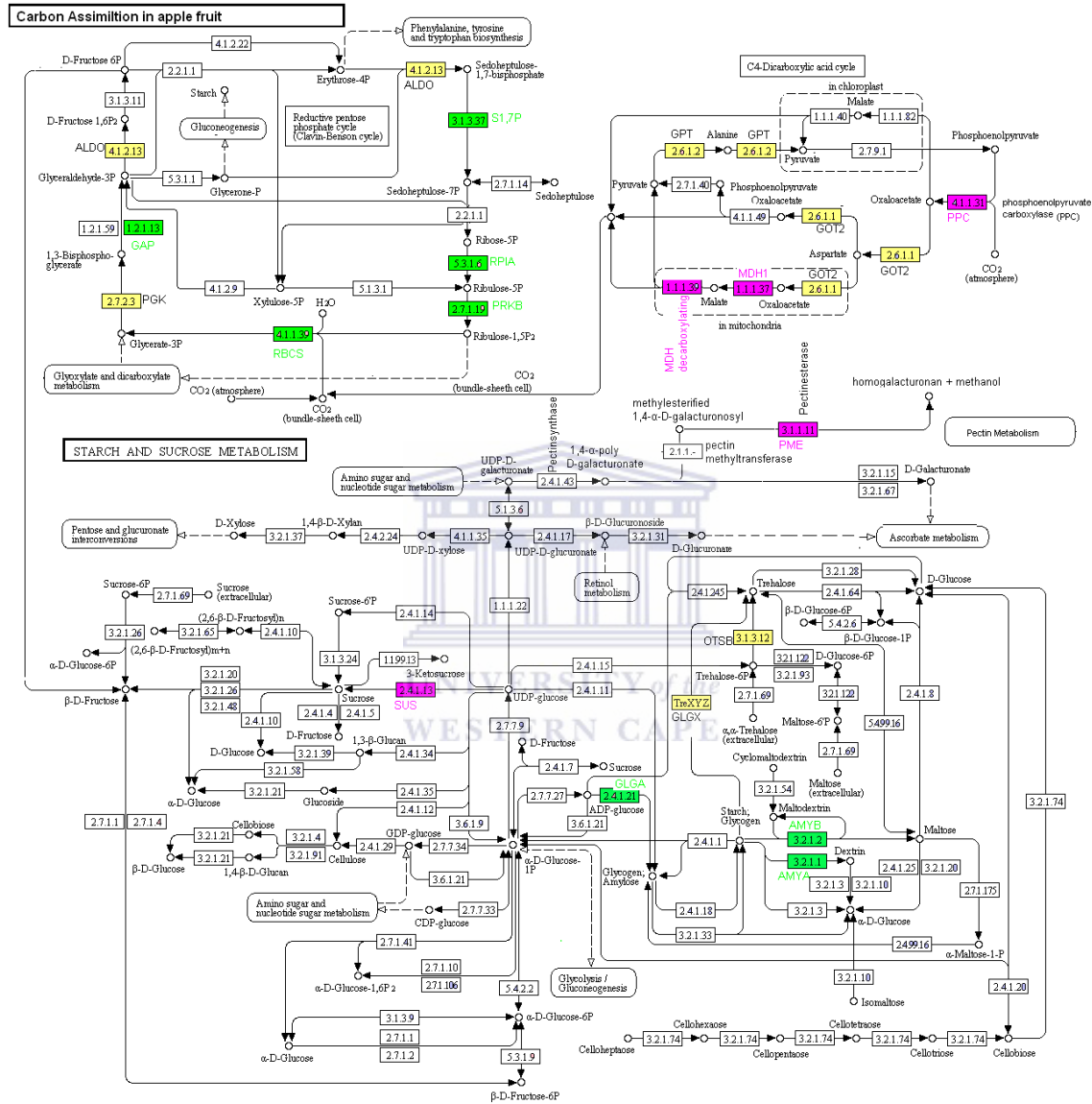
**Figure 6. 18 Localisation of carbon assimilation pathways enzymes enriched in peel and pulp apple fruit tissues.** The Figure shows a modified combined carbon fixation, starch and sucrose metabolism KEGG pathway reference map. Boxes coloured pink, green and yellow represent sub pathway genes enriched in 'Golden Delicious' fruit pulp, peel and in both tissues (peel and pulp) respectively. Names of tissue-enriched enzymes are written close to the EC number of the enzyme.

Pulp genes encoding Calvin cycle enzymes decreased in expression at daa60 but increased the remainder of fruit development period (Figure 6.19A). Interestingly, the

observation was an inverse of the period of peak starch accumulation. The result suggested that as CO_2 was fixed and converted to starch, the pulp-carbon sink strength for fixing more CO_2 decreased and was lowest during peak starch accumulation. However, as the fruit entered the climacteric phase, starch was degraded to augment the energy reserves, hence reducing starch levels. The decrease in starch might then have renewed the carbon-sink strength of the pulp, thereby increasing CO_2-fixation, as corroborated by the expression profiles of the Calvin cycle genes (Figure 6.19A).
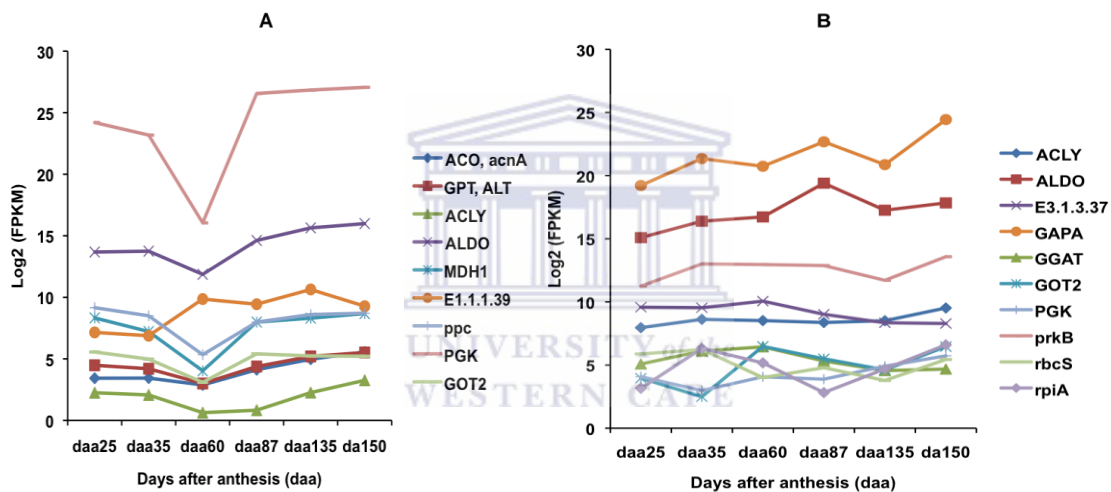
Figure 6. 19 Expression profiles of transcripts encoding the Calvin cycle and C$_4$ carbon fixation pathway enzymes. Expression patterns of pulp (A) and peel (B) greater than two-fold up regulated transcripts encoding dark reaction enzymes of photosynthesis. Expression values (FPKM) were log2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

What was however interesting, was the observed possible dual carbon fixation mechanisms in apple fruit tissues. The apple plant is known to be a C_3 plant, and C_3 activity was corroborated by the expression of C_3 transcripts in the peel tissue as was expected. However, C_4 carbon fixation has never been reported in apple, let alone the existence of dual C_3 and C_4 mechanisms in the same plant but different tissues. In a study

on the physiology of ex-vitro pineapple under crassulacean acid metabolism (CAM) and C_3 carbon fixing promoting conditions, Aragón, *et al,* (2012) observed that the carbon fixing mechanism could switch from CAM to C_3 depending on environmental conditions (Aragóń, *et al.*, 2012). Thus plastic morphology and physiology is indeed possible, but for a plant to exhibit tissue difference in carbon fixation mechanism in response to tissue micro-conditions has never been reported. The enrichment of the C_4 pathway genes in the pulp might therefore be testimony to a new discovery of the existence of a functional C_4 carbon-fixing pathway in addition to the C_3 in apple fruit. However, Aubry, *et al*. (2011) reviewed the role of C_3 proteins prior to recruitment into C_4 pathway, and it was noted that enzymes required for C_4 pathway are present in C_3 plants, but change in function or rate of activity after recruitment into C_4 pathway (Aubry, *et al.*, 2011). For example, in a normal C_3 plant PPC function to supply carbon skeletons to the tricarboxylic acid (TCA) cycle, operating in malate homeostasis during drought stress, supplying carbon skeletons to allow ammonium assimilation, and regulating stomatal conductance (Aubry, *et al.*, 2011; Covshoff and Hibberd, 2012). The increase in PPC could result in malate accumulation in pulp cells. The accumulation of malate in pulp cell acts as an osmotica increasing turgor required for cell elongation (Aubry, *et al.*, 2011). Nonetheless, the increased expression of the C_4 pathway genes in the pulp might have been in response to increased respiration in the pulp as indicated by high enrichment of oxidative phosphorylation pathway encoded enzymes. In order to salvage the accumulating CO_2 in the pulp, the expression levels of PPC might have increased accordingly to fix the respired CO_2.This theory was corroborated by the increase in PPC expression with increase in oxidative phosphorylation pathway genes as the fruit matured. In seedpods of

rice, $^{14}CO_2$-labeling experiments showed that PPC participated in the fixation of leaf respired CO_2. The pome fruit is much more complex than a leaf, with possibly much more rates of respiration, hence the pulp tissue might have modified its physiological processes to address the challenges in its microenvironment.

6.2.6.2 Starch and sucrose metabolism

The expression levels of alpha-amylase (amyA, malS, EC 3.2.1.1) encoding peel transcripts had a transitory peak at daa60 followed by rapid increase in expression from daa135 to daa150 (Figure 6.18, 20A). Peel transcripts encoding ß-amylase (EC 3.2.1.2), on the other hand were highly expressed and increased in expression progressively with fruit development (Figure 6.20A), a trend similar to glycogen operon protein (treX, glgX, EC 3.2.1.-). ß-amylase (AMYB, EC 3.2.1.2), and amyA are starch-degrading enzymes, which form maltose and dextrin products respectively. The high and progressive increase in ß-amylase (AMYB, EC 3.2.1.2), encoding transcripts expression levels might suggest accumulation of maltose sugar with apple fruit development. Sucrose synthase (SUS, EC 2.4.1.13) encoding genes were up regulated in the pulp tissue (Figure 6.18, 6.20A). The progressive increase in ß-amylase, amyA, glgX and SUS suggested an increase in total soluble solids with fruit development, corroborating the increased Brix˚ score with fruit development (Figure 4.1). Peak expression of starch synthase (glgA, EC 2.4.1.21) and SUS encoding transcripts at daa60 suggested that starch accumulation might have reached maximum accumulation about daa60. These findings were in concert with the period of peak starch accumulation as observed by Janseen, *et al*, (2008). The enrichment of glgX and trehalose 6-phosphate phosphatase (otsB, EC 3.1.3.12) in the apple peel

tissue suggested a possible accumulation of trehalose, through starch degradation and trehalose-6P dephosphorylation respectively (Figure 6.18, 20A, B). Trehalose is a non-reducing sugar, which like sorbitol and mannitol is a suitable compatible solute ideal for a plant cell role as an osmotica, improving desiccation tolerance (Ribaut, 2006).

The accumulation of trehalose in apple fruit tissues might therefore have functioned in improving fruit desiccation tolerance, considering that with fruit development stomata plasticity decreased due to suberisation. Desiccation protection possibly is not tissue or cell specific, since each fruit cell might be subject to desiccation, hence the observed enrichment of trehalose synthesis genes in both peel and pulp tissues of apple fruit. In previous attempts to improving drought tolerance, the transgenic plants only improved in desiccation tolerance but not plant water use efficiency (Ribaut, 2006). Targeting fruit or fruit tissue osmotica accumulation, therefore might only protect the fruit from possible water loss shock, but not optimize on its use of the available water.

Pectinases break down the cell wall cement, pectin to pectate (Costa*, et al.*, 2010). This pectin degradation might be desirable depending on the stage of fruit development. During fruit maturation fruit cells divide and expand, and the loosening of the cell walls by pectinases assist the process (Costa*, et al.*, 2010). However, degradation of pectin after physiological maturity may not be desired, as it result in low cell wall integrity and hence fruit firmness. A less firm fruit is susceptible to pathogen attack and reduced shelf life (Marondedze and Thomas, 2011, 2012). The enrichment of pectinesterase (EC 3.1.1.11) in the pulp tissue was in accordance with the actively dividing and expanding process

251

which took place in the pulp during fruit development (Figure 6.18, 20B). The pectinesterase (EC 3.1.1.11) enrichment in the pulp might also suggest that the pulp tissue might be the first to succumb to loss of cell wall integrity, hence affecting fruit firmness.
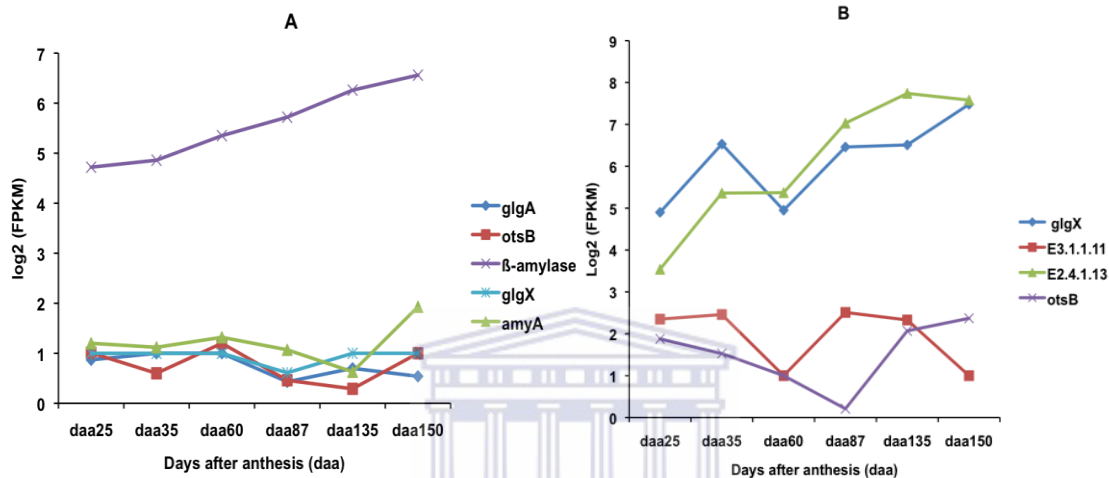
Figure 6. 20 Starch and sucrose biosynthesis pathway enzymes encoding transcripts expression profiles. A: Expression trend of apple fruit peel up regulated transcripts encoding starch synthesis and degradation, B: Expression trend of pulp tissue transcripts encoding starch synthesis, sucrose biosynthesis and pectinases. Expression values (FPKM) were log_2 transformed to normalize the data (section 2.2.17). Enrichment of an enzyme was calculated as the frequency of KO (KEGG Ontology) term annotations for each transcript (section 2.2.17).

6.2.7 Terpenoid backbone biosynthesis

Apple fruits produce an array of volatile compounds presumably to discourage premature seed dispersal and to provide a taste reward for seed dispersers after physiological maturity. Depending on the developmental fruit stage, the apple fruit produces volatiles, which act as signals to the internal and external environment. These volatiles include alcohols, aldehydes, esters, terpenes, terpenoids, and polyphenols derived from secondary metabolite pathways. Terpenoids are produced from a pool of isopentenyl diphosphate (IPP) synthesized form both the cytosolic mevalonate (MVA) and the plastidic 2-C-

252

methyl-d-erythritol 4-phosphate (MEP) pathways (Jomaa*, et al.*, 2011; León and Cordoba, 2012; Yu*, et al.*, 2012).

Genes encoding MVA and MEP pathway enzymes were enriched in both apple fruit peel and pulp tissues (Figure 6.21, 22A). However, enrichment of genes encoding enzymes up stream of the isopentenyl-diphosphate delta-isomerase (IDI, EC:5.3.3.2) in the MVA pathway, suggested that the MVA pathway was more represented, more so in peel than pulp (Figure 6.21). Both peel and pulp up regulated genes were enriched for 1-deoxy-D-xylulose-5-phosphate synthase (dxs, EC 2.2.1.7), which catalyses the rate limiting step of MEP pathway (León and Cordoba, 2012) (Figure 21, 22A). The expression of DXS in both peel and pulp tissues suggested that both MVA and MEP pathways were enriched and contributed to the pool of the terpene IPP in 'Golden Delicious' fruit.Enrichment of the IDI and farnesyl diphosphate synthase (FDPS, EC:2.5.1.1; 2.5.1.10) in peel suggested the balancing off of the IPP pool in the cytosol and plastids depending on the needs of the peel tissue depending on the conditions in the developing fruit.

In the peel tissue, genes encoding geranylgeranyl reductase (chlP, EC:1.3.1.83) were enriched (Figure 21, 22B).
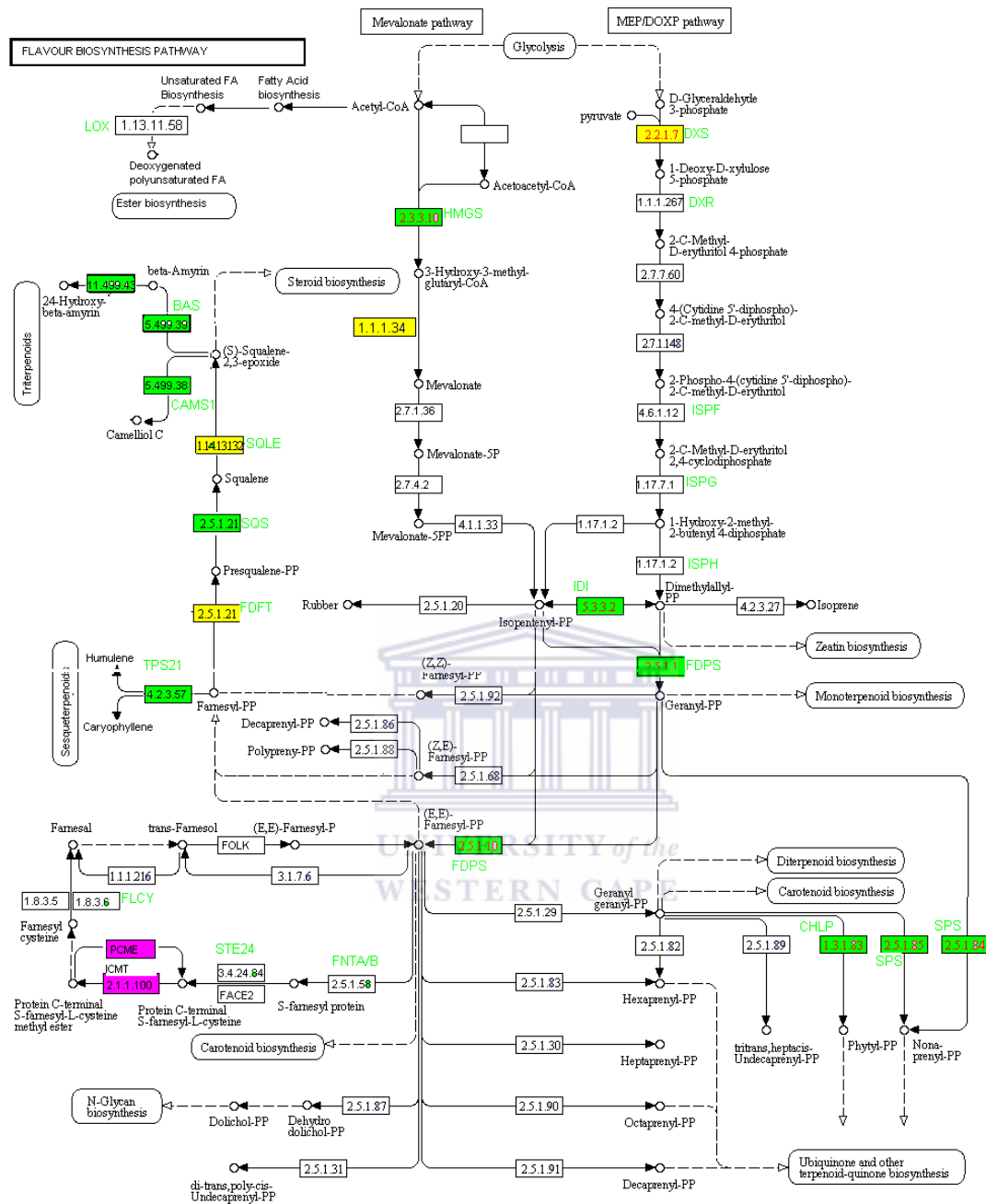
Figure 6. 21 Localisation of flavour forming pathway enzymes enriched in peel and pulp apple fruit tissues. The Figure shows a modified combined terpenoid, sesquiterpenoid and ester metabolism KEGG pathway reference map. Boxes coloured pink, green and yellow represent sub pathway genes enriched in 'Golden Delicious' fruit pulp, peel and in both tissues (peel and pulp) respectively. Names of tissue-enriched enzymes are written close to the EC number of the enzyme.

The chlP enzyme catalyse the formation of phytyl diphosphate, a precursor for vitamin E, phylloquinol, linoleate and chlorophyll b biosynthesis (Lippold, *et al.*, 2012). Genes

encoding all-trans-nonaprenyl-diphosphate synthase (SPS, EC:2.5.1.84; 2.5.1.85) enzyme were also enriched in the peel tissue (Figure 21, 22A). SPS catalyses the formation of all-trans-nonaprenyl diphosphate to a maximal unit length determined by the specific SPS enzyme (de Oliveira, *et al.*, 2012). The complete all-trans-nonaprenyl diphosphate unit is then incorporated into other molecules, e.g. quinone forming plastoquinone (de Oliveira, *et al.*, 2012).

Genes encoding prenylcysteine alpha-carboxyl methylesterase (PCME, EC:3.1.1.-) and protein-S-isoprenylcysteine O-methyltransferase (ICMT, STE14, EC:2.1.1.100) were enriched in the pulp tissue (Figure 21, 22A). PCME and ICMT enzymes and are involved in the protein post-translational modification with lipophilic molecules to aid in anchoring the protein to the phospholipid layer. The enrichment of ICMT and PMCE in the pulp tissue was in concert with the observed enrichment of the protein processing in the endoplasmic reticulum pathway (Figure 6.1), suggesting protein post-translational modification was higher in pulp than peel tissue. Protein prenylation is important for localization and function of many membrane-bound proteins (Hrycyna, *et al.*, 2011)
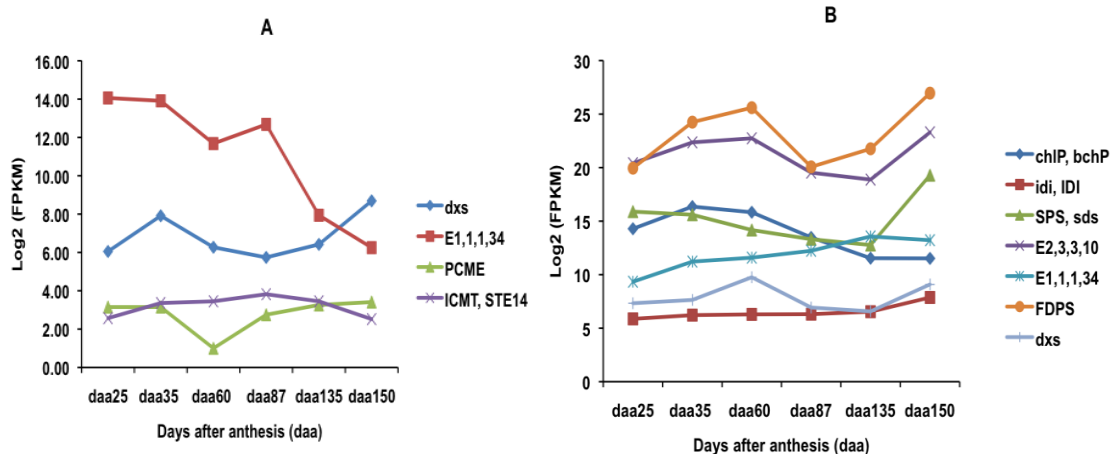
Figure 6. 22 Developmental expression profiles of terpenoid backbone biosynthesis genes encoded for by transcripts with more than two-fold up regulation. A: Expression trends of transcripts up regulated in apple pulp tissue encoding terpenoid backbone biosynthesis enzymes of the MEP, MVA and protein post-translational modification pathways; B: Peel enriched transcripts encoding MVA, MEP and formation of ubiquinone precursors. Expression values (FPKM) were $\log_2$ transformed to normalize the data (section 2.2.17).

6.2.7.1 Sesquiterpenoid biosynthesis

The enrichment of farnesyl-diphosphate farnesyltransferase (FDFT1, EC:2,5,1,21) and squalene monooxygenase (SMO, SQLE, ERG1, EC:1,14,13,132) in pulp suggested a possible accumulation of squalene and (S)-squalene 2,3-epoxides, precursors of triterpenoid, in the pulp. In the peel tissue, sesquiterpenoids (ß-caryophyllene, α--humulene) produced by alpha-humulene/beta-caryophyllene synthase (TPS21, EC:4,2,3,-4,2,3,57) and triterpenoids (ß-amyrin and camelliol C) produced by ß-amyrin synthase (BAS/LUP4, EC:5,4,99,39) and camelliol C synthase (CAMS1, LUP3, EC:5,4,99,38) respectively were produced. It is possible the terpenoids produced in the pulp might be transported to the peel tissue for further processing before release. Peel enriched genes encoding both sesquiterpenoids and triterpenoids biosynthesis pathway enzymes had transitory peak expression at daa60 followed by an increase from daa87 to daa150 (Figure 6.23B). However, pulp genes encoding FDFT1 had peak expression at daa60 but decreased afterwards (Figure 6.23A). SQLE encoding pulp genes on the other had an initial high expression, which decreased to daa60 followed by peak expression at daa87 to daa135 (Figure 6.23A). The high expression levels of genes encoding ß-amyrin synthase, which catalyses the formation of the triterpenoid ß-amyrin suggest that ß-amyrin might be the major triterpenoid in 'Golden Delicious' apple fruit with transitory expression at daa60 and an increase in expression with fruit maturity (Figure 6.23B).
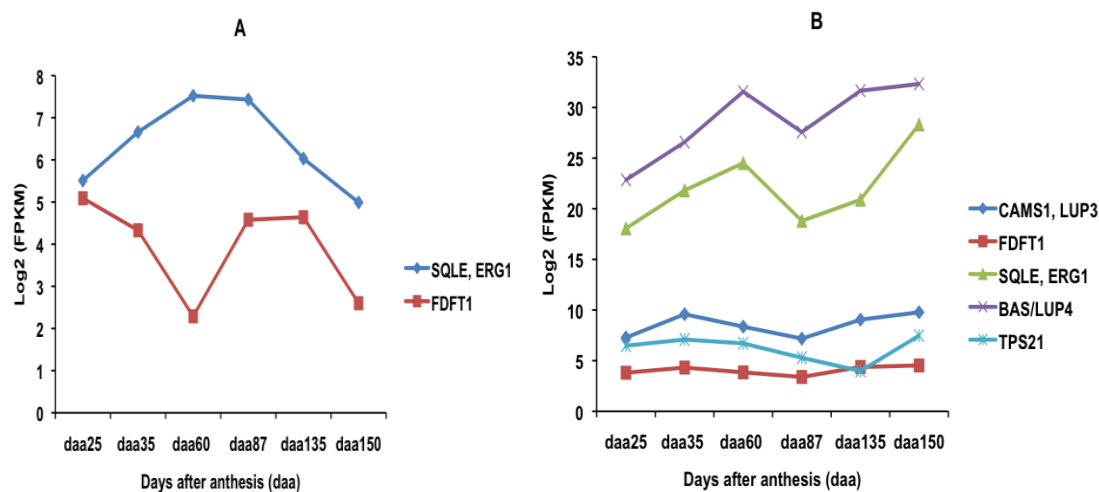
Figure 6. 23 Developmental expression profiles of transcripts encoding sesquiterpenoid and triterpenoid biosynthesis pathway enzymes. A: Expression trends of transcripts up regulated in pulp encoding triterpenoid biosynthesis enzymes; B: Expression trends of transcripts up regulated in peel tissue encoding sesquiterpenoid biosynthesis pathway enzymes, which further process triterpenoids into sesquiterpenoids. Expression values were $\log_2$ transformed to normalise the data (section 2.2.17).

The sesquiterpenoids α-farnesene (E,E and Z,E), ß-caryophyllene, and germacrene D were reported to be produced in developing apple fruit (Schaffer, *et al.*, 2007). However, in this study genes encoding biosynthesis of the sesquiterpenoids (ß-caryophyllene and α-humulene) and triterpenoids (ß-amyrin and camelliol C) to be enriched in 'Golden Delicious' apple fruit tissues were identified. The gene *TPS21,* despite coding for alpha-humulene/beta-caryophyllene synthase (TPS21, EC:4,2,3,- 4,2,3,57) also codes for an inactive (E,E)-alpha-farnesene synthase, suggesting TPS21 may also be involved in farnesene biosynthesis. The accumulation of triterpenoids and sesquiterpenoids is invaluable to the fruit, where it functions as an internal and external messenger as well as regulating seed dispersal and defence. The accumulation of farnesene in enclosed storage systems causes development of apple scald. Apple fruit scald is a tissue disorder arising as a result of prolonged and uncontrolled atmosphere apple fruit storage conditions.

257

Farnesene when oxidized produces trienes, which accumulate on the upper surfaces of the fruit. Trienes, however, are toxic to plant cells and causes the death of epidermal cells resulting in unevenness of fruit colour and texture. Scald apple fruit disorder can however be controlled by fruit immersion in squalene and farnesene, as well as controlling the storage atmosphere to reduce farnesene accumulation.

Though this study reports terpenoid-based compounds as contributors to flavour, it should be noted that terpenoid-based compounds are indeed not the only flavour contributing substances. On the contrary, esters were reported to be the major compounds produced in ripe apple fruit. Several studies have also corroborated the contribution of straight chain and branched c-chain esters to flavour (Schaffer, *et al.*, 2007; Soukoulis, *et al.*, 2012).

6.3 Conclusion

Spatial and temporal differential gene expressions in 'Golden Delicious' fruit define the properties of peel and pulp tissues. In the previous chapters mRNA profiling of pulp and peel tissues was done. However, a comparison of the transcripts up regulated between the peel and pulp tissue provides more insight into the physiological processes up regulated or exclusively activated in a particular tissue. Peel tissue up regulated transcripts were enriched for plant pathogen interaction, terpenoid, carbon fixation, fatty acid metabolism, photosynthesis, carotenoid and flavonoid metabolic pathways. On the other hand protein processing, oxidative phosphorylation and plant hormone signal transduction were more enriched in fruit pulp. Photosynthesis pathway related transcripts were only expressed in

the peel and had peak expression at daa35 to daa60, suggesting that 'Golden Delicious' fruit's photosynthetic capacity reached maximum between daa35 and daa60 and gradually decreased with fruit maturity.

Pulp tissue was more enriched for oxidative phosphorylation pathway genes, predominantly V-type and F-type ATPases, with lowest expression at daa60, possibly a period of lowest energy demand. V-type ATPases enrichment in the pulp could be for regulating cellular solute content, hence turgidity and optimum pH balance for optimum physiological cell functioning. The peel tissue on the other hand expressed AHA2, a functionally redundant equivalent of V-type ATPases.

Peel genes encoding fatty acid biosynthesis pathway enzymes had peak expression at daa60, suggesting that at daa60 the formation of the precursors for cutin, suberin and wax had reached their maximum. Low expression of fatty acid biosynthesis genes in the pulp suggested possible contribution of the pulp tissue towards epicuticular wax formation.

Auxin and cytokinin signal transduction pathway genes were enriched in both peel and pulp tissues, but more so in pulp, suggesting auxin and cytokinin-auto-regulation in these tissues. The accumulation of BZR1 /2 in the pulp might have functioned to regulate nucleus gene expression and plant development by promoting cell division and enlargement. ABA signal transduction pathway genes were more enriched in the pulp with some increasing with fruit development, possibly in response to increased stress due to increased oxidative phosphorylation. In the peel, ABF enrichment might have

259

functioned in stomata control. The enrichment of the ethylene signal transduction gene MPK6 in the pulp did not show a clear progressive increase with fruit development. The enrichment of DELLA in the peel tissue suggested that the peel tissue contained low GA, and that DELLA inhibited GA signal transduction. Enrichment of GID/SLY1 complex in the pulp tissue suggested the existence of GID/SLY1-mediated DELLA protein degradation and hence GA signal transduction in the pulp. Inactivation of GA signal by DELLA in the peel and activation of the GA signal in the pulp, corroborated pulp as an actively dividing and expanding tissue, accounting for most of the fruit increases in size.

Flavonoid biosynthesis pathway genes were more enriched in the peel, possibly to function as screen against UV-light and as antioxidants. The peel tissue was more enriched for carotenoid biosynthetic pathway genes, suggesting the peel to accumulate more carotenoids with fruit development explaining the transition to golden yellow colouration with fruit development.

The enrichment of the C_4-pathway genes in the pulp might be testimony to the existence of a functional C_4-carbon fixing pathway in apple fruit. The enrichment of the Calvin cycle genes in the peel suggests direct CO_2 fixation to glycerate 3-phosphate in the peel. The progressive increase in ß-amylase, amyA, glgX and sucrose synthase suggested an increase in total soluble solids with fruit development. Peak glgA expression at daa60 suggested maximum starch accumulation to be at daa60.

Aroma compounds synthesis terpenoid and sesquiterpenoid pathway genes were more enriched in the peel. The pulp tissue showed enrichment of protein prenylation genes suggesting post-translational protein modification to aid in protein localization and function.

The study has thus provided an in depth catalogue and elucidation of the pathways enriched in peel and pulp 'Golden Delicious' tissues, which might function in defining the properties of the tissues.

7   CHAPTER 7

7.0 OVERALL DISCUSSION

The apple fruit is a unique aspect of plant development and is a part of human diet. For these reasons, a molecular study was undertaken to elucidate mechanisms underlying the morpho-physiological transitions in developing 'Golden Delicious' fruit towards unravelling the complex interactions leading to fruit quality. A number of previous studies partially elucidated apple fruit cortex, whole fruit, leaves, and peel plus cortex transcriptomes using cDNA micro arrays and other PCR based technologies (Han, Bendik*, et al.*, 2007; Hattasch*, et al.*, 2008; He*, et al.*, 2008; Janssen*, et al.*, 2008; Mintz-Oron*, et al.*, 2008; Soglio*, et al.*, 2009). These studies, however, remain limited in throughput and specificity in revealing transcripts of low abundance. Hence, the need for the application of a high throughput technique to elucidate the full mRNA transcriptome of the 'Golden Delicious' fruit peel and pulp tissues.

7.1 Full-length transcript reconstruction

The aim of this experiment was to reconstruct full-length transcripts from RNA-seq data pooled from peel and pulp 'Golden Delicious' fruit tissues harvested at 25, 35, 60, 87, 135, and 150 days after anthesis (daa). Pooling of RNA-seq data from each time point and tissue type was essential to increasing coverage of the lowly expressed genes, hence their chances of full-length reconstruction (Grabherr*, et al.*, 2011; Trapnell*, et al.*, 2012). The gapped nature of the existing *M. x domestica* genome v1.0 (Velasco*, et al.*, 2010)

meant that some reads would not map to the reference genome. As such, to optimise on the transcripts that could be fully reconstructed *de novo* (Trinity) and genome-guided (Cufflinks) assemblers were used (Trapnell*, et al.*, 2012). The choice of these two algorithms was based on the recent study by Grabherr, *et al*, (2011), who ranked Trinity as the best *de novo* full-length transcript reconstruction algorithm, Trapnell *et al*, (2012), reported, Cufflinks, as one of the best genome-guided transcript assemblers.

A comparison of the accuracy of the transcripts reconstructed by genome-guided (Cufflinks) and *de novo* (Trinity), which mapped to the genome, revealed that Cufflinks reconstructed transcripts had higher count and breadth coverage of *M. x domestica* genome v1.0 gene models and ESTs gene set compared to Trinity reconstructed transcripts. The higher count in genome-guided transcripts could have arisen as a result of more isoforms being reconstructed compared to Trinity, where transcripts with more than 95% identity were collapsed (Grabherr*, et al.*, 2011). Both approaches extended a total of 25155 gene models in both the 3´and 5´ends further improving the annotation of the apple genome.

After genome-guided transcript assembly a substantial number of reads remained unmapped. These reference genome unmapped reads were assembled into transcripts using Trinity and anchored to the unanchored contigs of the *M. x domestica* genome v1.0. The study was done to observe how much gene space resided in the unanchored contigs. As expected a substantial gene space resided in the unanchored contigs of the v1.0 apple genome. What was even worrying was the fact that an even bigger number of the

reconstructed transcripts still could not map to either the genome or the unanchored contigs. These unanchored transcripts possibly constitute a yet to be sequenced apple genome gene space. It should be noted however, that some of the unmapped transcripts could be artefacts, meaning they are not representatives of any coding gene. It was interesting to note that despite so much gene discovery in apple fruit in the previous studies (Troggio*, et al.*, 2012), a substantial gene number in this study remained unknown.

Annotation of the combined (genome-guided and reference genome-unmapped *de novo* assembled) transcripts revealed that 29% of the combined peel and pulp transcriptome had no hits to NCBI and Swiss-prot databases, and 21% had unclassified function based on the MIPS Arabidopsis best hits functional categories (Mewes*, et al.*, 2008). Among the transcripts with known function, transcripts encoding proteins with binding function were predominant. It was evident that MIPS functional categories based on the *Arabidopsis* categories had limitations, hence the further annotation of the 'No-hit' transcripts using domain search. "No-hit" transcripts showed annotation of 98 Pfam domains, which composed mostly of leucine rich repeat, WD domain-G beta repeat and teratricopeptide repeat domains.
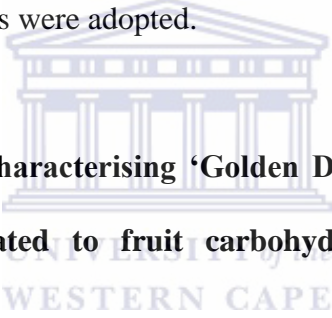
The existence of chloroplast DNA (cpDNA), mitochondria DNA (mtDNA) and nuclear DNA (nucDNA) genomes in plants is unique but suggest the existence of complex interaction of these three genomes (Hancock, 2012; Seaver*, et al.*, 2012; Wendel*, et al.*, 2012). Some of the reconstructed transcripts had homology to sequence regions on the

264

cpDNA and mtDNA. This evidence of homology suggests the presence of intricate regulatory networks among nucDNA, cpDNA and mtDNA for possibly specialised functions such as photosynthesis and respiration (Seaver, *et al.*, 2012) respectively. The existence of fragmented full-length genes of cpDNA or mtDNA origin is evident of the high rates of horizontal gene transfer (HGT).

Development of tools for transcript abundance estimation has grown in leaps and bounds over the previous years. The advent of high-throughput RNA-seq approaches increased sensitivity and range of transcript detection and abundance estimation (Trapnell, *et al.*, 2012). Since RNA-seq is still at its infancy, the adopted approach in RNA-seq studies is to validate the expression levels with those obtained by RT-qPCR (Fox, *et al.*, 2012). However, considering the limited range of detection when using RT-qPCR as well as the absence of 'ideal' reference gene for the validation of RNA-seq to RT-qPCR might only be used to observe the general transcript expression trend rather than precise abundance estimation. Nonetheless, a positive correlation of RSEM transcripts abundance estimates to RT-qPCR across the six time points, only when reference gene geometric mean was used, was observed. The RT-qPCR estimates could be increased in sensitivity by using more than three reference genes for normalising of the expression levels. The limitation however would be the cost of using multiple reference genes for a small target gene set.

In the previous protocols for library preparation the library enrichment step was characterised by multiple PCR cycles, which in most cases resulted in accumulation of significant PCR duplicates. Failure to significantly enrich for mRNA also resulted in

libraries with significant amount of rRNA. PCR duplicates and rRNA contamination both reduce the effective library concentration, hence the effective mRNA sequenced read set. In this study ploy-A enriched mRNA was DNase treated and amplified for only ten cycles. However after transcript reconstruction ribosomal RNA transcripts were observed, prompting *in-silico* rRNA treatment. The effect of rRNA treatment, read de-duplication as well as read sub-setting on transcript abundance estimation was thus evaluated. Transcript levels of expressions were affected by removing rRNA; de-duplicate as well as sub-setting read count. Read de-duplication and sub-setting resulted in a large number of transcripts without read support; as such rRNA read treated transcript abundance estimations were adopted.

7.2 Profiling and characterising 'Golden Delicious' fruit pulp expressed transcripts related to fruit carbohydrate content and hormonal changes
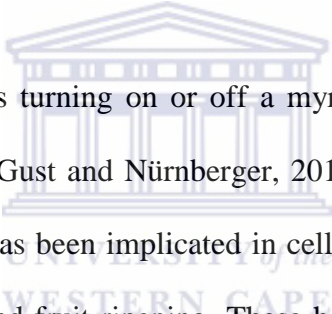
The parenchymatus mesophyll cells of the pulp tissue have been reported as the centre of most of the physiological process characteristic of the climacteric phase (Gonkiewicz, 2011). The pulp constitutes the major portion of the apple fruit that is consumed as well as a reservoir of compounds, such as carbohydrates, that attract seed dispersers (Gonkiewicz, 2011). To gain a molecular understanding underlying changes in physiological processes in the pulp tissue conferring sugar content, an in depth understanding of the patterns of expression of genes related to the carbohydrate accumulation and hormonal changes and regulation was undertaken. In the work described in chapter 4, sugar content (Brix˚) in 'Golden Delicious' positively correlated

266

with an increase in fruit fresh weight, with highest sugar content at daa150. In previous postharvest studies on apple fruit, sugar content was noted to increase even further as a consequent of continued degradation of carbohydrates (Geigenberger, 2011; Gonkiewicz, 2011; Han, Seo, *et al.*, 2007).

From a total of 186498 reconstructed transcripts, 27255 pulp tissue significantly expressed transcripts were obtained by performing a two-way analysis of variance (ANOVA). These pulp significantly expressed transcripts were clustered into 20 closely correlated clusters, with eleven main groups of transcript expression trends showing: an increase or decrease throughout fruit development; transitory peaks or decreases at daa35, daa60, daa87, and daa135. Most of the transcripts that decreased with fruit development encoded genes in the plant hormone signal transduction, plant pathogen interaction, ribosome, photosynthesis and amino sugar and nucleotide sugar metabolism pathways. Transcripts that increased in expression with fruit development on the other had encoded genes in the plant hormone signal transduction, carbon fixation, amino sugar nucleotide sugar metabolism, peroxisome, plant pathogen interaction and oxidative phosphorylation KEGG pathways. MIPS functional classification of the pulp transcripts revealed that daa150 had the highest frequency of transcripts in all the functional categories. Functional categories of protein with binding function, metabolism, interaction with environment and cell defence were the most enriched categories.

Pathway enrichment analysis showed that genes encoding Calvin cycle, sucrose and starch biosynthetic pathway enzymes had peak expression between daa60 and daa87, a

period characterised by Janssen, *et al,* (2008) as a having the highest starch accumulation levels. The progressive maturation of the fruit is characterised by a transition from the pre-climacteric to the climacteric phase, resulting in increased degradation of energy reserves to meet the energy requirements of the physiological processes during the climacteric (Janssen*, et al.*, 2008). This possibly explains the observed increase in expression of genes encoding starch and sucrose degrading enzymes with fruit development. Farmers and merchants use the changes in starch, sucrose and other sugars as well as organic acids content in the developing apple fruit as maturity indices (Fleancu, 2007).

Plant hormones act as switches turning on or off a myriad of metabolic switch points throughout fruit development (Gust and Nürnberger, 2012; Pieterse*, et al.*, 2012). Gene regulation by plant hormones has been implicated in cell division, cell enlargement, cell differentiation, plant defence and fruit ripening. These hormonal changes have practical implications in the agronomy of the apple fruit. Examples of hormonal usage in apple are the production of parthenocarpic fruit by application of GA, hastening fruit ripening by application of ethaphon (ethylene) and fruit senescence by application of ABA. Nonetheless, a number of grey areas remain as to the mechanisms, effects and interaction of these plant hormone accumulations during fruit development. Chapter 4 elucidated expression profiles of ABA, auxin, ethylene, and BR biosynthesis pathway genes encoding pulp transcripts. ABA biosynthesis encoding transcripts increased in expression with fruit development suggesting possible increase in fruit stress with fruit development.

The observation might be corroborated by the high enrichment of oxidative phosphorylation pathway encoding transcripts in the pulp.

Transcripts encoding ethylene biosynthesis enzymes increased in expression with fruit development, except for ACS, which decreased with fruit development. The ethylene biosynthesis pathway has been extensively studied, but the intricate mechanisms of the regulation of ethylene production during pre-climacteric, climacteric and post climacteric remain unknown. It is however generally accepted that pre-climacteric phase ethylene production is driven by *de novo* synthesis of ethylene, a process catalysed by ACS (Van de Poel*, et al.*, 2012). This period in which ACS dominates ethylene production has been referred to as the system I. The produced ethylene then triggers the activation of ACO, which results in the burst in ethylene production forcing the fruit to enter the climacteric phase, also referred as the system II. However, autocatalysis has been reported to be regulated by the biological clock, which turns off autocatalysis, resulting in reduced ACO and increased ACS expression levels (Van de Poel*, et al.*, 2012). This post-climacteric down regulation of the ethylene production determines the shelf life of the fruit. Storage conditions therefore that force the fruit to stay in the autocatalysis phase drastically reduce the shelf life of the fruit.

The expression pattern of BRI1 encoding transcripts increased with fruit development peaking up at daa135 just after YUCCA peak expression at daa87, suggesting possible involvement of BR and auxin pathway genes in the final fruit cell expansion before physiological maturity.

7.3 Profiling and characterising 'Golden Delicious' fruit peel expressed transcripts related to epicuticular wax accumulation, pigmentation, flavour and cell-wall metabolism

The peel is a crucial component of the apple fruit, invaluable for fruit and plant's survival against the internal and external environment. Visible peel properties such as pigmentation define colour, while the development of a wax layer on the peel surface determine the glossiness of the fruit. Cell wall composition and integrity determines the fruits' firmness and hence the potential of the fruit to survive during packing and transportation, as well as susceptibility to abiotic and biotic environments.

To characterise on some of the properties that might define peel pigmentation, wax accumulation, flavour and cell wall metabolism, a set of transcripts that were significantly expressed in the peel tissue was analysed. The analysis revealed cell fate, interaction with environment; protein with binding function, development and metabolism to be the predominant MIPS functional categories in apple peel tissue. Fifty-nine % of the transcripts that were significantly expressed in peel increased or decreased in expression with fruit development. Transcripts that showed an increase in expression with fruit development were mostly enriched for protein processing in endoplasmic reticulum, plant pathogen interaction, carbon fixation, RNA transport and oxidative phosphorylation pathways.

The cuticular waxes are complexes of saturated and unsaturated fatty acids, very-long fatty acids, cutin and derivatives of fatty acids. The composition of cuticle has been

reported to change with fruit development and growing conditions (Rhee, *et al.*, 1998). Fatty acid, very-long fatty acid and cuticle biosynthesis encoding transcripts increased in expression with fruit development, suggesting an increase in epicuticular wax thickness with fruit development. These results are in agreement with the observed accumulation of total waxes in 'Red Fuji' apple, during fruit development. At harvesting, the fruit is coated by the maximum possible wax thickness, which however decreases in storage according to the observations made by Dong, *et al,* (2012). This study however, further qualified peak synthesis of wax to be at daa60 and a possible termination at daa135.

Fruit aroma is one of the traits that define a fruit cultivar's acceptance by consumers (Schaffer, *et al.*, 2007). During apple fruit development aroma volatile compounds accumulate mainly driven by the mevalonate (MVA), 2-C-methyl-D-erythritol4-phosphate (MEP) and ester biosynthesis pathways (Schaffer, *et al.*, 2007). In this study MVA and MEP pathway genes were encoded for by transcripts that increased in expression with time of harvesting, suggesting an increase in aroma compounds with fruit maturity. The high expression levels of the sesquiterpenoid ß-amyrin synthase (BAS) suggested ß-amyrin might be the predominant sesquiterpenoid accounting for most of the 'Golden Delicious' fruit aroma.
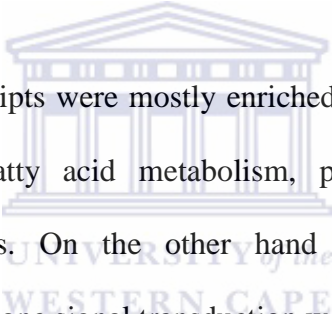
Transcripts encoding flavonoid and anthocyanin biosynthetic pathway structural genes had high expression during early fruit development, coinciding with the development of the reddish colouration early during fruit development. Loss of anthocyanin conferred colouration in fruit and other plant tissues has been attributed to the change in vacuolar

pH, degradation of the anthocyanin with changes in growth stages and environmental conditions(Castañeda-Ovando*, et al.*, 2009). One of the enzymes, phenol peroxidase (PPO) was encoded for by transcripts whose expression had a similar pattern to the anthocyanin structural genes, suggesting it might be responsible for the degradation of anthocyanins with fruit development. Transcripts encoding carotenoid biosynthesis genes increased with fruit maturity supporting the observed change to golden colour with fruit maturity.

Cell wall composition and integrity with fruit development is regarded as the major determinant of fruit firmness, susceptibility and fruit shelf life. An elucidation of cell wall formation and degradation mechanisms as well as temporal changes assists in targeting time and particular genes responsible for the changes. In this study transcripts encoding cell wall loosening genes decreased in expression with harvesting time. A low expression of the cell wall degrading genes at fruit harvesting in agronomy terms is desirable as it means cell walls are still intact, hence fruit will be firm. However, temporal expression of cell wall loosening proteins is required during cell division and enlargement. Expansins are some of the enzymes that have been implicated in cell wall loosening (Choe and Cosgrove, 2010; Goulao and Oliveira, 2008). Two expansins (A8 and A10) had peak expression at daa87 and daa60 respectively, suggesting they could be responsible for the final push in cell expansion when fruit cells attained their final size.

7.4 Comparative analyses of transcripts up regulated in peel and pulp tissues
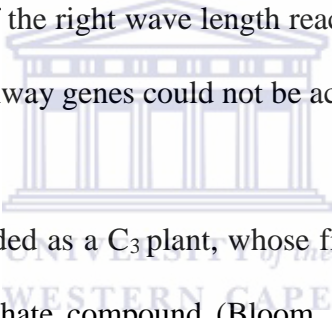
Spatial and temporal differential gene expressions in 'Golden Delicious' fruit define the properties of peel and pulp tissues. Several studies have corroborated the existence of spatial and temporal gene expression in apple fruit (Malladi and Johnson, 2011; Roongsattham, *et al.*, 2012; Seo, *et al.*, 2011). Unlike an analysis of individual tissues, a comparison of transcripts up regulated between the peel and pulp tissues provided more insight into physiological processes up regulated or exclusively activated in a particular tissue (chapter 6).

Peel tissue up regulated transcripts were mostly enriched for plant pathogen interaction, terpenoid, carbon fixation, fatty acid metabolism, photosynthesis, carotenoid and flavonoid metabolic pathways. On the other hand protein processing, oxidative phosphorylation and plant hormone signal transduction were more enriched in fruit pulp.

Plants have the ability to capture and convert light energy (in photons) into chemical energy usable by the plant. The light reaction process ('Z-scheme') of photosynthesis converts photons to ATP and NADPH, which are used in CO_2-fixation. Substantial knowledge has been gained from previous studies, yet more fascinating discoveries in concert or disagreement to the establish models continue to be uncovered(Solovchenko, *et al.*, 2010). Most of the studies on the light reaction of photosynthesis are well established in leaves, yet little is known about photosynthesis in fruits (Solovchenko, *et al.*, 2010). Transcripts encoding photosynthesis pathway proteins were only expressed in

the peel and had peak expression at daa35 to daa60, suggesting that 'Golden Delicious' fruit's photosynthetic capacity reached maximum between daa35 and daa60 and gradually decreased with fruit maturity. The period when fruit photosynthetic capacity starts declining is crucial in agronomy as the fruit becomes dependent on phloem import, thus all factors affecting source sink relationship have to be optimum to obtain good quality fruits (Blanke, *et al.*, 2012; Iqbal, *et al.*, 2011; Morandi, *et al.*, 2011). Though the mesophyll cells contained green colouration, possibly due to chlorophyll, they were not enriched for photosynthesis pathway genes. The absence of transcripts encoding photosynthesis pathway proteins in the pulp tissue could have been as a result of unavailability of light energy of the right wave length reaching these deeply located cells. As such the photosynthesis pathway genes could not be activated.

The apple plant has been regarded as a C_3 plant, whose first fixed carbon compound is a three-carbon glycerate 3-phosphate compound (Bloom, 2009; Ducat and Silver, 2012; Eckardt, 2011). As in other C_3 plants, this mechanism is only ideal in cool, wet, and cloudy climates, where light levels may be low. Because the metabolic pathway is more energy efficient, and water is normally plentiful, the stomata can stay open and let in more carbon dioxide. However, with the spread of these C_3 plants to more hotter and drier marginal areas, the C_3 mechanism efficiency decreased, owing to increased photorespiration. In marginal areas, C_4 plants have adopted an alternative mechanism to circumvent the direct use of the inefficient CO_2-fixing ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) enzyme, by concentrating CO_2 around Rubisco (Bloom, 2009; Ducat and Silver, 2012; Eckardt, 2011). Interestingly in this study C_4-

pathway genes were enriched in the pulp, which might be testimony to the existence of a functional C_4-carbon fixing pathway in apple fruit. The enrichment of the Calvin cycle genes in the peel suggests direct CO_2 fixation to glycerate 3-phosphate in the peel in concert with the C_3 carbon fixing mechanism in apple. No studies thus far have alluded to the existence of dual carbon fixing mechanisms in apple, but the enrichment of the C_4-carbon fixation pathway genes in the pulp suggested the existence of C_4 pathway in 'Golden Delicious' fruit pulp tissue.

The progressive increase in *ß-amylase, amyA, glgX* and *sucrose synthase* suggested an increase in total soluble solids with fruit development. Peak glgA expression at daa60 suggested maximum starch accumulation to be at daa60, which is in agreement with findings by Janssen *et al*., (2008) in developing apple fruit.

Fruit aroma, is a complex trait, determined volatile compounds from the fruit. Consumers use aroma characteristics of fruit for acceptance, as such aroma is an important marketing trait in apple. Fruit produced volatiles have been reported to be enriched in the peel of apple fruit. Genes encoding production of volatile compounds in apple fruit have been shown in this study to exhibit temporal and spatial expression in peel and pulp tissues. Terpenoid, sesquiterpenoid as well as ester biosynthesis pathways have been implicated in the production of volatile compounds that give aroma in the fruit. Genes encoding these three pathway enzymes were more enriched in the peel than pulp tissues of the developing apple fruit. The terpenoid genes enriched in the pulp tissue encoded protein

prenylation enzymes suggesting post-translational protein modification to aid in protein localization and function.

Respiration rate is a desirable factor in the provision of energy for physiological processes, but undesirable to the apple fruit grower in the expenditure of dry matter. Factors such as temperature regulate enzyme activity, and high temperature accelerates the rates of respiration in apple fruit resulting in reduced fruit dry weight, hence fruit yield. Though respiration is one of the components and determinants of apple fruit yield, manipulation of this parameter by the grower is not feasible. However, understanding of the energy generating oxidative phosphorylation helps in the elucidation of the possible control points, for optimizing temperature tolerance. Apple fruit peel and pulp tissue comparison revealed that the pulp tissue was more enriched for oxidative phosphorylation pathway genes, predominantly V-type and F-type ATPases, with lowest expression at daa60, possibly a period of lowest energy demand. The enrichment of V-type ATPases in the pulp could be for regulating cellular solute content; hence turgidity and optimum pH balance for optimum physiological cell functioning. The peel tissue on the other hand expressed AHA2, a functionally redundant equivalent of V-type ATPases. High oxidative phosphorylation pathway gene enrichment in the pulp suggested that of the two tissues physiological processes characteristic of the pulp tissue required more energy and could be the determinant of fruit shelf life.

The protective waxy layer on apple fruit surfaces has been characterized by molecular and biochemical approaches. Genes encoding wax formation in apple fruit are

predominantly expressed in the peel tissue. Peel transcripts encoding fatty acid biosynthesis pathway genes had peak expression at daa60, suggesting that by about daa60 the formation of the precursors for cutin, suberin and wax had reached their maximum. Low expression of fatty acid biosynthesis genes in the pulp suggested possible contribution of the pulp tissue towards epicuticular wax formation.

Auxin and cytokinin signal transduction pathway genes were enriched in both peel and pulp tissues, but more so in pulp, suggesting auxin and cytokinin-auto-regulation in these tissues. The accumulation of BZR1 /2 in the pulp might have functioned to regulate nucleus gene expression and plant development by promoting cell division and enlargement. ABA signal transduction pathway genes were more enriched in the pulp with some increasing with fruit development, possibly in response to increased stress due to increased oxidative phosphorylation. In the peel, ABF enrichment might have functioned in stomata control. The enrichment of the ethylene signal transduction gene MPK6 in the pulp did not show a clear progressive increase with fruit development. The enrichment of DELLA in the peel tissue suggested that the peel tissue contained low GA, and that DELLA inhibited GA signal transduction. Enrichment of GID/SLY1 complex in the pulp tissue suggested the existence of GID/SLY1-mediated DELLA protein degradation and hence GA signal transduction in the pulp. Inactivation of GA signal by DELLA in the peel and activation of the GA signal in the pulp, corroborated pulp as an actively dividing and expanding tissue, accounting for most of the fruit increases in size.

Flavonoid biosynthesis pathway genes were more enriched in the peel, possibly to function as screen against UV-light and as antioxidants. The peel tissue was more enriched for anthocyanin and carotenoid biosynthetic pathway genes, suggesting the peel accumulated more anthocyanins and carotenoids with fruit development.

7.5 Future work

Relying on the assembled genome in genome-guided transcript reconstruction, suggest that the reconstructed transcripts can only be as good as the assembled genome. The current apple genome v1.0 still has up to 37% of the genome denoted as ambiguous bases. The fragmented nature of the genome necessitates refinement of the genome, by genome reassembly or gap filling. A more refined genome will allow for better reconstruction and localisation of transcripts within the apple genome towards assisting in transcript/genome annotation and expression quantitative loci (eQTLs) studies.

In this study *de novo* and genome-guided algorithms were evaluated and showed differential sensitivity in the detection and reconstruction of transcripts, but complimented in reconstructing of unmapped reads. Consistence in transcript detection regardless of the algorithm will aid in the reproducibility of transcript reconstruction across algorithms, which will allow validation of transcript reconstruction. Several improvements continue to be effected in the assembly algorithms towards this goal.

In this study several gene transcripts were identified and reported to show temporal and spatial expression in peel and pulp tissues of the 'Golden Delicious' apple fruit. These

gene transcripts provide the foundation for fruit and tissue specific candidate gene studies, such as regulation of wax accumulation in peel and starch accumulation in the pulp tissue. Such studies also allows for gain/loss of function analyses, further elucidating the functions and mechanisms of the respective genes in defining fruit and tissue properties.

The differential gene expressions noted during apple fruit development require corroboration from either proteomic or metabolomics studies to confidently assign the observed transcript changes to a phenotype, hence function.

7.6 Conclusion

This study has thus provided validation to existing apple genome v1.0 gene models as well as improved and provided new gene models. A large number of genes not mapping to the pseudo molecules suggested the existence of a substantial gene space unaccounted for by the current apple genome v1.0. As such the genome requires further improvements.

Analysis of the apple fruit pulp during fruit development demonstrated intricate relationships with carbohydrate accumulation and hormonal metabolism and perception. Significantly peel enriched pathways and gene changes defining peel colour, flavour, wax formation and cell wall integrity exhibited temporal expression with fruit development. Peel and pulp tissue comparison revealed the existence of temporal and spatial gene

expression, which might function in defining the properties of the tissues. The study has thus provided an in depth catalogue and elucidation of the pathways enriched in peel and pulp 'Golden Delicious' tissues. The extensive catalogue of gene expression patterns defined here will serve as an invaluable reference for future fruit quality investigations including the exploration of the complex transcriptional regulatory hierarchies that govern apple pulp development and differentiation. The catalogue will also allow targeted pathway and gene studies towards apple fruit improvement.

Despite the achievements of this study, the major limitation in comprehensive annotation of the reconstructed transcripts remains the depth of the publicly available non-redundant databases for interrogating.
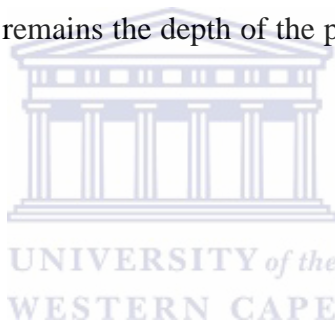
8 REFERENCE

Abassi, N., Kushad, M., and Endress, A. (1998). Active oxygen-scavenging enzymes activities in developing apple flowers and fruits. *Scientia Horticulturae, 74*(3), 183-194.

Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C., and Venter, J. C. (1993). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genetics, 4*(4), 373-380.

Adams-Phillips, L., Barry, C., and Giovannoni, J. (2004). Signal transduction systems regulating fruit ripening. *Trends in Plant Science, 9*(7), 331-338.

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., Van Arkel, G., and Pereira, A. (2004). The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in Arabidopsis. *The Plant Cell Online, 16*(9), 2463-2480.

Ajjawi, I., Lu, Y., Savage, L. J., Bell, S. M., and Last, R. L. (2010). Large-scale reverse genetics in *Arabidopsis*: case studies from the Chloroplast 2010 Project. *Plant Physiology, 152*(2), 529-540.

Alayón-Luaces, P., Ponce, N., Mroginski, L. A., Stortz, C. A., and Sozzi, G. O. (2011). Compositional changes in cell wall polysaccharides from apple fruit callus cultures modulated by different plant growth regulators. *Plant Science, 185*, 169–175.

Albert, Z., Ivanics, B. z., Molnár, A., Miskó, A. s., Tóth, M., and Papp, I. n. (2013). Candidate genes of cuticle formation show characteristic expression in the fruit skin of apple. *Plant Growth Regulation, 20*(1), 71-78.

Albrecht, G., Brenner, S., DuBridge, R. B., Lloyd, D. H., and Pallas, M. C. (2012). Massively parallel signature sequencing by ligation of encoded adaptors. Patent number US6013445: Google Patents.

Alexander, L., and Grierson, D. (2002). Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. *Journal of Experimental Botany, 53*(377), 2039-2055.

Allan, A. C., Hellens, R. P., and Laing, W. A. (2008). MYB transcription factors that colour our fruit. *Trends in Plant Science, 13*(3), 99-102.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences, USA, 74*(12), 5350-5354.

Andrews, S. (2010). A quality control tool for high throughput sequence data. [http://www. bioinformatics. bbsrc. ac. uk/projects/fastqc/].

Aragón, C., Carvalho, L., González, J., Escalona, M., and Amancio, S. (2012). The physiology of ex vitro pineapple (*Ananas comosus* L. Merr. var MD-2) as CAM or C3 is regulated by the environmental conditions. *Plant Cell Reports*, 1-13.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., and Eppig, J. T. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics, 25*(1), 25-29.

Ashour, M., Wink, M., and Gershenzon, J. (2011). Biochemistry of Terpenoids: Monoterpenes, Sesquiterpenes and Diterpenes. *Annual Plant Reviews Volume 40: Biochemistry of Plant Secondary Metabolism, Second Edition*, 258-303.

Ashraf, M., Akram, N., Arteca, R., and Foolad, M. (2010). The physiological, biochemical and molecular roles of brassinosteroids and salicylic acid in plant processes and salt tolerance. *Critical Reviews in Plant Sciences, 29*(3), 162-190.

Ashrafi, H., Kinkade, M. P., Merk, H. L., and Foolad, M. R. (2012). Identification of novel quantitative trait loci for increased lycopene content and other fruit quality traits in a tomato recombinant inbred line population. *Molecular Breeding*, 1-19.

Atkinson, R. G., Bolitho, K. M., Wright, M. A., Iturriagagoitia-Bueno, T., Reid, S. J., and Ross, G. S. (1998). Apple ACC-oxidase and polygalacturonase: ripening-specific gene expression and promoter analysis in transgenic tomato. *Plant Molecular Biology, 38*(3), 449-460.

Atkinson, R. G., Sutherland, P. W., Johnston, S. L., Gunaseelan, K., Hallett, I. C., Mitra, D., Brummell, D. A., Schroder, R., Johnston, J. W., and Schaffer, R. J. (2012). Down-regulation of POLYGALACTURONASE 1 alters firmness, tensile strength and water loss in apple (*Malus x domestica*) fruit. *BMC Plant Biology, 12*(1), 129-142.

Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A., Moulton, G., Nordle, A., Paine, K., and Taylor, P. (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research, 31*(1), 400-402.

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C3 plants prior to their recruitment into the C4 pathway. *Journal of Experimental Botany, 62*(9), 3049-3059.

R Core Team,. (2013). A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Bai, S., Wang, A., Igarashi, M., Kon, T., Fukasawa-Akada, T., Li, T., Harada, T., and Hatsuyama, Y. (2012). Distribution of MdACS3 null alleles in apple (*Malus x domestica* Borkh.) and its relevance to the fruit ripening characters. *Breeding Science, 62*(1), 46-52.

Bajguz, A. (2011). Brassinosteroids, occurence and chemical structures in plants *Brassinosteroids: A Class of Plant Hormone* (pp. 1-27). Netherlands: Springer.

Baldi, P., Patocchi, A., Zini, E., Toller, C., Velasco, R., and Komjanc, M. (2004). Cloning and linkage mapping of resistance gene homologues in apple. *Theoretical and Applied Genetics, 109*(1), 231-239.

Bangerth, F. K., Song, J., and Streif, J. (2012). Physiological Impacts of Fruit Ripening and Storage Conditions on Aroma Volatile Formation in Apple and Strawberry Fruit: A Review. *HortScience, 47*(1), 4-10.

Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiology, 155*(4), 1520-1532.

Barry, C. S., Llop-Tous, M. I., and Grierson, D. (2000). The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the

transition from system-1 to system-2 ethylene synthesis in tomato. *Plant Physiology, 123*(3), 979-986.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., and Sonnhammer, E. L. L. (2004). The Pfam protein families database. *Nucleic Acids Research, 32*(suppl 1), D138-D141.

Batut, P. J., Dobin, A., Plessy, C., Carninci, P., and Gingeras, T. R. (2012). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Research, 23*(1), 169-180.

Beisson, F., Li-Beisson, Y., and Pollard, M. (2012). Solving the puzzles of cutin and suberin polymer biosynthesis. *Current Opinion in Plant Biology, 15*(3), 329-337.

Ben-Dor, A., Shamir, R., and Yakhini, Z. (1999). Clustering gene expression patterns. *Journal of Computational Biology, 6*(3-4), 281-297.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., and Bignell, H. R. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*(7218), 53-59.

Berckmans, B., Vassileva, V., Schmid, S. P. C., Maes, S., Parizot, B., Naramoto, S., Magyar, Z., Kamei, C. L. A., Koncz, C., and Bögre, L. (2011). Auxin-dependent cell cycle reactivation through transcriptional regulation of Arabidopsis E2Fa by lateral organ boundary proteins. *The Plant Cell Online, 23*(10), 3671-3683.

Berüter, J. (1985). Sugar accumulation and changes in the activities of related enzymes during development of the apple fruit. *Journal of Plant Physiology, 121*(4), 331-341.

Beyenbach, K. W., and Wieczorek, H. (2006). The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. *Journal of Experimental Biology, 209*(4), 577-589.

Blanke, M., Kunz, A., Aliev, T., Solomakhin, A., and Klad, A. (2012). *Shading as an Effective Means for Crop Load Management and Fruit Quality Enhancement in Apple Trees.* Paper presented at the VII International Symposium on Light in Horticultural Systems 956.

Bloom, A. J. (2009). Plant Responses to Rising Carbon Dioxide and Nitrogen Relations. *UC Davis: Department of Plant Sciences, UC Davis. Retrieved from: http://www.escholarship.org/uc/item/58f012gm.*

Bogs, J., Jaffé, F. W., Takos, A. M., Walker, A. R., and Robinson, S. P. (2007). The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. *Plant Physiology, 143*(3), 1347-1361.

Börner, T., Parlitz, S., Kunze, R., Mueller-Roeber, B., Balazadeh, S., Köllmer, I., Werner, T., Schmülling, T., Holst, K., and Zubo, Y. O. (2011). Regulation of plant primary metabolism. *Journal of Plant Physiology, 168*(12), 1309-1310.

Boyer, J., and Liu, R. H. (2004). Apple phytochemicals and their health benefits. *Nuture Journal, 3*, 5-20.

Bridges, H. R., Birrell, J. A., and Hirst, J. (2011). The mitochondrial-encoded subunits of respiratory complex I (NADH: ubiquinone oxidoreductase): identifying residues important in mechanism and disease. *Biochemical Society Transactions, 39*(3), 799-806.

Brummell, D., and Harpster, M. (2001). Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. *Plant Molecular Biology, 47*(1), 311-339.

Burstenbinder, K., and Sauter, M. (2012). Early events in the ethylene biosynthetic pathway, Regulation of the pools of methionine and S-adenosylmethionine. *Annual Plant Reviews, The Plant Hormone Ethylene, 44*, 22.

Byun, M., Abhyankar, A., Lelarge, V., Plancoulaine, S., Palanduz, A., Telhan, L., Boisson, B., Picard, C., Dewell, S., and Zhao, C. (2010). Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *The Journal of Experimental Medicine, 207*(11), 2307-2312.

Cajuste, J. F., González-Candelas, L., Veyrat, A., García-Breijo, F. J., Reig-Armiñana, J., and Lafuente, M. T. (2010). Epicuticular wax content and morphology as related to ethylene and storage performance of 'Navelate' orange fruit. *Postharvest Biology and Technology, 55*(1), 29-35.

Caligiuri, M., and Bauerle, R. (1991). Identification of amino acid residues involved in feedback regulation of the anthranilate synthase complex from Salmonella typhimurium. Evidence for an amino-terminal regulatory site. *Journal of Biological Chemistry, 266*(13), 8328-8335.

Cantacessi, C., Campbell, B. E., and Gasser, R. B. (2011). Key strongylid nematodes of animals, Impact of next-generation transcriptomics on systems biology and biotechnology. *Biotechnology Advances, 30*(3), 469-488.

Castañeda-Ovando, A., Pacheco-Hernández, M. L., Páez-Hernández, M. E., Rodríguez, J. A., and Galán-Vidal, C. A. (2009). Chemical studies of anthocyanins: A review. *Food Chemistry, 113*(4), 859-871.

Cevallos-Casals, B. A., and Cisneros-Zevallos, L. (2004). Stability of anthocyanin-based aqueous extracts of Andean purple corn and red-fleshed sweet potato compared to synthetic and natural colorants. *Food Chemistry, 86*(1), 69-77.

Chagné, D., Krieger, C., Rassam, M., Sullivan, M., Fraser, J., André, C., Pindo, M., Troggio, M., Gardiner, S. E., and Henry, R. A. (2012). QTL and candidate gene mapping for polyphenolic composition in apple fruit. *BMC Plant Biology, 12*(1), 12-27.

Chalker-Scott, L. (1999). INVITED REVIEWS-Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology, 70*(1), 1-9.

Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter, 11*(2), 113-116.

Chapman, E. J., and Estelle, M. (2009). Mechanism of auxin-regulated gene expression in plants. *Annual Review of Genetics, 43*, 265-285.

Chase, C. (2011). *Identification of Fertility Restorers for S Male-Sterile Maize: Beyond PPRs.* Paper presented at the Plant and Animal Genome XX Conference (January 14-18, 2012).

Chen, G., Deng, W., Truksa, M., Peng, F. Y., and Weselake, R. J. (2012). The B_{sister} MADS-box proteins have multiple regulatory functions in plant development. *Biocatalysis and Agricultural Biotechnology, 1*(3), 203–206.

Choe, H. T., and Cosgrove, D. J. (2010). Expansins as agents in hormone action. In P. J. Davies (Ed.), *Plant Hormones* (pp. 262-281). Netherlands: Springer.

Choi, H. S., Rom, C. R., and Gu, M. (2011). Effects of different organic apple production systems on seasonal nutrient variations of soil and leaf. *Scientia Horticulturae, 129*(1), 9-17.

Ciardi, J. A., Tieman, D. M., Lund, S. T., Jones, J. B., Stall, R. E., and Klee, H. J. (2000). Response to *Xanthomonas campestris* pv. vesicatoria in tomato involves regulation of ethylene receptor gene expression. *Plant Physiology, 123*(1), 81-92.

Clifton, S. W., and Mitreva, M. (2009). Strategies for Undertaking Expressed Sequence Tag (EST) Projects. *Methods in Molecular Biology (Clifton, NJ), 533*, 13-32.

Clouse, S. D. (2011). Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development. *The Plant Cell Online, 23*(4), 1219-1230.

Coart, E., Van-Glabeke, S., De-Loose, M., Larsen, A. S., and Roldan-Ruiz, I. (2006). Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Molecular Ecology, 15*(8), 2171-2182.

Conant, G. C., and Wolfe, K. H. (2008). GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics, 24*(6), 861-862.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21*(18), 3674-3676.

Corpet, F., Servant, F., Gouzy, J., and Kahn, D. (2000). ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Research, 28*(1), 267-269.

Costa, F., Alba, R., Schouten, H., Soglio, V., Gianfranceschi, L., Serra, S., Musacchi, S., Sansavini, S., Costa, G., and Fei, Z. (2010). Use of homologous and heterologous gene expression profiling tools to characterize transcription dynamics during apple fruit maturation and ripening. *BMC Plant Biology, 10*(1), 229-245.

Cova, V., Perini, D., Soglio, V., Komjanc, M., Weg, E., Gessler, C., and Gianfranceschi, L. (2012). Exploiting expressed sequence tag databases for mapping markers associated with fruit development and fruit quality in apple. *Molecular Breeding, 29*(3), 699-715.

Covshoff, S., and Hibberd, J. M. (2012). Integrating C4 photosynthesis into C3 crops to increase yield potential. *Current Opinion in Biotechnology, 23*(2), 209–214.

Crowell, E. F., Bischoff, V., Desprez, T., Rolland, A., Stierhof, Y. D., Schumacher, K., Gonneau, M., Höfte, H., and Vernhettes, S. (2009). Pausing of Golgi bodies on microtubules regulates secretion of cellulose synthase complexes in *Arabidopsis*. *The Plant Cell Online, 21*(4), 1141-1154.

Curry, E. (2008). Effects of 1-MCP applied postharvest on epicuticular wax of apples (*Malus domestica* Borkh.) during storage. *Journal of the Science of Food and Agriculture, 88*(6), 996-1006.

Curry, E. A. (2009). *Growth-Induced Microcracking and Repair Mechanisms of Fruit Cuticles*. Paper presented at the Proceedings of the SEM Annual Conference.

Cutter, E. G. (1971). *Plant anatomy, Experiment and Interpretation, part 2 organs.* . London: Edward Arnold.

Dal Cin, V., Barbaro, E., Danesin, M., Murayama, H., Velasco, R., and Ramina, A. (2009). Fruitlet abscission: A cDNA-AFLP approach to study genes differentially expressed during shedding of immature fruits reveals the involvement of a putative auxin hydrogen symporter in apple (*Malus domestica* L. Borkh). *Gene, 442*(1-2), 26-36.

Davies, P. J. (2004). *Plant hormones: biosythesis, signal transduction, action.* Netherlands: Springer.

Davies, P. J. (2010). *The plant hormones: their nature, occurrence, and functions*. Netherlands: Springer.

De Lorenzo, G., D'Ovidio, R., and Cervone, F. (2001). The role of polygalacturonase-inhibiting proteins (PGIPs) in defense against pathogenic fungi. *Annual Review of Phytopathology, 39*(1), 313-335.

de Oliveira, L. S., Gregoracci, G. B., Silva, G. G. Z., Salgado, L. T., Gilberto Filho, A., Alves-Ferreira, M. A., Pereira, R. C., and Thompson, F. L. (2012). Transcriptomic analysis of the red seaweed Laurencia dendroidea (Florideophyceae, Rhodophyta) and its microbiome. *BMC Genomics, 13*(1), 487-500.

Deluc, L. G., Grimplet, J., Wheatley, M. D., Tillett, R. L., Quilici, D. R., Osborne, C., Schooley, D. A., Schlauch, K. A., Cushman, J. C., and Cramer, G. R. (2007). Transcriptomic and metabolite analyses of Cabernet Sauvignon grape berry development. *BMC Genomics, 8*(1), 429.

Demnitz-King, A., Ho, L., and Baker, D. (1997). Activity of sucrose hydrolysing enzymes and sugar accumulation during tomato fruit development. *Plant Growth Regulation, 22*(3), 193-201.

den Boer, B. G. W., and Murray, J. A. H. (2000). Control of plant growth and development through manipulation of cell-cycle genes. *Current Opinion in Biotechnology, 11*(2), 138-145.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genetics, 14*(4), 457-460.

Deshpande, N., Wilkins, M. R., Packer, N., and Nevalainen, H. (2008). Protein glycosylation pathways in filamentous fungi. *Glycobiology, 18*(8), 626-637.

Dong, Q., Schlueter, S. D., and Brendel, V. (2004). PlantGDB, plant genome database and analysis tools. *Nucleic Acids Research, 32*(suppl 1), D354-D359.

Dong, X., Rao, J., Huber, D. J., Chang, X., and Xin, F. (2012). Wax composition of 'Red Fuji' apple fruit during development and during storage after 1-methylcyclopropene treatment. *Horticulture, Environment, and Biotechnology, 53*(4), 288-297.

Ducat, D. C., and Silver, P. A. (2012). Improving carbon fixation pathways. *Current Opinion in Chemical Biology, 16*(3-4), 337–344.

Dussi, M., Sugar, D., and Wrolstad, R. (1995). Characterizing and quantifying anthocyanins in red pears and the effect of light quality on fruit color. *Journal of the American Society for Horticultural Science, 120*(5), 785-789.

Eckardt, N. A. (2011). How to Make a C4 Plant: Insight from Comparative Transcriptome Analysis. *The Plant Cell Online, 23*(6), 2009-2009.

Elaine R, M. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics, 24*(3), 133-141.

Ermolaeva, O., Rastogi, M., Pruitt, K. D., Schuler, G. D., Bittner, M. L., Chen, Y., Simon, R., Meltzer, P., Trent, J. M., and Boguski, M. S. (1998). Data management and analysis for gene expression arrays. *Nature Genetics, 20*(1), 19-24.

Esau, K. (1977). Anatomy of seed plants 2nd edn New York John Wiley & Sons.

FAO (2010). FAOSTAT: apple production. Retrieved 11 September 2010, from FAO Statistics Division: http://faostat.fao.org:

FAO (2011). FAOSTAT: apple production. Retrieved 28 February 2013, from FAO Statistical Division: http://faostat.fao.org:

Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., Porceddu, A., Venturini, L., Bicego, M., and Murino, V. (2012). The Grapevine Expression Atlas Reveals a Deep Transcriptome Shift Driving the Entire Plant into a Maturation Program. *The Plant Cell Online, 24*(9), 3489-3505.

Fattouch, S., Caboni, P., Coroneo, V., Tuberoso, C., Angioni, A., Dessi, S., Marzouki, N., and Cabras, P. (2008). Comparative analysis of polyphenolic profiles and antioxidant and antimicrobial activities of tunisian pome fruit pulp and peel aqueous acetone extracts. *Journal of Agricultural and Food Chemistry, 56*(3), 1084-1090.

Finkelstein, R. R., Gampala, S. S. L., and Rock, C. D. (2002). Abscisic acid signaling in seeds and seedlings. *The Plant Cell Online, 14*(Supplement 1), S15-S45.

Finkelstein, R. R., and Rock, C. D. (2002). *Abscisic acid biosynthesis and response*: 1: e0058. Published online 2002 September 30. doi: 10.1199/tab.0058.

Fleancu, M. (2007). Correlations among some physiological processes in apple fruit during growing and maturation processes. *International Journal of Agriculture and Biology, 9*(4), 613–616.

Foolad, M. R. (2009). High lycopene content tomato plants and markers for use in breeding for same: WO2009117423 A2 Google Patents.

Foreman, J., Demidchik, V., Bothwell, J. H. F., Mylona, P., Miedema, H., Torres, M. A., Linstead, P., Costa, S., Brownlee, C., and Jones, J. D. G. (2003). Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature, 422*(6930), 442-446.

Forte, A. V., Ignatov, A. N., Ponomarenko, V. V., Dorokhov, D. B., and Savel'ev, N. I. (2002). Phylogeny of the *Malus* (apple tree) species, inferred from its morphological traits and molecular DNA analysis. *Genetika, 38*(10), 1357-1369.

Fox, B. C., Devonshire, A. S., Baradez, M. O., Marshall, D., and Foy, C. A. (2012). Comparison of RT-qPCR methods and platforms for single cell gene expression analysis. *Analytical Biochemistry, 427*(2), 178-264.

Foyer, C. H., and Shigeoka, S. (2011). Understanding oxidative stress and antioxidant functions to enhance photosynthesis. *Plant Physiology, 155*(1), 93-100.

Fukuda, Y., Aguilar-Bryan, L., Vaxillaire, M., Dechaume, A., Wang, Y., Dean, M., Moitra, K., Bryan, J., and Schuetz, J. D. (2011). Conserved intramolecular disulfide bond is critical to trafficking and fate of ATP-binding cassette (ABC) transporters ABCB6 and sulfonylurea receptor 1 (SUR1)/ABCC8. *Journal of Biological Chemistry, 286*(10), 8481-8492.

Fullwood, M., Wei, C., Liu, E., and Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research, 19*(4), 521-553.

Gao, M., Tao, R., Miura, K., Dandekar, A. M., and Sugiura, A. (2001). Transformation of Japanese persimmon (Diospyros kaki Thunb.) with apple cDNA encoding NADP-dependent sorbitol-6-phosphate dehydrogenase. *Plant Science, 160*(5), 837-845.

Gao, Z., Maurousset, L., Lemoine, R., Yoo, S. D., van Nocker, S., and Loescher, W. (2003). Cloning, expression, and characterization of sorbitol transporters from developing sour cherry fruit and leaf sink tissues. *Plant Physiology, 131*(4), 1566-1575.

Gasic, K. G., Thimmapuram, D. O., Liu, J., Malnoy, L., Gong, M., Han, G., Vodkin, Y., Aldwinckle, L. O., Carroll, H. S., and Orvis, N. J. (2009). Comparative analysis and functional annotation of a large expressed sequence tag collection of apple. *The Plant Genome, 2*(1), 23-38.

Gasser, C. S., and Dean, C. (2009). Growth and development: a broad view of fine detail. *Current Opinion in Plant Biology, 12*(1), 1-3.

Geigenberger, P. (2011). Regulation of starch biosynthesis in response to a fluctuating environment. *Plant Physiology, 155*(4), 1566-1577.

Gill, R. W., and Sanseau, P. (2000). Rapid in silico cloning of genes using expressed sequence tags (ESTs). *Biotechnology Annual Review, 5*, 25-44.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: a developmental perspective. *The Plant Cell, 5*(10), 1439–1451.

Giovannoni, J. (2001). Molecular biology of fruit maturation and ripening. *Annual Review of Plant Biology, 52*(1), 725-749.

Giovannoni, J., DellaPenna, D., Bennett, A., and Fischer, R. (1989). Expression of a chimeric polygalacturonase gene in transgenic rin (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. *The Plant Cell Online, 1*(1), 53-63.

Giovannoni, J. J. (2004). Genetic regulation of fruit development and ripening. *The Plant Cell Online, 16*(suppl_1), S170-S180.

Gomi, K., and Matsuoka, M. (2003). Gibberellin signalling pathway. *Current Opinion in Plant Biology, 6*(5), 489-493.

Gonkiewicz, A. (2011). Ethylene evolution intensity and apple fruit setting after thinning with NAA and ethephon. *Journal of Fruit and Ornamental Plant Research, 19*(1), 63-72.

Gordon, A. (2011). FASTX-Toolkit. *Computer program distributed by the author, website http://hannonlab. cshl. edu/fastx_toolkit/index. html [accessed 15 July 2011].*

Gou, J.-Y., Wang, L.-J., Chen, S.-P., Hu, W.-L., and Chen, X.-Y. (2007). Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Research, 17*(5), 422-434.

Goulao, L. F., and Oliveira, C. M. (2008). Cell wall modifications during fruit ripening: when a fruit is not the fruit. *Trends in Food Science & Technology, 19*(1), 4-25.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., and Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology, 29*(7), 644-652.

Gray, J., Picton, S., Shabbeer, J., Schuch, W., and Grierson, D. (1992). Molecular biology of fruit ripening and its manipulation with antisense genes. *Plant Molecular Biology, 19*(1), 69-87.

Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harbor Perspectives in Biology, 4*(9).

Green, S., Friel, E. N., Matich, A., Beuning, L. L., Cooney, J. M., Rowan, D. D., and MacRae, E. (2007). Unusual features of a recombinant apple alpha-farnesene synthase. *Phytochemistry, 68*(2), 176-188.

Grimplet, J., Deluc, L. G., Tillett, R. L., Wheatley, M. D., Schlauch, K. A., Cramer, G. R., and Cushman, J. C. (2007). Tissue-specific mRNA expression profiling in grape berry tissues. *BMC Genomics, 8*(1), 187-209.

Gucci, R., Corelli, G., Tustin, S., and Ravaglia, G. (1995). The effect of defruiting at different stages of fruit development on leaf photosynthesis of "Golden Delicious" apple. *Tree Physiology, 15*(1), 35-40.

Gust, A. A., and Nürnberger, T. (2012). Plant immunology: A life or death switch. *Nature, 486*(7402), 198-199.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., and Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols, 8*(8), 1494-1512.

Hagen, G., Guilfoyle, T. J., and Gray, W. M. (2010). Auxin signal transduction. *Plant Hormones*, 282-307.

Hager, A. (2003). Role of the plasma membrane H+-ATPase in auxin-induced elongation growth: historical and new aspects. *Journal of Plant Research, 116*(6), 483-505.

Han, S. E., Seo, Y. S., Kim, D., Sung, S. K., and Kim, W. T. (2007). Expression of MdCAS1 and MdCAS2, encoding apple beta-cyanoalanine synthase homologs, is concomitantly induced during ripening and implicates MdCASs in the possible role of the cyanide detoxification in Fuji apple (Malus domestica Borkh.) fruits. *Plant Cell Reports, 26*(8), 1321-1331.

Han, Y., Bendik, E., Sun, F. J., Gasic, K., and Korban, S. S. (2007). Genomic isolation of genes encoding starch branching enzyme II (SBEII) in apple: toward characterization of evolutionary disparity in SbeII genes between monocots and eudicots. *Planta, 226*(5), 1265-1276.

Hancock, J. F. (2012). *Plant evolution and the origin of crop species*: Cab International.

Harb, J., Gapper, N. E., Giovannoni, J. J., and Watkins, C. B. (2012). Molecular analysis of softening and ethylene synthesis and signaling pathways in a non-softening apple cultivar,'Honeycrisp' and a rapidly softening cultivar, 'McIntosh'. *Postharvest Biology and Technology, 64*(1), 94-103.

Harris, S. A., Robinson, J. P., and Juniper, B. E. (2002). Genetic clues to the origin of the apple. *Trends in Genetics, 18*(8), 426-430.

Hattasch, C., Flachowsky, H., Kapturska, D., and Hanke, M. V. (2008). Isolation of flowering genes and seasonal changes in their transcript levels related to flower induction and initiation in apple (*Malus domestica*). *Tree Physiology, 28*(10), 1459-1466.

Hauvermale, A. L., Ariizumi, T., and Steber, C. M. (2012). Gibberellin Signaling: A Theme and Variations on DELLA Repression. *Plant Physiology, 160*(1), 83-92.

He, L., Ban, Y., Miyata, S., Kitashiba, H., and Moriguchi, T. (2008). Apple aminopropyl transferase, MdACL5 interacts with putative elongation factor 1-alpha and S-adenosylmethionine synthase [corrected]. *Biochemical and Biophysical Research Communications, 366*(1), 162-167.

Hedden, P. (2012). Gibberellin biosynthesis. *Encyclopedia of Life Sciences, DOI: 10.1002/9780470015902.a0023720*.

Henson, J., Tischler, G., and Ning, Z. (2012). Next-generation sequencing and large genome assemblies. *Pharmacogenomics, 13*(8), 901-915.

Higuchi, R., Fockler, C., Dollinger, G., and Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (NY), 11*, 1026-1030.

Hirayama, T., and Shinozaki, K. (2007). Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. *Trends in Plant Science, 12*(8), 343-351.

Hiremath, P. J., Farmer, A., Cannon, S. B., Woodward, J., Kudapa, H., Tuteja, R., Kumar, A., BhanuPrakash, A., Mulaosmanovic, B., and Gujaria, N. (2012). Large-scale transcriptome analysis in chickpea *(Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnology Journal, 9*(8), 922-931.

Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999). The PROSITE database, its status in 1999. *Nucleic Acids Research, 27*(1), 215-219.

Hrycyna, C., Bergo, M., and Tamanoi, F. (2011). *Protein Prenylation*: Academic Press.

Hu, J., Baker, A., Bartel, B., Linka, N., Mullen, R. T., Reumann, S., and Zolman, B. K. (2012). Plant peroxisomes: biogenesis and function. *The Plant Cell Online, 24*(6), 2279-2303.

Hu, M., and Polyak, K. (2006). Serial analysis of gene expression. *Nature Protocols, 1*(4), 1743-1760.

Hulme, A. C., Rhodes, M. J., Galliard, T., and Wooltorton, L. S. (1968). Metabolic Changes in Excised Fruit Tissue. IV. Changes Occurring in Discs of Apple Peel During the Development of the Respiration Climacteric. *Plant Physiology, 43*(7), 1154-1161.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., and Duquenne, L. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Research, 37*(suppl 1), D211-D215.

Idury, R. M., and Waterman, M. S. (1995). A new algorithm for DNA sequence assembly. *Journal of Computational Biology, 2*(2), 291-306.

Inze, D. (2003). Why should we study the plant cell cycle? *Journal of Experimental Botany, 54*(385), 1125-1131.

Iqbal, N., Nazar, R., Khan, M. I. R., Masood, A., and Khan, N. A. (2011). Role of gibberellins in regulation of source-sink relations under optimal and limiting environmental conditions. *Current Science(Bangalore), 100*(7), 998-1007.

Iseli, C., Jongeneel, C. V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *7*, 138-148.

Jahns, P., and Holzwarth, A. R. (2012). The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1817*(1), 182-193.

Janick, J. (1972). *Horticultural Science*. San Francisco: WH Freeman.

Janick, J. (1986). *Horticultural Science*. San Francisco: WH Freeman.

Janick, J., Schery, R. W., Woods, F. W., and Ruttan, V. W. (1981). *Plant Science, An introduction to world crops* (Third edition ed.). San Francisco: WH Freeman and company.

Janicka-Russak, M. (2011). Plant plasma membrane H+-ATPase in adaptation of plants to abiotic stresses. *Abiotic stress response in Plants-physiological, biochemical and genetic perspectives. Rijeka, Croatia: InTech*, 197-218.

Janisiewicz, W. J., and Korsten, L. (2002). Biological control of postharvest diseases of fruits. *Annual Review of Phytopathology, 40*(1), 411-441.

Janssen, B. J., Thodey, K., Schaffer, R. J., Alba, R., Balakrishnan, L., Bishop, R., Bowen, J. H., Crowhurst, R. N., Gleave, A. P., Ledger, S., McArtney, S., Pichler, F. B., Snowden, K. C., and Ward, S. (2008). Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. *BMC Plant Biology, 8*, 16-44.

Jomaa, H., Eberl, M., and Altincicek, B. (2011). Inactivation of genes of the MEP pathway, *Patent US7875279*: Google Patents.

Jones, B., Gunnerås, S. A., Petersson, S. V., Tarkowski, P., Graham, N., May, S., Dolezal, K., Sandberg, G., and Ljung, K. (2010). Cytokinin regulation of auxin synthesis in Arabidopsis involves a homeostatic feedback loop regulated via auxin and cytokinin signal transduction. *The Plant Cell Online, 22*(9), 2956-2969.

Jones, D. C., Ruzzo, W. L., Peng, X., and Katze, M. G. (2012). A new approach to bias correction in RNA-Seq. *Bioinformatics, 28*(7), 921-928.

Juniper, B. E., and Mabberley, D. J. (2006). *The story of the apple*: Timber Press.

Juniper, B. E., Watkins, R., and Harris, S. A. (1996). The origin of the apple. *In Eucarpia Symposium on Fruit Breeding and Genetics* (Vol. 484, pp. 27-34). : ISHS.

Kahle, K., Kraus, M., and Richling, E. (2005). Polyphenol profiles of apple juices. *Molecular Nutrition and Food Research, 49*(8), 797-806.

Kakimoto, T. (2003). Perception and signal transduction of cytokinins. *Annual Review of Plant Biology, 54*(1), 605-627.

Kanayama, Y., Mori, H., Imaseki, H., and Yamaki, S. (1992). Nucleotide sequence of a cDNA encoding NADP-sorbitol-6-phosphate dehydrogenase from apple. *Plant Physiology, 100*(3), 1607–1608.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Research, 32*(suppl 1), D277-D280.

Kaushik, R., Vashist, M., Jain, S., Sikka, V., and Kumar, S. (2012). Subtle structural differences crucial for function in similarly engineered ADP-glucose pyrophosphorylase larger subunit in rice and maize. *Journal of Plant Biochemistry and Biotechnology, 21*(2), 275-278.

Kellerhals, M. (2009). Introduction to Apple (Malus x domestica). *Genetics and Genomics of Rosaceae*, 73-84.

Kim, T. W., and Wang, Z. Y. (2010). Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annual Review of Plant Biology, 61*, 681-704.

Klee, H. J., and Clark, D. G. (2010). Ethylene signal transduction in fruits and flowers *Plant Hormones* (pp. 377-398). Netherlands: Springer

Kleine-Vehn, J., Dhonukshe, P., Swarup, R., Bennett, M., and Friml, J. (2006). Subcellular trafficking of the *Arabidopsis* auxin influx carrier AUX1 uses a novel pathway distinct from PIN1. *The Plant Cell Online, 18*(11), 3171-3181.

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends in Plant Science, 10*(5), 236-242.

Kohorn, B. D., and Kohorn, S. L. (2012). The Cell Wall-Associated Kinases, WAKs, Regulate Cell Expansion and the Stress Response. *Receptor-like Kinases in Plants*, 109-124.

Kondo, S., Tomiyama, H., Kittikorn, M., Okawa, K., Ohara, H., Yokoyama, M., Ifuku, O., Saito, T., Ban, Y., and Tatsuki, M. (2012). Ethylene production and 1-aminocyclopropane-1-carboxylate (ACC) synthase and ACC oxidase gene expression in apple fruit are affected by 9, 10-ketol-octadecadienoic acid (KODA). *Postharvest Biology and Technology, 72*, 20-26.

Korban, S. S., and Skrivin, R. M. (1984). Nomenclature of the cultivated apple. *Hort Science, 19*, 177-180.

Kosma, D. K., Parsons, E. P., Isaacson, T., Lü, S., Rose, J. K. C., and Jenks, M. A. (2010). Fruit cuticle lipid composition during development in tomato ripening mutants. *Physiologia Plantarum, 139*(1), 107-117.

Krost, C., Petersen, R., and Schmidt, E. R. (2012). The transcriptomes of columnar and standard type apple trees (*Malus x domestica*) - A comparative study. *Gene, 498*(2), 223-253.

Krumpochova, P., Sapthu, S., Brouwers, J. F., de Haas, M., de Vos, R., Borst, P., and van de Wetering, K. (2012). Transportomics: screening for substrates of ABC transporters in body fluids using vesicular transport assays. *The FASEB Journal, 26*(2), 738-747.

Kühn, B. F., Bertelsen, M., and Sørensen, L. (2011). Optimising quality-parameters of apple cv., Pigeon, by adjustment of nitrogen. *Scientia Horticulturae, 129*(3), 369-375.

Kühn, C., and Grof, C. P. L. (2010). Sucrose transporters of higher plants. *Current Opinion in Plant Biology, 13*(3), 287-297.

Kunst, L., and Samuels, L. (2009). Plant cuticles shine: advances in wax biosynthesis and export. *Current Opinion in Plant Biology, 12*(6), 721-727.

Kuzyakov, Y., and Gavrichkova, O. (2011). REVIEW: Time lag between photosynthesis and carbon dioxide efflux from soil: A review of mechanisms and controls. *Global Change Biology, 16*(12), 3386-3406.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology, 10*(3), R25.

Lata, B., and Tomala, K. (2007). Apple peel as a contributor to whole fruit quantity of potentially healthful bioactive compounds. Cultivar and year implication. *Journal of Agricultural and Food Chemistry, 55*(26), 10795-10802.

Lau, J., Cooper, N., Robinson, D., and Korban, S. (2009). Sequence and In Silico Characterization of the Tomato Polygalacturonase (PG) Promoter and Terminator Regions. *Plant Molecular Biology Reporter, 27*(3), 250-256.

Le, D. T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L. S. P. (2011). Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Research, 18*(4), 263-276.

Lee, Y. P., Yu, G. H., Seo, Y. S., Han, S. E., Choi, Y. O., Kim, D., Mok, I. G., Kim, W. T., and Sung, S. K. (2007). Microarray analysis of apple gene expression engaged in early fruit development. *Plant Cell Reports, 26*(7), 917-926.

Leisso, R., Buchanan, D., Lee, J., Mattheis, J., and Rudell, D. (2013). Cell wall, cell membrane, and volatile metabolism are altered by antioxidant treatment, temperature shifts, and peel necrosis during apple fruit storage. *Journal of Agricultural and Food Chemistry, 61*(6), 1373-1387.

Lelièvre, J. M., Latche, A., Jones, B., Bouzayen, M., and Pech, J. C. (1997). Ethylene and fruit ripening. *Physiologia Plantarum, 101*(4), 727-739.

León, P., and Cordoba, E. (2012). Understanding the Mechanisms that Modulate the MEP Pathway *Isoprenoid Synthesis in Plants and Microorganisms: New Concepts and Experimental Approaches* (pp. 457-464). New York: Springer

Lewis, P. A., and Alessi, D. R. (2012). Deciphering the function of leucine-rich repeat kinase 2 and targeting its dysfunction in disease1. *Biochemical Society Transactions, 40*(5), 1039-1041.

Leyser, O. (2006). Dynamic integration of auxin transport and signalling. *Current Biology, 16*(11), R424-R433.

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*(1), 323.

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., and Dewey, C. N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics, 26*(4), 493-500.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, M., Feng, F., and Cheng, L. (2012). Expression patterns of genes involved in sugar metabolism and accumulation during apple fruit development. *PLoS One, 7*(3), e33055.

Liere, K., Weihe, A., and Börner, T. (2011). The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. *Journal of Plant Physiology, 168*(12), 1345-1360.

Lin, F., Ding, H., Wang, J., Zhang, H., Zhang, A., Zhang, Y., Tan, M., Dong, W., and Jiang, M. (2009). Positive feedback regulation of maize NADPH oxidase by mitogen-activated protein kinase cascade in abscisic acid signalling. *Journal of Experimental Botany, 60*(11), 3221-3238.

Lin-Wang, K., Micheletti, D., Palmer, J., Volz, R., Lozano, L., Espley, R., Hellens, R. P., Chagne, D., Rowan, D. D., and Troggio, M. (2011). High temperature reduces apple fruit colour via modulation of the anthocyanin regulatory complex. *Plant, Cell and Environment, 34*(7), 1176-1265.

Lippold, F., vom Dorp, K., Abraham, M., Hölzl, G., Wewer, V., Yilmaz, J. L., Lager, I., Montandon, C., Besagni, C., and Kessler, F. (2012). Fatty Acid Phytyl Ester Synthesis in Chloroplasts of *Arabidopsis*. *The Plant Cell Online, 24*(5), 2001-2014.

Lister, C. E., Lancaster, J. E., Sutton, K. H., and Walker, J. R. L. (2006). Developmental changes in the concentration and composition of flavonoids in skin of a red and a green apple cultivar. *Journal of the Science of Food and Agriculture, 64*(2), 155-161.

Lu, Y., Wang, L., Liu, P., Yang, P., and You, M. (2012). Gene-expression signature predicts postoperative recurrence in stage I non-small cell lung cancer patients. *PLoS One, 7*(1), e30880.

MacDaniels, L. H. (1940). The morphology of the apple and other pome fruits. *Cornell University, Agricultural Experiment Station Memoir, 230*, 1-32.

Malladi, A., and Johnson, L. K. (2011). Expression profiling of cell cycle genes reveals key facilitators of cell production during carpel development, fruit set, and fruit growth in apple (Malus x domestica Borkh.). *Journal of Experimental Botany, 62*(1), 205–219.

Manzoor, M., Anwar, F., Saari, N., and Ashraf, M. (2012). Variations of Antioxidant Characteristics and Mineral Contents in Pulp and Peel of Different Apple (Malus domestica Borkh.) Cultivars from Pakistan. *Molecules, 17*(1), 390-407.

Mardis, E. R. (2008a). The impact of next-generation sequencing technology on genetics. *Trends in Genetics, 24*(3), 133-141.

Mardis, E. R. (2008b). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics, 9*, 387-402.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., and Chen, Z. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*(7057), 376-380.

Marondedze, C., and Thomas, L. A. (2011). Insights into fruit function from the proteome of the hypanthium. *Journal of Plant Physiology, 169*(1), 12-19.

Marondedze, C., and Thomas, L. A. (2012). Apple Hypanthium Firmness: New Insights from Comparative Proteomics. *Applied Biochemistry and Biotechnology, 168* (2), 306-326.

Martin, J. A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics, 12*, 671-682.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal, 17*(1), 10-12.

Masia, A., Ventura, M., Gemma, H., and Sansavini, S. (1998). Effect of some plant growth regulator treatments on apple fruit ripening. *Plant Growth Regulation, 25*(2), 127-134.

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. . *Nucleic acids research 40*(10), 4288-4297.

McGlasson, W. B., and Hulme, A. C. (1970). *The Biochemistry of Fruits and their Products*. New York: Academic Press.

McIntyre, L. M., Lopiano, K. K., Morse, A. M., Amin, V., Oberg, A. L., Young, L. J., and Nuzhdin, S. V. (2011). RNA-seq: technical variability and sampling. *BMC Genomics, 12*(1), 293.

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiology, 138*(2), 1083-1096.

Merkle, T. (2011). Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant Cell Reports, 30*(2), 153-176.

Merriman, B., Torrent, I., and Rothberg, J. M. (2012). Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis, 33*(23), 3397-3417.

Merzlyak, M. N., Melø, T. B., and Naqvi, K. R. (2008). Effect of anthocyanins, carotenoids, and flavonols on chlorophyll fluorescence excitation spectra in apple fruit: signature analysis, assessment, modelling, and relevance to photoprotection. *Journal of Experimental Botany, 59*(2), 349-359.

Merzlyak, M. N., Solovchenko, A. E., Smagin, A. I., and Gitelson, A. A. (2005). Apple flavonols during fruit adaptation to solar radiation: spectral features and technique for non-destructive assessment. *Journal of Plant Physiology, 162*(2), 151-160.

Mewes, H., Dietmann, S., Frishman, D., Gregory, R., Mannhaupt, G., Mayer, K., Münsterkötter, M., Ruepp, A., Spannagl, M., and Stümpflen, V. (2008). MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Research, 36*(suppl 1), D196–D201.

Mewes, H. W., Frishman, D., Güldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Münsterkötter, M., Rudd, S., and Weil, B. (2002). MIPS: a database for genomes and protein sequences. *Nucleic Acids Research, 30*(1), 31.

Meyers, B. C., Vu, T. H., Tej, S. S., Ghazal, H., Matvienko, M., Agrawal, V., Ning, J., and Haudenschild, C. D. (2004). Analysis of the transcriptional complexity of Arabidopsis thaliana by massively parallel signature sequencing. *Nature Biotechnology, 22*(8), 1006-1011.

Milborrow, B. (2001). The pathway of biosynthesis of abscisic acid in vascular plants: a review of the present state of knowledge of ABA biosynthesis. *Journal of Experimental Botany, 52*(359), 1145-1164.

Miller, J. R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics, 95*(6), 315-327.

Min, X. J., Butler, G., Storms, R., and Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research, 33*(suppl 2), W677-W680.

Minoche, A. E., Dohm, J. C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology, 12*(11), R112.

Mintz-Oron, S., Mandel, T., Rogachev, I., Feldberg, L., Lotan, O., Yativ, M., Wang, Z., Jetter, R., Venger, I., and Adato, A. (2008). Gene expression and metabolism in tomato fruit surface tissues. *Plant Physiology, 147*(2), 823-851.

Mizuno, H., Kawahigashi, H., Kawahara, Y., Kanamori, H., Ogata, J., Minami, H., Itoh, T., and Matsumoto, T. (2012). Global transcriptome analysis reveals distinct expression among duplicated genes during sorghum-Bipolarissorghicola interaction. *BMC Plant Biology, 12*(1), 121.

Monagas, M., Urpi-Sarda, M., Sánchez-Patán, F., Llorach, R., Garrido, I., Gómez-Cordovés, C., Andres-Lacueva, C., and Begoña Bartolomé, B. (2010). Insights into the metabolism and microbial biotransformation of dietary flavan-3-ols and the bioactivity of their metabolites. *Food and Function, 1*(3), 233-253.

Moore, R., Clark, W. D., and Vodopich, D. S. (1998). Botany New York, USA: WCB: McGraw-Hill.

Moore, S., Vrebalov, J., Payton, P., and Giovannoni, J. (2002). Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. *Journal of Experimental Botany, 53*(377), 2023-2030.

Morandi, B., Zibordi, M., Losciale, P., Manfrini, L., Pierpaoli, E., and Grappadelli, L. C. (2011). Shading decreases the growth rate of young apple fruit by reducing their phloem import. *Scientia Horticulturae, 127*(3), 347-352.

Morozova, O., Hirst, M., and Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics, 10*, 135-151.

Muchuweti, M., and Chikwambi, Z. (2008). Isolation and Identification of Anthocyanins in the Fruit Peels of Starkrimson and Marx Red Bartlett Common Pear Cultivars and Their Bud Mutants. *American Journal of Food Technology, 3*(1), 1-12.

Muench, S. P., Trinick, J., and Harrison, M. A. (2011). Structural divergence of the rotary ATPases. *Quarterly Reviews of Biophysics, 44*(03), 311-356.

Muranaka, T., and Ohyama, K. (2013). Functional Analysis of HMG-CoA Reductase and Oxidosqualene Cyclases in *ArabidopsisIsoprenoid Synthesis in Plants and Microorganisms* (pp. 465-474): Springer.

Myers, E. W. (1995). Toward simplifying and accurately formulating fragment assembly. *Journal of Computational Biology, 2*(2), 275-290.

Naganeeswaran, S. A., Subbian, E. A., and Ramaswamy, M. (2012). Analysis of expressed sequence tags (ESTs) from cocoa (Theobroma cacao L) upon infection with Phytophthora megakarya. *Bioinformation, 8*(2), 65-69.

Nagaraj, S. H., Deshpande, N., Gasser, R. B., and Ranganathan, S. (2007). ESTExplorer: an expressed sequence tag (EST) assembly and annotation platform. *Nucleic Acids Research, 35*(suppl 2), W143-W147.

Namitha, K. K., Archana, S. N., and Negi, P. S. (2011). Expression of carotenoid biosynthetic pathway genes and changes in carotenoids during ripening in tomato (*Lycopersicon esculentum*). *Food and Function, 2*(3-4), 168-173.

Nemhauser, J. L., Mockler, T. C., and Chory, J. (2004). Interdependency of brassinosteroid and auxin signaling in *Arabidopsis*. *PLoS Biology, 2*(9), e258.

Neuspiel, M., Schauss, A. C., Braschi, E., Zunino, R., Rippstein, P., Rachubinski, R. A., Andrade-Navarro, M. A., and McBride, H. M. (2008). Cargo-selected transport from the mitochondria to peroxisomes is mediated by vesicular carriers. *Current Biology, 18*(2), 102-108.

Newcomb, R. D., Crowhurst, R. N., Gleave, A. P., Rikkerink, E. H., Allan, A. C., Beuning, L. L., Bowen, J. H., Gera, E., Jamieson, K. R., Janssen, B. J., Laing, W. A., McArtney, S., Nain, B., Ross, G. S., Snowden, K. C., Souleyre, E. J., Walton, E. F., and Yauk, Y. K. (2006). Analyses of expressed sequence tags from apple. *Plant Physiol, 141*(1), 147-166.

Nitsch, J. P. (1951). Growth and development in vitro of excised ovaries. *American Journal of Botany, 38*(7), 566-577.

Nobusawa, T., Okushima, Y., Nagata, N., Kojima, M., Sakakibara, H., and Umeda, M. (2013). Synthesis of Very-Long-Chain Fatty Acids in the Epidermis Controls Plant Organ Growth by Restricting Cell Proliferation. *PLoS Biology, 11*(4), e1001531.

Nookaraju, A., Upadhyaya, C. P., Pandey, S. K., Young, K. E., Hong, S. J., Park, S. K., and Park, S. W. (2010). Molecular approaches for enhancing sweetness in fruits and vegetables. *Scientia Horticulturae, 127*(1), 1–15.

Norelli, J. L., Farrell, R. E., Bassett, C. L., Baldo, A. M., Lalli, D. A., Aldwinckle, H. S., and Wisniewski, M. E. (2009). Rapid transcriptional response of apple to fire blight disease revealed by cDNA suppression subtractive hybridization analysis. *Tree Genetics and Genomes, 5*(1), 27-40.

Olofsson, L., Engstrom, A., Lundgren, A., and Brodelius, P. (2011). Relative expression of enzymes of terpene metabolism in different tissues of Artemisia annua L. *BMC Plant Biology, 11*(1), 45.

Oren-Shamir, M. (2009). Does anthocyanin degradation play a significant role in determining pigment concentration in plants? *Plant Science, 177*(4), 310-316.

Ortiz, A., Graell, J., and Lara, I. (2011a). Cell wall-modifying enzymes and firmness loss in ripening ' Golden Reinders' apples: a comparison between calcium dips and ulo storage. *Food Chemistry, 128*(4), 1072–1079.

Ortiz, A., Graell, J., and Lara, I. (2011b). Preharvest calcium applications inhibit some cell wall-modifying enzyme activities and delay cell wall disassembly at commercial harvest of 'Fuji Kiku-8' apples. *Postharvest Biology and Technology, 62*(2), 161–167.

Palma, J. M., Corpas, F. J., and del Río, L. A. (2011). Proteomics as an approach to the understanding of the molecular physiology of fruit development and ripening. *Journal of Proteomics, 74*(8), 1230-1243.

Pandey, V., Nutter, R. C., and Prediger, E. (2008). Applied Biosystems SOLiD System: Ligation-Based Sequencing. In M. Janitz (Ed.), *Next Generation Genome Sequencing: Towards Personalized Medicine* (pp. 29-42). Chicago: Published Online: 30 OCT. DOI: 10.1002/9783527625130.ch3.

Park, S., Sugimoto, N., Larson, M. D., Beaudry, R., and van Nocker, S. (2006). Identification of genes with potential roles in apple fruit development and biochemistry through large-scale statistical analysis of expressed sequence tags. *Plant Physiology, 141*(3), 811-824.

Park, Y. B., and Cosgrove, D. J. (2012). Changes in cell wall biomechanical properties in the xyloglucan-deficient xxt1/xxt2 mutant of Arabidopsis. *Plant Physiology, 158*(1), 465-475.

Parkinson, J., and Blaxter, M. (2009). Expressed sequence tags: An overview. *Methods in Molecular Biology (Clifton, NJ), 533*, 1-12.

Passam, H. C., Karapanos, I. C., and Alexopoulos, A. A. (2011). The Biological Basis of Fruit Quality. *In Breeding for Fruit Quality,* (pp. 3 - 38). Athens, Greece: Purdue University, Center for Plant Environmental Stress Physiology, West Lafayette, Indiana; Agricultural University of Athens, Laboratory of Plant Breeding and Biometry.

Peaucelle, A., Braybrook, S., and Höfte, H. (2012). Cell wall mechanics and growth control in plants: the role of pectins revisited. *Frontiers in Plant Science, 3*, 121.

Pedersen, C. N. S., Axelsen, K. B., Harper, J. F., and Palmgren, M. G. (2012). Evolution of plant P-type ATPases. *Frontiers in Plant Science, 3*, 31.

Perez-Gil, J., Uros, E. M., Sauret-Güeto, S., Lois, L. M., Kirby, J., Nishimoto, M., Baidoo, E. E. K., Keasling, J. D., Boronat, A., and Rodriguez-Concepcion, M. (2012). Mutations in Escherichia coli aceE and ribB Genes Allow Survival of Strains Defective in the First Step of the Isoprenoid Biosynthesis Pathway. *PLoS One, 7*(8), e43775.

Peter, E., Wallner, T., Wilde, A., and Grimm, B. (2011). Comparative functional analysis of two hypothetical chloroplast open reading frames (ycf) involved in chlorophyll biosynthesis from *Synechocystis* sp. PCC6803 and plants. *Journal of Plant Physiology, 168*(12), 1380-1386.

Pieterse, C. M. J., Van der Does, D., Zamioudis, C., Leon-Reyes, A., and Van Wees, S. C. M. (2012). Hormonal modulation of plant immunity. *Annual Review of Cell and Developmental Biology, 28*, 28.21-28.33.

Pinheiro, C., and Chaves, M. (2011). Photosynthesis and drought: can we make metabolic connections from available data? *Journal of Experimental Botany, 62*(3), 869-882.

Piwien-Pilipuk, G., Huo, J. S., and Schwartz, J. (2011). Growth hormone signal transduction. *Journal of Pediatric Endocrinology and Metabolism, 15*(6), 771-786.

Prasanna, V., Prabha, T., and Tharanathan, R. (2007). Fruit ripening phenomena-an overview. *Critical Reviews in Food Science and Nutrition, 47*(1), 1-19.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., and Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics, 13*(1), 341.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842.

R Core Team,. (2013). A Language and Environment for Statistical Computing. R Foundation for Staistical Computing, Vienna, Austria, http//www.R-progect.org.

Rajjou, L., Duval, M., Gallardo, K., Catusse, J., Bally, J., Job, C., and Job, D. (2012). Seed germination and vigor. *Annual Review of Plant Biology, 63*, 507-533.

Ranganathan, S., Menon, R., and Gasser, R. B. (2009). Advanced in silico analysis of expressed sequence tag (EST) data for parasitic nematodes of major socio-economic importance--fundamental insights toward biotechnological outcomes. *Biotechnology Advances, 27*(4), 439-448.

Raven, P. H., Evert, R. F., and Eichhorn, S. E. (1999). Biology of Plants (sixth edition ed.). New York: WH Freeman and Company.

Reinartz, J., Bruyns, E., Lin, J. Z., Burcham, T., Brenner, S., Bowen, B., Kramer, M., and Woychik, R. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics and Proteomics, 1*(1), 95-104.

Reis, S. F., Rocha, S. l. M., Barros, A. n. S., Delgadillo, I., and Coimbra, M. A. (2009). Establishment of the volatile profile of 'Bravo de Esmolfe' apple variety and identification of varietal markers. *Food Chemistry, 113*(2), 513-521.

Rhee, Y., Hlousek-Radojcic, A., Ponsamuel, J., Liu, D., and Post-Beittenmiller, D. (1998). Epicuticular wax accumulation and fatty acid elongation activities are induced during leaf development of leeks. *Plant Physiology, 116*(3), 901-911.

Ribaut, J. M. (2006). *Drought adaptation in cereals*. Routledge, Chicago: CRC.

Riederer, M., and Burghardt, M. (2006). Cuticular transpiration. *Biology of the Plant Cuticle. Blackwell Publishing, Oxford, UK, 23*, 292-311.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology, 29*(1), 24-26.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Roongsattham, P., Morcillo, F., Jantasuriyarat, C., Pizot, M., Moussu, S., Jayaweera, D., Collin, M., Gonzalez-Carranza, Z., Amblard, P., and Tregear, J. W. (2012). Temporal and spatial expression of polygalacturonase gene family members reveals divergent regulation during fleshy fruit ripening and abscission in the monocot species oil palm. *BMC Plant Biology, 12*(1), 150.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., and Edwards, M. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature, 475*(7356), 348-352.

Ruan, Y., Le Ber, P., Hui Ng, H., and Liu, E. T. (2004). Interrogating the transcriptome. *Trends in Biotechnology, 22*(1), 23-30.

Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Güldener, U., Mannhaupt, G., and Münsterkötter, M. (2004). The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Research, 32*(18), 5539-5545.

Rushton, D. L., Tripathi, P., Rabara, R. C., Lin, J., Ringler, P., Boken, A. K., Langum, T. J., Smidt, L., Boomsma, D. D., and Emme, N. J. (2011). WRKY transcription factors: key components in abscisic acid signalling. *Plant Biotechnology Journal, 10*(1), 2-11.

Rutkowski, K. P., Michalczuk, B., and Konopacki, P. (2008). Nondestructive determination of 'Golden Delicious' apple quality and harvest maturity. *Journal of Fruit and Ornamental Plant Research, 16*, 39-52.

Saeed, A., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., and Thiagarajan, M. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques, 34*(2), 374-378.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA, 74*(12), 5463–5467.

Saroussi, S., and Nelson, N. (2009). Vacuolar H+-ATPase, An enzyme for all seasons. *Pflugers Archiv European Journal of Physiology, 457*(3), 581-587.

Sayanova, O., Ruiz-Lopez, N., Haslam, R. P., and Napier, J. A. (2012). The role of Δ6- desaturase acyl- carrier specificity in the efficient synthesis of long- chain polyunsaturated fatty acids in transgenic plants. *Plant Biotechnology Journal, 10*(2), 195-206.

Schaffer, R. J., Friel, E. N., Souleyre, E. J., Bolitho, K., Thodey, K., Ledger, S., Bowen, J. H., Ma, J. H., Nain, B., Cohen, D., Gleave, A. P., Crowhurst, R. N., Janssen, B. J., Yao, J. L., and Newcomb, R. D. (2007). A genomics approach reveals that aroma production in apple is controlled by ethylene predominantly at the final step in each biosynthetic pathway. *Plant Physiology, 144*(4), 1899-1912.

Schaller, G. E., Shiu, S. H., and Armitage, J. P. (2011). Two-component systems and their co-option for eukaryotic signal transduction. *Current Biology, 21*(9), R320-R330.

Schröter, Y., Steiner, S., Matthäi, K., and Pfannschmidt, T. (2010). Analysis of oligomeric protein complexes in the chloroplast sub, proteome of nucleic acid, binding proteins from mustard reveals potential redox regulators of plastid gene expression. *Proteomics, 10*(11), 2191-2204.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Research, 28*(1), 231-234.

Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics, 28*(8), 1086-1092.

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nature, 5*(1), 16-18.

Seaver, S. M. D., Henry, C. S., and Hanson, A. D. (2012). Frontiers in metabolic reconstruction and modeling of plant genomes. *Journal of Experimental Botany, 63*(6), 2247-2258.

Sekimoto, H., Seo, M., Kawakami, N., Komano, T., Desloire, S., Liotenberg, S., Marion-Poll, A., Caboche, M., Kamiya, Y., and Koshiba, T. (1998). Molecular cloning and characterization of aldehyde oxidases in Arabidopsis thaliana. *Plant and Cell Physiology, 39*(4), 433-442.

Senadheera, P., and Maathuis, F. J. M. (2009). Differentially regulated kinases and phosphatases in roots may contribute to inter-cultivar difference in rice salinity tolerance. *Plant Signaling and Behavior, 4*(12), 1163-1165.

Seo, M., Aoki, H., Koiwai, H., Kamiya, Y., Nambara, E., and Koshiba, T. (2004). Comparative studies on the Arabidopsis aldehyde oxidase (AAO) gene family revealed a major role of AAO3 in ABA biosynthesis in seeds. *Plant and Cell Physiology, 45*(11), 1694-1703.

Seo, M., Hanada, A., Kuwahara, A., Endo, A., Okamoto, M., Yamauchi, Y., North, H., Marion Poll, A., Sun, T., and Koshiba, T. (2006). Regulation of hormone metabolism in Arabidopsis seeds: phytochrome regulation of abscisic acid metabolism and abscisic acid regulation of gibberellin metabolism. *The Plant Journal, 48*(3), 354-366.

Seo, M., and Koshiba, T. (2002). Complex regulation of ABA biosynthesis in plants. *Trends in Plant Science, 7*(1), 41-48.

Seo, Y. S., Kim, S. J., Harn, C. H., and Kim, W. T. (2011). Ectopic expression of apple fruit homogentisate phytyltransferase gene (MdHPT1) increases tocopherol in transgenic tomato (*Solanum lycopersicum* cv. Micro-Tom) leaves and fruits. *Phytochemistry, 72*(4), 321-329.

Seo, Y. S., and Kim, W. T. (2009). A Genomics Approach Using Expressed Sequence Tags and Micro arrays in Ripening Apple Fruit (*Malus domestica* Borkh.). *Journal of Plant Biology, 52*(1), 35-40.

Sharpe, P. L., and DiCosimo, D. J. (2012). Construction and Utilization of Carotenoid Reporter Systems: Identi fication of Chromosomal Integration Sites That Support Suitable Expression of Biosynthetic Genes and Pathways. *Methods in Molecular Biology (Clifton, NJ), 892*, 219-243.

Shay, P. E., and Kubien, D. S. (2012). Field analysis of photoprotection in co-occurring cool climate C3 and C4 grasses. *Physiologia Plantarum, 147*, 316–328.

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D., and Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science, 309*(5741), 1728-1732.

Soglio, V., Costa, F., Molthoff, J. W., Weemen-Hendriks, W. M. J., Schouten, H. J., and Gianfranceschi, L. (2009). Transcription analysis of apple fruit development using cDNA micro arrays. *Tree Genetics and Genomes, 5*(4), 685-698.

Soglio, V., Schouten, H., Costa, F., Perini, D., and Gianfranceschi, L. (2007). Apple fruit: Towards the identification of genes showing modulated expression during development. *Acta Horticulturae, 814*, 675-680.

Solovchenko, A., and Merzlyak, M. (2003). Optical properties and contribution of cuticle to UV protection in plants: experiments with apple fruit. *Photochemistry Photobiology Science, 2*(8), 861-866.

Solovchenko, A. E., Chivkunova, O. B., Gitelson, A. A., and Merzlyak, M. N. (2010). Non-destructive estimation pigment content ripening quality and damage in apple fruit with spectral reflectance in the visible range. *Fresh Produce, 4*(1), 91-102.

Soukoulis, C., Cappellin, L., Aprea, E., Costa, F., Viola, R., Märk, T. D., Gasperi, F., and Biasioli, F. (2012). PTR-ToF-MS, A Novel, Rapid, High Sensitivity and Non-Invasive Tool to Monitor Volatile Compound Release During Fruit Post-Harvest Storage: The Case Study of Apple Ripening. *Food and Bioprocess Technology*, 1-13.

Spartz, A. K., Lee, S. H., Wenger, J. P., Gonzalez, N., Itoh, H., Inzé, D., Peer, W. A., Murphy, A. S., Overvoorde, P. J., and Gray, W. M. (2012). The SAUR19 subfamily of SMALL AUXIN UP RNA genes promote cell expansion. *The Plant Journal, 70*(6), 978–990.

Stepanova, A. N., Yun, J., Robles, L. M., Novak, O., He, W., Guo, H., Ljung, K., and Alonso, J. M. (2011). The Arabidopsis YUCCA1 flavin monooxygenase functions in the indole-3-pyruvic acid branch of auxin biosynthesis. *The Plant Cell Online, 23*(11), 3961-3973.

Stern, D. (2009). *The Chlamydomonas sourcebook: organellar and metabolic processes* (Vol. 2): Access Online via Elsevier.

Strickler, S. R., Bombarely, A., and Mueller, L. A. (2012). Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. *American Journal of Botany, 99*(2), 257-266.

Sugita, M., and Gruissem, W. (1987). Developmental, organ-specific, and light-dependent expression of the tomato ribulose-1, 5-bisphosphate carboxylase small subunit gene family. *Proceedings of the National Academy of Sciences, USA, 84*, 7104-7108.

Sun, Q., Zhou, G., Cai, Y., Fan, Y., Zhu, X., Liu, Y., He, X., Shen, J., Jiang, H., and Hu, D. (2012). Transcriptome analysis of stem development in the tumourous stem mustard Brassica juncea var. tumida Tsen et Lee by RNA sequencing. *BMC Plant Biology, 12*(1), 53.

Sung, S. K., Jeong, D. H., Nam, J., Kim, S. H., Kim, S. R., and An, G. (1998). Expressed sequence tags of fruits, peels, and carpels and analysis of mRNA expression levels of the tagged cDNAs of fruits from the Fuji apple. *Molecules and Cells, 8*(5), 565-577.

Suzuki, Y. and Dandekar, A.M. (2014). Sucrose inducess expression of the sorbitol-6-phosphate dehydrogenase gene in source leaves of loquat. *Physiologia plantarum*, 150(3), 355-362.

Suzuki, N., Miller, G., Morales, J., Shulaev, V., Torres, M. A., and Mittler, R. (2011). Respiratory burst oxidases: the engines of ROS signaling. *Current Opinion in Plant Biology, 14*(6), 691-699.

Suzuki, Y., and Makino, A. (2012). Availability of Rubisco Small Subunit Up-Regulates the Transcript Levels of Large Subunit for Stoichiometric Assembly of Its Holoenzyme in Rice. *Plant Physiology, 160*(1), 533-540.

Tada, H., Mitsui, T., Kiyonaga, T., Akita, T., and Tanaka, K. (2006). All-solid-state Z-scheme in CdS–Au–TiO2 three-component nanojunction system. *Nature Materials, 5*(10), 782-786.

Takos, A. M., Ubi, B. E., Robinson, S. P., and Walker, A. R. (2006). Condensed tannin biosynthesis genes are regulated separately from other flavonoid biosynthesis genes in apple fruit skin. *Plant Science, 170*(3), 487-499.

Tan, D., Li, T., and Wang, A. (2013). Apple 1-aminocyclopropane-1-carboxylic acid synthase genes, MdACS1 and MdACS3a, are expressed in different systems of ethylene biosynthesis. *Plant Molecular Biology Reporter, 31*(1), 204-209.

Tang, W., Kim, T. W., Oses-Prieto, J. A., Sun, Y., Deng, Z., Zhu, S., Wang, R., Burlingame, A. L., and Wang, Z. Y. (2008). BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. *Science Signalling, 321*(5888), 557-560.

Tank, J., and Thaker, V. (2011). Cyclin dependent kinases and their role in regulation of plant cell cycle. *Biologia Plantarum, 55*(2), 201-212.

Tapiero, H., Tew, K., Nguyen Ba, G., and Mathe, G. (2002). Polyphenols: do they play a role in the prevention of human pathologies? *Biomedicine and Pharmacotherapy, 56*(4), 200-207.

Telias, A., Bradeen, J. M., Luby, J. J., Hoover, E. E., and Allen, A. C. (2011). Regulation of anthocyanin accumulation in apple peel. *Horticultural Reviews*, 357-391.

Telias, A., Lin-Wang, K., Stevenson, D. E., Cooney, J. M., Hellens, R. P., Allan, A. C., Hoover, E. E., and Bradeen, J. M. (2011). Apple skin patterning is associated with differential expression of MYB10. *BMC Plant Biology, 11*(1), 93.

Teo, G., Suzuki, Y., Uratsu, S. L., Lampinen, B., Ormonde, N., Hu, W. K., DeJong, T. M., and Dandekar, A. M. (2006). Silencing leaf sorbitol synthesis alters long-distance partitioning and apple fruit quality. *Proceedings of the National Academy of Sciences, USA, 103*(49), 18842-18847.

Tholl, D., and Lee, S. (2011). Elucidating the Metabolism of Plant Terpene Volatiles: Alternative Tools for Engineering Plant Defenses? *The Biological Activity of Phytochemicals*, 159-178.

Toei, M., Saum, R., and Forgac, M. (2010). Regulation and isoform function of the V-ATPases. *Biochemistry, 49*(23), 4715-4723.

Torres, M. A., and Dangl, J. L. (2005). Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. *Current Opinion in Plant Biology, 8*(4), 397-403.

Torti, S., Fornara, F., Vincent, C., Andrés, F., Nordström, K., Göbel, U., Knoll, D., Schoof, H., and Coupland, G. (2012). Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. *The Plant Cell Online, 24*(2), 444-462.

Tozawa, Y., Hasegawa, H., Terakawa, T., and Wakasa, K. (2001). Characterization of Rice Anthranilate Synthase -Subunit Genes OASA1 and OASA2. Tryptophan Accumulation in Transgenic Rice Expressing a Feedback-Insensitive Mutant of OASA1. *Plant Physiology, 126*(4), 1493-1506.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics, 25*(9), 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols, 7*(3), 562-578.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology, 28*(5), 511-515.

Treberg, J. R., Quinlan, C. L., and Brand, M. D. (2011). Evidence for two sites of superoxide production by mitochondrial NADH-ubiquinone oxidoreductase (complex I). *Journal of Biological Chemistry, 286*(31), 27103-27110.

Troggio, M., Gleave, A., Salvi, S., Chagné, D., Cestaro, A., Kumar, S., Crowhurst, R., and Gardiner, S. (2012). Apple, from genome to breeding. *Tree Genetics and Genomes, 8*(3), 509-529.

Tsantili, E., Gapper, N. E., Arquiza, J. M., Whitaker, B. D., and Watkins, C. B. (2007). Ethylene and alpha-farnesene metabolism in green and red skin of three apple cultivars in response to 1-methylcyclopropene (1-MCP) treatment. *Journal of Agricultural and Food Chemistry, 55*(13), 5267-5276.

Tu, Q., Dong, H., Yao, H., Fang, Y., Dai, C., Luo, H., Yao, J., Zhao, D., and Li, D. (2008). Global identification of significantly expressed genes in developing

endosperm of rice by Expressed sequence tags and cDNA array approaches. *Journal of Integrative Plant Biology, 50*(9), 1078-1088.

Ubi, B. E., Honda, C., Bessho, H., Kondo, S., Wada, M., Kobayashi, S., and Moriguchi, T. (2006). Expression analysis of anthocyanin biosynthetic genes in apple skin: effect of UV-B and temperature. *Plant Science, 170*(3), 571-578.

Uematsu, K., Suzuki, N., Iwamae, T., Inui, M., and Yukawa, H. (2012). Expression of Arabidopsis plastidial phosphoglucomutase in tobacco stimulates photosynthetic carbon flow into starch synthesis. *Journal of Plant Physiology, 169*(15), 1454–1462.

Van de Poel, B., Bulens, I., Markoula, A., Hertog, M., Dreesen, R., Wirtz, M., Vandoninck, S., Oppermann, Y., Keulemans, W., and Hell, R. (2012). Targeted systems biology profiling of tomato fruit reveals coordination of the Yang cycle and a distinct regulation of ethylene biosynthesis during post-climacteric ripening. *Plant Physiology, 160*(3), 1498-1514.

Van Rossum, G., and Drake, F. L. (2009). PYTHON 2.6 Reference Manual. *docs.python.org/release/2.6/.*

Vandenbussche, F., Petrásek, J., Zádníková, P., Hoyerová, K., Pesek, B., Raz, V., Swarup, R., Bennett, M., Zazímalová, E., Benková, E., and Van Der Straeten, D. (2010). The auxin influx carriers AUX1 and LAX3 are involved in auxin-ethylene interactions during apical hook development in *Arabidopsis thaliana* seedlings. *Development, 137*(4), 597-606.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology, 3*(7), research0034.

Vanneste, S., and Friml, J. (2009). Auxin: a trigger for change in plant development. *Cell, 136*(6), 1005-1016.

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *TRENDS in Biotechnology, 27*(9), 522-530.

Vavilov, N. I. (1930). *The Problems of the Origins of Cultivated Plants and Domestic Animals as Conceived at the Present Time.* Paper presented at the U.S.S.R. Congr. Genet. Plant- and Animal-Breed.

Vazquez-Pianzola, P., and Suter, B. (2012). Conservation of the RNA Transport Machineries and Their Coupling to Translation Control across Eukaryotes. *Comparative and Functional Genomics, 2012:287852.*

Veberic, R., Trobec, M., Herbinger, K., Hofer, M., Grill, D., and Stampar, F. (2005). Phenolic compounds in some apple *Malus domestica* Borkh. cultivars of organic and integrated production. *Journal of the Science of Food and Agriculture, 85*(10), 1687-1694.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S. K., Troggio, M., and Pruss, D. (2010). The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics, 42*(10), 833-839.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science, 270*(5235), 484-487.

Vert, G., Walcher, C. L., Chory, J., and Nemhauser, J. L. (2008). Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proceedings of the National Academy of Sciences, USA, 105*(28), 9829-9834.

Vogt, J., Schiller, D., Ulrich, D., Schwab, W., and Dunemann, F. (2013). Identification of lipoxygenase (LOX) genes putatively involved in fruit flavour formation in apple (*Malus x domestica*). *Tree Genetics & Genomes*, 9(6), 1493-1511

Vorstermans, B., and Creemers, P. (2007). Screening preharvest/postharvest strategies to prevent fruit rot decay. *Communications in Agricultural and Applied Biological Sciences, 72*(4), 909-915.

Wagner, I., and Weeden, N. F. (1999). *Isozymes in Malus sylvestris, Malus domestica and in related Malus species.* Paper presented at the Eucarpia Symposium on Fruit Breeding and Genetics.

Wall, L., Christiansen, T., and Orwant, J. (2000). *Programming perl*: O'Reilly Media, Incorporated.

Walley, J. W., Kliebenstein, D. J., Bostock, R. M., and Dehesh, K. (2013). Fatty acids and early detection of pathogens. *Current Opinion in Plant Biology, 13*, S1369-5266.

Wang, F., Sanz, A., Brenner, M. L., and Smith, A. (1993). Sucrose synthase, starch accumulation, and tomato fruit sink strength. *Plant Physiology, 101*(1), 321–327.

Wang, L., Xu, Y., Li, J., Powell, R. A., Xu, Z., and Chong, K. (2007). Transgenic rice plants ectopically expressing AtBAK1 are semi-dwarfed and hypersensitive to 24-epibrassinolide. *Journal of Plant Physiology, 164*(5), 655-664.

Wang, S., Wang, X., He, Q., Liu, X., Xu, W., Li, L., Gao, J., and Wang, F. (2012). Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. *Plant Cell Reports, 31*(8), 1437-1447.

Wang, Z., Bai, M., Oh, E., and Zhu, J. (2012). Brassinosteroid Signaling Network and Regulation of Photomorphogenesis. *Annual Review of Genetics, 46*, 701-724.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57-63.

Wei, J., Ma, F., Shi, S., Qi, X., Zhu, X., and Yuan, J. (2010). Changes and postharvest regulation of activity and gene expression of enzymes related to cell wall degradation in ripening apple fruit. *Postharvest Biology and Technology, 56*(2), 147-154.

Wendel, J. F., Greilhuber, J., Dolezel, J., and Leitch, I. J. (2012). *Plant Genome Diversity* (Vol. 1): Springer.

Whitaker, B. (2012). *Genetic and biochemical bases of superficial scald storage disorder in apple and pear fruits.* Paper presented at the Southeast Asia Symposium on Quality Management in Postharvest Systems and Asia Pacific Symposium on Postharvest Quality 989.

White, P. J. (2002). Recent advances in fruit development and ripening: an overview. *Journal of Experimental Botany, 53*(377), 1995-2000.

Wickham, H. (2009). *Ggplot2: elegant graphics for data analysis*: Springer Publishing Company, Incorporated.

Wiersma, P. A., Zhang, H., Lu, C., Quail, A., and Toivonen, P. (2007). Survey of the expression of genes for ethylene synthesis and perception during maturation and

ripening of 'Sunrise' and 'Golden Delicious' apple fruit. *Postharvest Biology and Technology, 44*(3), 204-211.

Wisniewski, M., Bassett, C., Norelli, J., Macarisin, D., Artlip, T., Gasic, K., and Korban, S. (2008). Expressed sequence tag analysis of the response of apple (*Malus x domestica*, Royal Gala) to low temperature and water deficit. *Physiologia Plantarum, 133*(2), 298-317.

Wolf, S., Hématy, K., and Höfte, H. (2012). Growth control and cell wall signaling in plants. *Annual Review of Plant Biology, 63*, 381-407.

Wolfe, K., Wu, X., and Liu, R. H. (2003). Antioxidant activity of apple peels. *Journal of Agricultural and Food Chemistry, 51*(3), 609-614.

Wolfe, K. L., Kang, X., He, X., Dong, M., Zhang, Q., and Liu, R. H. (2008). Cellular antioxidant activity of common fruits. *Journal of Agricultural and Food Chemistry, 56*(18), 8418-8426.

Wrolstad, R., Putnam, T. P., and Varseveld, G. (2006). Color quality of frozen strawberries: effect of anthocyanin, pH, total acidity and ascorbic acid variability. *Journal of Food Science, 35*(4), 448-452.

Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics, 21*(9), 1859-1875.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C. Y., and Wei, L. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research, 39*(suppl 2), W316-W322.

Yamashino, T., and Mizuno, T. (2009). Cytokinin signal transduction with special reference to two-component regulatory system in plants. *Kagaku to Seibutsu, 47*(5), 312-322.

Yeung, K. Y., Haynor, D. R., and Ruzzo, W. L. (2001). Validating clustering for gene expression data. *Bioinformatics, 17*(4), 309-318.

Yruela, I., and Contreras-Moreira, B. (2012). Protein disorder in plants: a view from the chloroplast. *BMC Plant Biology, 12*(1), 165.

Yu, F., Thamm, A. M. K., Reed, D., Villa-Ruano, N., Quesada, A. L., Gloria, E. L., Covello, P., and De Luca, V. (2012). Functional characterization of amyrin synthase involved in ursolic acid biosynthesis in *Catharanthus roseus* leaf epidermis. *Phytochemistry, 91*, 122-127.

Zeeman, S. C., Kossmann, J., and A.M., S. (2010). Starch: Its metabolism, evolution, and biotechnological modification in plants. *Annual Review of Plant Biology, 61*, 209-234.

Zegzouti, H., Jones, B., Frasse, P., Marty, C., Maitre, B., LatchÈ, A., Pech, J. C., and Bouzayen, M. (1999). Ethylene regulated gene expression in tomato fruit: characterization of novel ethylene responsive and ripening related genes isolated by differential display. *The Plant Journal, 18*(6), 589-600.

Zeng, G. J., Li, C. M., Zhang, X. Z., Han, Z. H., Yang, F. Q., Gao, Y., Chen, D. M., Zhao, Y. B., Wang, Y., and Teng, Y. L. (2010). Original Article: Differential proteomic analysis during the vegetative phase change and the floral transition in *Malus domestica*. *Development, Growth and Differentiation, 52*(7), 635-644.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research, 18*(5), 821-829.

Zhang, C. B., Yuan, G. Z., Wang, J., Pan, G. T., Rong, T. Z., and Cao, M. J. (2011). Genetic analysis of maize cytoplasmic male sterile mutants obtained by space flight. *Yi chuan= Hereditas/Zhongguo yi chuan xue hui bian ji, 33*(2), 175-181.

Zhang, L. Y., Peng, Y. B., Pelleschi-Travier, S., Fan, Y., Lu, Y. F., Lu, Y. M., Gao, X. P., Shen, Y. Y., Delrot, S., and Zhang, D. P. (2004). Evidence for apoplasmic phloem unloading in developing apple fruit. *Plant Physiology, 135*(1), 574-586.

Zhang, Y., Zhu, J., and Dai, H. (2012). Characterization of transcriptional differences between columnar and standard apple trees using RNA-Seq. *Plant Molecular Biology Reporter, 30*(4), 957-965.

Zhao, Y. (2010). Auxin biosynthesis and its role in plant development. *Annual Review of Plant Biology, 61*, 49-64.

Zheng, H. Z., Kim, Y. I., and Chung, S. K. (2012). A profile of physicochemical and antioxidant changes during fruit growth for the utilisation of unripe apples. *Food Chemistry, 131*(1), 106-110.

Zheng, Y., Anderson, S., Zhang, Y., and Garavito, R. M. (2011). The structure of sucrose synthase-1 from *Arabidopsis thaliana* and its functional implications. *Journal of Biological Chemistry, 286*(41), 36108-36118.

Zhou, C., Lakso, A. N., Robinson, T. L., and Gan, S. (2008). Isolation and characterization of genes associated with shade-induced apple abscission. *Molecular Genetics and Genomics, 280*(1), 83-92.

Zhou, Y., Tang, J., Walker, M. G., Zhang, X., Wang, J., Hu, S., Xu, H., Deng, Y., Dong, J., Ye, L., Lin, L., Li, J., Wang, X., Pan, Y., Lin, W., Tian, W., Liu, J., Wei, L., Liu, S., Yang, H., and Yu, J. (2003). Gene identification and expression analysis of 86,136 Expressed Sequence Tags (EST) from the rice genome. *Genomics Proteomics Bioinformatics, 1*(1), 26-42.

Zhu, Z., Liu, R., Li, B., and Tian, S. (2013). Characterisation of genes encoding key enzymes involved in sugar metabolism of apple fruit in controlled atmosphere storage. *Food chemistry*, 141(4), 3323-3328.

Zolnerciks, J. K., Andress, E. J., Nicolaou, M., and Linton, K. J. (2011). Structure of ABC transporters. *Essays in Biochemistry, 50*(1), 43-61.

Zook, J. M., Samarov, D., McDaniel, J., Sen, S. K., and Salit, M. (2012). Synthetic Spike-in Standards Improve Run-Specific Systematic Error Analysis for DNA and RNA Sequencing. *PLoS One, 7*(7), e41356.

Zude, M., Herold, B., Roger, J. M., Bellon-Maurel, V., and Landahl, S. (2006). Non-destructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *Journal of Food Engineering, 77*(2), 254-260.

8.0 APPENDICES

Appendix 8. 1 Reconstructed transcripts statistics calculator using fasta_sequence_stats.pl

```perl
#!/usr/bin/perl -w
# count_fasta.pl
# AUTHOR : Zedias Chikwambi: (Modified from Joseph Fass) July 2012
use strict;
use POSIX;
use Getopt::Std;
my $usage = "\n\nusage: $0 [-i <interval size in # of residues>] <fasta file(s)>\n".
"Produces a histogram of sequence lengths and other measures to standard out.\n\n".
"-i #        specify bin size for histogram (default 100)\n\n";
our($opt_i); # histogram interval
getopts('i:') or die $usage;
if (!defined($opt_i) or !($opt_i =~ m/^[0-9]+$/)) {$opt_i = 100;}

if( ( $#ARGV + 1 ) < 1 ) {
        die $usage;
}
# Read in sequences from one or more fasta files
my @data_files;
for(my $i = 0; $i < ($#ARGV + 1); $i++){
        $data_files[$i] = $ARGV[$i];
}
my $Id;
my %seq;
foreach my $file (@data_files){
        open(FASTA, $file) or die"Can't open file $file\n";
        while (<FASTA>) {
#       if (/^>(.+)\s/)  { $Id = $1; }  # overwrites if id up to 1st whitespace is non-unique in file
if (/^>(.*)$/)  { $Id = $1; }
#                       if (/(\w+)/) {
                        elsif (/^(\S+)$/)    { $seq{$Id} .= $1 if $Id; }
        }
        close (FASTA);
}
# Count the number of sequences in the file and create a histogram of the distribution
my $n = 0;
my $int;
my $totalLength = 0;
my $gCount=0;
my $cCount=0;
my $aCount=0;
my $tCount=0;
my $nCount=0;
my $trans_to_Gmodel_identity=0;
```

```perl
my $gcatCount =0;
my $gcCount = 0;
my $gcPercentage=0;
my %len= ();
my @seqLengths;
foreach my $Id (keys %seq) {
        push @seqLengths, length($seq{$Id}); # record length for N50 calc's
        $n++;
        $int = floor( length($seq{$Id})/$opt_i );
        $totalLength = length($seq{$Id});
        $gCount = ($seq{$Id} =~ tr/gG/gG/);
        $cCount = ($seq{$Id} =~ tr/cC/cC/);
        $aCount = ($seq{$Id} =~ tr/aA/aA/);
        $tCount = ($seq{$Id} =~ tr/tT/tT/);
        $nCount = ($seq{$Id} =~ tr/nN/nN/);
        $trans_to_Gmodel_identity = ((($totalLength - $nCount)/$totalLength) *100);
        $gcPercentage = (($gCount + $cCount)/$totalLength )*100;
        if( !defined($len{$int}) ) {
                $len{$int} = 1;
        } else {
                $len{$int}++;
        }
# Calculate N25, N50, and N75 and counts
my $N25; my $N50; my $N75;
my $N25count=0; my $N50count=0; my $N75count=0;
my $frac_covered = $totalLength;
@seqLengths = reverse sort { $a <=> $b } @seqLengths;
$N25 = $seqLengths[0];
while ($frac_covered > $totalLength*3/4) {
        $N25 = shift(@seqLengths);
        $N25count++; $N50count++; $N75count++;
        $frac_covered -= $N25;
}
$N50 = $N25;
while ($frac_covered > $totalLength/2) {
        $N50 = shift(@seqLengths);
        $N50count++; $N75count++;
        $frac_covered -= $N50;
}
$N75 = $N50;
while ($frac_covered > $totalLength/4) {
        $N75 = shift(@seqLengths);
        $N75count++;
        $frac_covered -= $N75;
}
print
"$totalLength\t$gCount\t$cCount\t$aCount\t$tCount\t$nCount\t$trans_to_Gmodel_identity\t$gcPercentage"
;
}
```
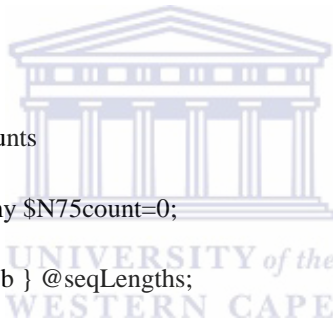
Appendix 8. 2 Blast_best.pl script. Adopted from Jennifer Menneghin (2007)

```perl
#!/usr/bin/perl –w
# Jennifer Meneghin
# 08/14/2007
# This script removes duplicate records from a "short" format BLAST output file, and keeps only the "best"
records  (sorts by smallest e-value and then biggest percent identity)
# Usage: blast_best.pl <input file><output file>
#-------------------------------------------------------------------------------------------------------------
#Deal with passed parameters
#-------------------------------------------------------------------------------------------------------------
#If no arguments are passed, show usage message and exit program.
if ($#ARGV == -1) {
usage("BLAST BEST 1.0 2007");
exit;
}
#get the names of the input file (first argument passed) and output file (second argument passed)
$in_file = $ARGV[0];
$out_file = $ARGV[1];
#Open the input file for reading, open the output file for writing.
#If either are unsuccessful, print an error message and exit program.
unless ( open(IN, "$in_file") ) {
usage("Got a bad input file: $in_file");
exit;
}
unless ( open(OUT, ">$out_file") ) {
usage("Got a bad output file: $out_file");
exit;
}
#Everything looks good. Print the parameters we've found.
print "Parameters:\ninput file = $in_file\noutput file = $out_file\n\n";
#-------------------------------------------------------------------------------------------------------------
#The main event
#-------------------------------------------------------------------------------------------------------------
$counter = 0;
$total_counter = 0;
print "De-duplicating File...\n";
@in = <IN>;
#Do stuff for each line of text in the input file.
foreach $line (@in) {
   #if the line starts with a pound symbol, it is not real data, so skip this line.
if ( $line =~ /^#/ ) {
        next;
   }
   #Count the total number of data lines in the file.
```

```perl
    $total_counter++;
    #The chomp commands removes any new line (and carriage return) characters from the end of the line.
chomp($line);
    #Split up the tab delimited line, naming only the variables we are interested in.
    ($id, $junk, $percent_identity, $junk, $junk, $junk, $junk, $junk, $junk, $junk, $evalue, $junk) = split(/\t/,
$line);
    #check to see if the id label is already in the list of ids (called dedupe)
    #if its not there, add it.
if ( $dedupe{$id} ) {
        #if it is, look at the old line to see if it is still "better" than the new one.
        ($junk, $junk, $list_percent_identity,, $junk, $junk, $junk, $junk, $junk, $junk, $junk, $list_evalue,
$junk) = split(/\t/,$dedupe{$id});
        #if the new evalue is better than the old one, change the value of this id to the new line.
        #otherwise, if the new evalue is the same, and the percent_identity is better, change the value of this
id to the new line.
        #otherwise, don't do anything (keep the old line).
        if ( $evalue < $list_evalue ) {
            $dedupe{$id} = $line;
        }
        elsif ( $evalue == $list_evalue ) {
        if ( $percent_identity > $list_percent_identity ) {
                $dedupe{$id} = $line;
            }
        }
    }
else {
        $dedupe{$id} = $line;
        #count the number of non-duplicated lines we have.
        $counter++;
    }
}
print "Total # records = $total_counter\nBest only # records = $counter\n";
print "Writing to output file...\n";
#Print the final "dedupe" list to the new file (adding the new line back on the end).
foreach $id (sort keys %dedupe) {
print OUT "$dedupe{$id}\n";
}
#Close the files.
close(IN);
close(OUT);
print "Done.\n";
sub usage {
my($message) = @_;
print "\n$message\n";
print "\nThis script removes duplicate records from a \"short\" format BLAST output file, and keeps only the
\"best\" records.\nIt sorts by smallest e-value and then biggest percent identity.\n";
}
```

Appendix 8. 3 Merge_blasts.pl script used to merge BLAST out puts to transcript ID list

```perl
#! /usr/bin/perl
#Copywright Chikwambi Zedias, ARC biotechnology platform, 2012
use strict;
my $daa_blast = shift;
my $q_fpkm = shift;
my %hash;
open(F, $daa_blast);
while (<F>) {
        chomp;
        my @info = split(/\t+/);
        my $trans_id = $info[0];
        my $gene_id = $info[1];
        my $C_2 = $info[2];
        my $C_3 = $info[3];
        my $C_4 = $info[4];
        my $C_5 = $info[5];
        my $C_6 = $info[6];
        my $C_7 = $info[7];
        my $C_8 = $info[8];
        my $C_9 = $info[9];
        my $C_10 = $info[10];
        my $C_11 = $info[11];
        my $desc = join("\t", $info[1], $info[2], $info[3], $info[4], $info[6], $info[7], $info[8], $info[9], $info[10], $info[11]);
        $hash{$trans_id}->{'gn'} = $trans_id;
        $hash{$trans_id}->{'dsc'} = $desc;
#       print "$trans_id\t".$hash{$trans_id}->{'dsc'}."\n";
}
close(F);
open(F, $q_fpkm);
while (<F>) {
        chomp;
        my @line = split(/\s+/);
        my $trans_id = shift @line;
        if ($hash{$trans_id}->{'gn'}) {
                $trans_id =~/^(\S+)/;
                my $trans_id = $1;
                print "$trans_id\t".$hash{$trans_id}->{'dsc'}."\n";

        } else {
                printf join('%s',"$trans_id\t","No-hit")."\n";
                # exit;
        }
}
close(F);
```

Appendix 8. 4 Retrieve replaced or merged uniprotKB accessions

#!/bin/sh

Created by Zedias Chikwambi on 2011/08/018.
Copyright 2011 __MyCompanyName__. All rights reserved.

$ for i in <replaced/merged accessions separated by space>;do echo
$i;curl "http://www.uniprot.org/uniprot/${i}.list" --compressed -L -s -o /dev/stdout echo;done

#The output is written to standard out.

Appendix 8. 5 Extracting fasta sequences corresponding to a list of sequence Ids using split_contigs.py

```
#!/usr/bin/python
# this script uses BioPython's SeqIO. For more on BioPython
from Bio import SeqIO
import sys
import os.path
if len(sys.argv) != 5:
  # if arguments are not the right number, print some helpful info
  # sys.argv[0] contains the name of the script, os.path.basename() is
  # used to trim it to just the script name, not its full path
sys.stderr.write('Usage: %s <FASTA file><Malus contig ID file><NULL FASTA output file><Malus
FASTA output file>\n' % os.path.basename(sys.argv[0]))
sys.exit(1)
# open input and output files. if the open fails, it throws
# and exception causing the script to exit. thus, unlike in perl,
# there is no need to check for failure (although doing so
# would make the errors a bit more user-friendly)
fasta_input = open(sys.argv[1])
#null_id_input = open(sys.argv[2])
Malus_id_input = open(sys.argv[2])
null_fasta_output = open(sys.argv[3],'w')
Malus_fasta_output = open(sys.argv[4],'w')

# use a hash data structure to store a list of the NULL ids.
# a hash is used because hash lookup is fast (the null_ids.has_key()
# used later in the script)
#null_ids = {}
#for line in null_id_input:
#  line = line.strip() # remove whitespace, including newlines
#  id = line.lstrip('>') # id = the line except for the leading >
#  null_ids[id] = 1
```

313

```python
# and now a hash data structure for the Malus ids.
# python filehandles (like Malus_id_input) are 'interable', that
# means you can use the for ... in ... : construct to iterate
# through them, just like you would with e.g. an array
Malus_ids = {}
for line in Malus_id_input:
line = line.strip() # remove whitespace
id = line.lstrip('>')
Malus_ids[id] = 1
#null_id_input.close()
Malus_id_input.close()
missing_ids = []
# for BioPython's sequence input and output look at Chapter 5 of the
# Biopython Tutorial and Cookbook
sequences = SeqIO.parse(fasta_input, 'fasta')
# algorithm:
# 1. read in a sequence
# 2. check if the id is in the null_ids hash
# 3.   if yes, write to the null_fasta_output file in FASTA format
# 4.  if no, check if the id is in the Malus ids hash
# 5.    if yes, write to the Malus_fasta_output file in FASTA format
# 6.    if no, append it to the list of 'missing ids'
# 7. if there are still sequences available, go to step 1.
for seq in sequences:
sys.stdout.write('.') # print some dots to show progress
ifMalus_ids.has_key(seq.id):
SeqIO.write(seq, Malus_fasta_output, 'fasta')
else:
SeqIO.write(seq, null_fasta_output, 'fasta')
print # print a newline
# if we have 'missing ids' that means there are sequences in the
# input FASTA file that are neither in the NULL ids list nor in the
# Malus ids list. Let the user know which these are
if len(missing_ids) > 0:
sys.stderr.write('Sequence IDs in input FASTA that are in neither set: ')
for id in missing_ids:
sys.stderr.write('%s ' % id)
sys.stderr.write('\n')
# close all the filehandles
fasta_input.close()
null_fasta_output.close()
Malus_fasta_output.close()
```

Appendix 8. 6 Subsetting mapped read in sam format using

mapped_ribo_free_reads-sam.pl

```perl
#! /usr/bin/perl
## this script is memory intensive, if you have big files, use high memory computers to reduce run time
```

```perl
use strict;
my $cuff_list = shift;
my $accession_track = shift;
my %hash;
open(F, $cuff_list);
while (<F>) {
        chomp;
        my @info = split(/\t+/);
        my $read_id = $info[0];
        $hash{$read_id}->{'rd'} = $read_id;
        #$hash{$read_id}->{'dsc'} = $desc;
}
close(F);
open(F, $accession_track);
while (<F>) {
        chomp;
        my @line = split(/\t+/);
        my $read_id = shift @line;
        my $C_2 = shift @line;
        my $C_3 = shift @line;
        my $C_4 = shift @line;
        my $C_5 = shift @line;
        my $C_6 = shift @line;
        my $C_7 = shift @line;
        my $C_8 = shift @line;
        my $C_9 = shift @line;
        my $C_10 =shift @line;
        my $C_11 = shift @line;
        my $C_12 = shift @line;
        my $C_13 = shift @line;
        my $C_14 = shift @line;
        my $C_15 = shift @line;
        my $C_16 = shift @line;
        my $string = join("\t", $read_id, $C_2, $C_3, $C_4, $C_5, $C_6, $C_7, $C_8, $C_9, $C_10,
$C_11, $C_12, $C_13, $C_14, $C_15, $C_16);
        if ($hash{$read_id}->{'rd'}) {
                $read_id =~/^(\S+)/;
                my $read_id = $1;
                print "$string","\n";
        } else {
                #printf join('%s',"No-hit")."\n";
 #              # exit;
        }
 }
close(F);
```
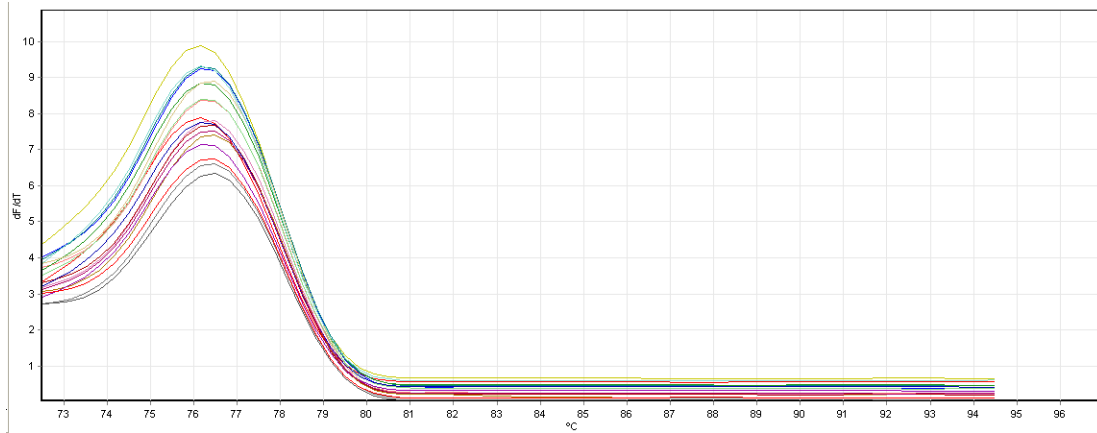
Appendix 8. 7A representative melt curve analysis from the reference gene ß-Actin using Rotor gene Q software v2.0.2. All qPCR amplifications were followed by a melt curveanalysis to check for the amplification of a single product. This example shows a uniform melt temperature of 76.5 ºC and confirms specific amplification. This method enables determination of non-specific amplification and can distinguish PCR products from primer dimmers. If the melt curve analysis indicated multiple PCR products, reaction mixes from qPCR reactions were run and visualised on agarose gels.

A)

B)
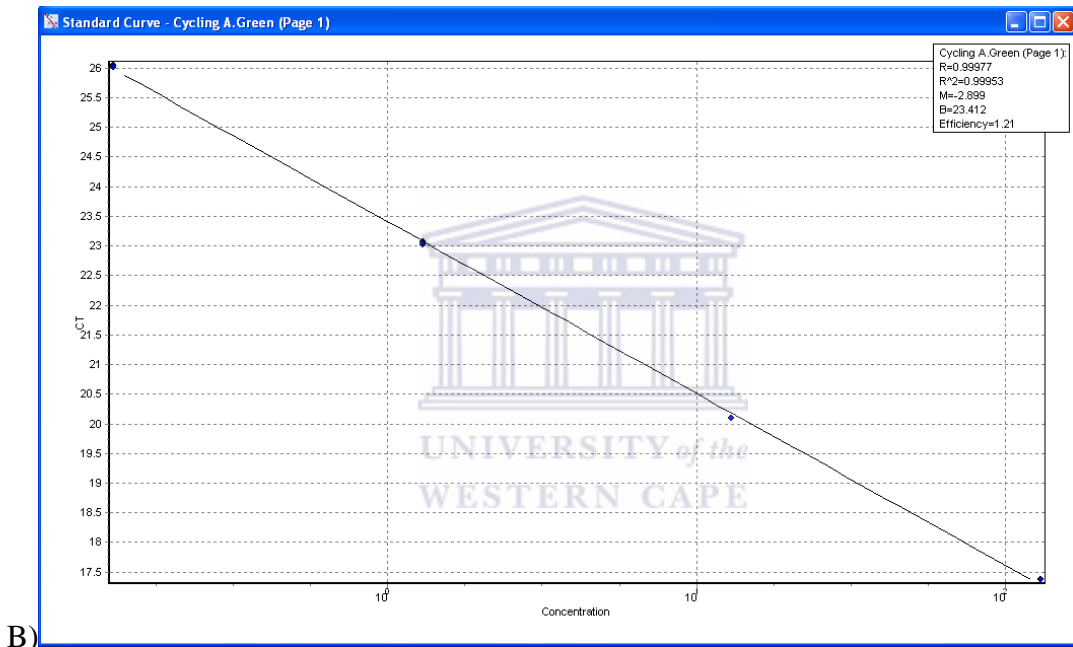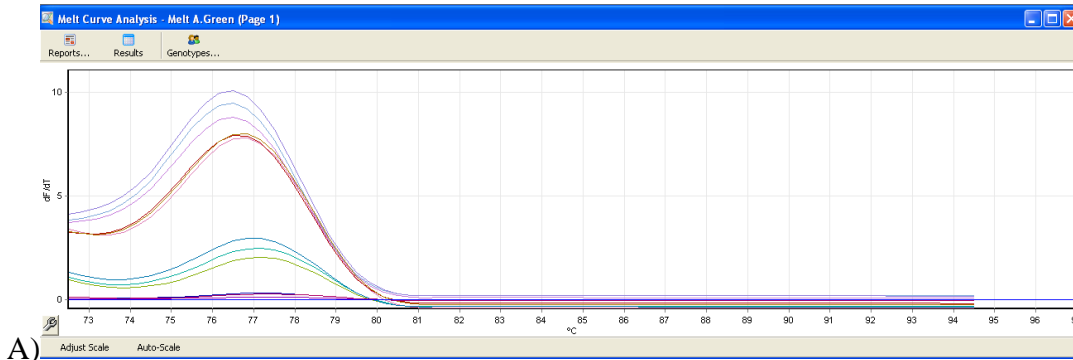
Appendix 8. 8 Serial dilutions of cDNA template used to generate a standard curve representing primer efficiency in Rotor gene Q software v2.0.2.(A) Representative example for *ß-Actin* shows a 10-fold dilution series over 5 orders of magnitude. Efficient amplification of the template was seen at all five dilutions over the 25 cycles PCR. (B) Raw qPCR data was used to construct a curve representing the efficiency of the primer pair to amplify the target DNA.
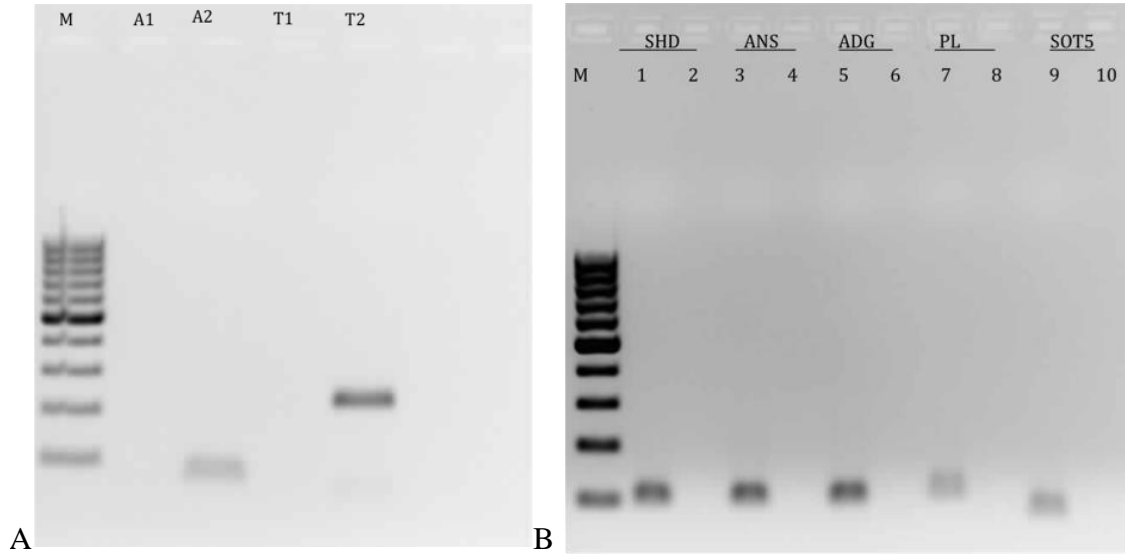
Appendix 8. 9 RT-qPCR products electrophoresis on 1.5% agarose gel stained with ethidium bromide. Gel picture A confirms that the reference genes ß-actin (A1 and A2) and tubulin produced only single bands corresponding to the expected fragment sizes. Gel picture B shows target gene RT-qPCR products. Lanes A1, T1, 2, 4, 6, 8, and 10 RT-qPCR products of no template controls

Appendix 8. 10 Post read mapping treatment effects on transcript abundance estimation: ANOVA table

Source of Variation	SS	df	MS	F	p-level	F crit
Treatment	6,1228E+13	5	1,22456E+13	3,37869	0,0154	2,53355
Daa	8,01E+13	5	1,60E+13	5,35041	0,00124	2,53355
Treatment. Daa	2,86E+13	25	1,14E+12			
Total	1,69959E+14	35				

Fruit tissue type and developmental effects on estimating transcript abundance was evaluated by two-way
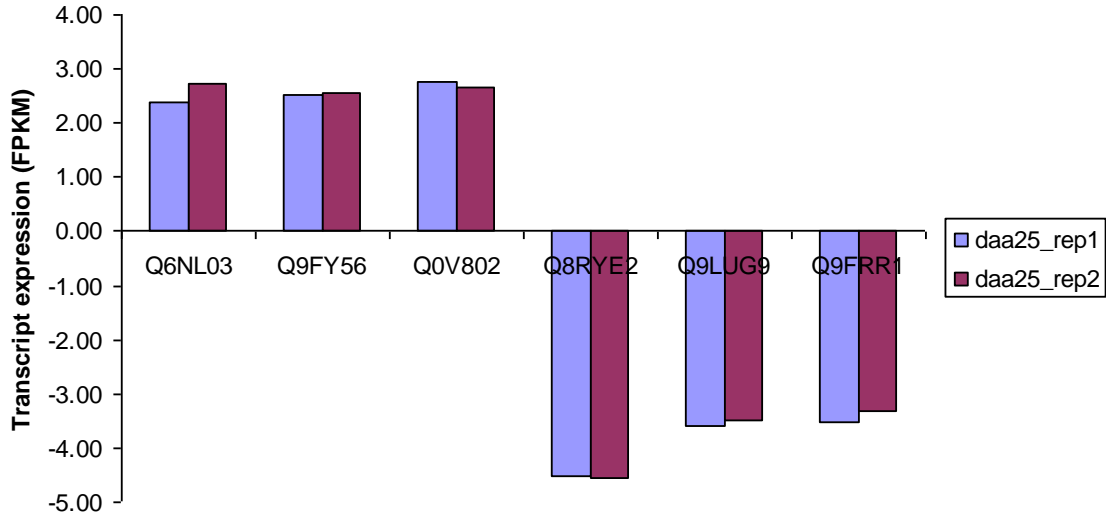
Appendix 8. 11 A comparison of biological replicates for uniformity in estimating transcript expression levels. The Figure shows almost uniform expression levels (FPKM) of six annotated transcripts estimated using RSEM at daa25 from two biological replicates (rep 1 and rep2). Each replicate consisting of ten apple fruits sampled around the tree canopy. The uniformity in the estimated expression values between the biological replicates suggests a consistancy in transcript expression estimation, as well as possible uniform physiological state of the biological replicates. As such the transcript reconstruction and expression estimates were acceptable.

A)

blast.ncbi.nlm.**nih.gov**/Blast.cgi — malus x domestica genome v1.0

Download ˅ GenBank Graphics

Eriobotrya japonica cultivar Jiefangzhong mixed amyrin synthase mRNA, complete cds
Sequence ID: gb|JX173279.2| Length: 2700 Number of Matches: 1

Range 1: 2493 to 2663 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
255 bits(138)	8e-65	160/171(94%)	0/171(0%)	Plus/Plus

```
Query  1     AGAGTTGATGAACTTGTAACTGCTGAAGAACGAATAATCTATGCAAATGTATTCTTTTGT  60
             |||||||| |||||||||||||||||||| ||||||||||||||||||||||||| |||||
Sbjct  2493  AGAGTTGACGAACTTGTAACTGCTGAACAACGAATAATCTATGCAAATGTGTTCTTTTGT  2552

Query  61    GATTTGTAACCGCTAAAGAACGAATAATCTGTGAAAATATATACATATTTATGTTACGTA  120
             |||||||||||||||||||||||||||||| |||||||||||||||||| ||||||| ||
Sbjct  2553  GATTTGTAACCGCTAAAGAACGAATAATCTATGAAAATGTATACATATTTATGTTACTCA  2612

Query  121   TAGACAACATGTATCATGCATTGCGAGATGTCACATCTCATCAAAATAAAT  171
             | |||||||||||||||||||||| |||||| |||||||||||||| ||||
Sbjct  2613  TGCACAACATGTATCATGCATTGCGAGATATCACATCTCATCAAATTAAAT  2663
```

Download ˅ GenBank Graphics

B)

blast.ncbi.nlm.**nih.gov**/Blast.cgi — malus x domestica genome v1.0

Download ˅ GenBank Graphics

Malus domestica beta-amyrin synthase (OSC1), mRNA
Sequence ID: ref|NM_001294017.1| Length: 2643 Number of Matches: 1
▶ See 1 more title(s)

Range 1: 2513 to 2632 GenBank Graphics ▼ Next Match ▲ Prev

Score	Expect	Identities	Gaps	Strand
222 bits(120)	8e-55	120/120(100%)	0/120(0%)	Plus/Plus

```
Query  1     AGAGTTGATGAACTTGTAACTGCTGAAGAACGAATAATCTATGCAAATGTATTCTTTTGT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  2513  AGAGTTGATGAACTTGTAACTGCTGAAGAACGAATAATCTATGCAAATGTATTCTTTTGT  2572

Query  61    GATTTGTAACCGCTAAAGAACGAATAATCTGTGAAAATATATACATATTTATGTTACGTA  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2573  GATTTGTAACCGCTAAAGAACGAATAATCTGTGAAAATATATACATATTTATGTTACGTA  2632
```

C)

Download ˅ GenBank Graphics

Malus domestica beta-amyrin synthase (OSC1), mRNA
Sequence ID: ref|NM_001294017.1| Length: 2643 Number of Matches: 1
▶ See 1 more title(s)

Range 1: 1 to 118 GenBank Graphics ▼ Next Match ▲ Pre

Score	Expect	Identities	Gaps	Strand
219 bits(118)	1e-53	118/118(100%)	0/118(0%)	Plus/Plus

```
Query  44    CTCAGCATCTGTGCCAAGCAACTCAAAGCAAACTCTTTCCTTCCTCACCATCACAGCCAG  103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     CTCAGCATCTGTGCCAAGCAACTCAAAGCAAACTCTTTCCTTCCTCACCATCACAGCCAG  60

Query  104   TTTTTGCAATCGGTACTAGAAAAAAGATTATTTTTGTACTACTAATCCAGTGATCAAG  161
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    TTTTTGCAATCGGTACTAGAAAAAAGATTATTTTTGTACTACTAATCCAGTGATCAAG  118
```

Appendix 8.12 Blast alignement comparison of the reconstructed BAS. The Figure shows alignments of the reconstructed BAS against two perfect hits in the NCBI (nr) nucleotide database. A and B are alignments against the 3' fragment while C shows an alignment against the 5 fragments. All the matches were against mRNA, coding sequence, as such the 3' and 5' extension of the reconstructed BAS transcript possibly code for a functional protein. Date of blast search: 28/08/2014