STRAWBERRY AND BEYOND:
A NOVEL AND COMPREHENSIVE INVESTIGATION OF FRUIT MATURATION AND RIPENING

Promotoren:	Prof. dr. J.N.M. Mol, hoogleraar moleculaire genetica van planten, Vrije Universiteit Amsterdam
	Prof. dr. Willem J. Stiekema, hoogleraar bio-informatica m.b.a.v. genomics
Co-promotor:	Dr. A. P. O'Connell, voormalig senior-onderzoeker, Plant Research International
Promotiecommissie:	Prof. dr. A. J. van Tunen Swammerdam Institute, Amsterdam
	Prof. dr. C. Martin John Innes Centre (UK)
	Prof. dr. R. J. Bino, Plant Research International, Wageningen
	Prof. dr. M. Koornneef Wageningen Universiteit

STRAWBERRY AND BEYOND: A NOVEL AND COMPREHENSIVE INVESTIGATION OF FRUIT MATURATION AND RIPENING

Asaph Aharoni

Proefschrift
ter verkrijging van de graad van doctor
op gezag de rector magnificus
van Wageningen Universiteit,
Prof. dr. ir. L. Speelman,
in het openbaar te verdedigen
op woensdag 16 october 2002
des namiddags te vier uur in de Aula.

A. Aharoni – Strawberry and beyond: a novel and comprehensive investigation of fruit maturation and ripening - 2002

Thesis Wageningen University, Wageningen, The Netherlands – 250 p.

Key words: **ripening, fruit, microarray, differential expression, functional genomics, metabolic pathway, transcription factor, MYB factor, hormonal regulation, achene, development, mass spectrometry, FT-MS, flavour, ester, alcohol acyltransferase, linalool, terpene**

ISBN: **90-5808-725-5**

Table of Contents

This thesis is dedicated to
Lena Albam and Sarah Aharoni

CHAPTER 1

General Introduction and Scope of the Thesis

GENERAL INTRODUCTION AND SCOPE OF THE THESIS

Functional Genomics and the Post-Genomic Era

General

Observing the history of key scientific discoveries achieved to date denotes the development of new equipment and methologies as the main driving force behind them. Such breakthrough technologies should certainly include the establishment and improvements both in throughput and quality of DNA sequencing. Advancements in DNA sequencing in the last 10 years has provided us with the opportunity to investigate a large portion and even the complete primary genetic information present in a given organism. The ease by which we can now obtain gene sequence data has enabled us to shift our scientific efforts to the understanding of how genes and proteins act to initiate, develop, maintain and terminate the life of biological entities.

Research on model organisms such as the yeast (*Saccharomyces cerevisiae)*, the worm (*Caenorhabditis elegans*) the fly (*Drosophila melanogaster*) and the weed *(Arabidopsis thaliana*) benefited the most from the development of DNA sequencing technologies, since as a result one can now observe the genome of such organisms as a whole. The complete genome sequencing of other organisms including humans, the monocot rice and more than 60 microbial species have also recently been completed (www.ncbi.nlm.nih.gov). The dramatic increase in DNA sequence information and the wish to a obtain global view on biological systems has led to the establishment of a new scientific approach, often referred to as “functional genomics” which combines knowledge from the various levels of genetic regulation (Figure 1). The goal of functional genomics is too systematically determine the biological function of all genes and proteins in a given organism and their interactions. To date an array of functional genomics tools have been developed and utilized enabling the large-scale investigation of gene expression, protein expression, protein structure, protein localization, enzyme activity, protein interaction and metabolite expression (Delneri et al., 2001). In addition, tools for the production of large mutant populations of plant lines and yeast strains have also been generated which allow either gene knockout or activation on a genomic scale (Parinov and Sundaresan, 2000; Vidal, 2001). In order

to obtain the maximum benefits from such information, development of data analysis programmes and integrated databases is an absolute request.

Examples of Tools for Functional Genomics

Large-Scale Mutation Analyses

The study of genetics in any biological discipline is dependent on the use of variants, either natural or induced by mutagenesis. Analysis of the inheritance of a variation in a mapping population will allow the identification of the genetic factors responsible for it. The use of chemical or physical procedures such as Ethyl Methyl Sulfonate (EMS) mutagenesis can relatively easily achieve a mutation in every gene in a genome. A more common way to clone a gene of interest is by establishing its position in a genetic map (termed "positional cloning"). However, the establishment of a mapping population and the fine mapping of a mutant locus are limiting factors in this procedure. When using a model organism, the existence of dense genetic maps with many molecular markers and a complete physical map, provided in addition to a collection of overlapping DNA fragments cloned may accelerate the procedure (Lukowitz et al., 2000).

A recently described method termed Targeting Induced Local Lesions In Genomes (termed TILLING; McCallum et al., 2000) uses populations of chemical or physical mutant lines for a reverse genetics approach (from gene to obtain a mutant and a function). In this method DNA from EMS mutagenised lines are pooled and a region of interest is amplified by PCR. Denaturation and annealing is then performed and allows the formation of hetroduplexes, which are then detected as an extra peak in a denaturing HPLC (DHPLC) chromatogram. The mutant individual can then be identified and the PCR product sequenced. With the current TILLING methodology one knockout lesion could be identified using a single DHPLC device in about 1 to 2 weeks.

Several reverse genetics approaches including the formation of deletions in ORFs using either gene replacement, "site selection insertion mutants" or transposons *in-vitr*o and by double-stranded (ds)RNA -mediated interference (RNAi) have been used extensively to generate whole genome knockouts in yeast and *C. Elegans* (Kim, 2001; Winzeler et al., 1999; Delneri et al., 2001). In plants, these methods could not effectively be used for large-scale loss of function experiments. Instead, insertional mutagenesis using either transposable elements or *Agrobacterium tumefaciens* mediated T-DNA (transfer DNA) insertion is now widely applied (Pereira, 2000). In this forward genetics approach the sequence inserted will block gene expression and as a result in some cases a mutant phenotype will be detected. Cloning the inserted DNA flanking region allows one to identify the "tagged" insert that might be responsible for the phenotype. Such screens may be used for reverse genetics approaches as well in which the pools of DNAs from the insertion lines can be screened with a pair of primers, one annealing to the insertion sequence and the other to a region within the gene of interest. PCR on the DNA pools combined with a de-convolution step will allow the identification of a single line with an insertion in the "target" gene. The chance

of identifying an insert depends on the size of the insert population, the genome size and the size of the gene targeted (Speulman et al., 1999).

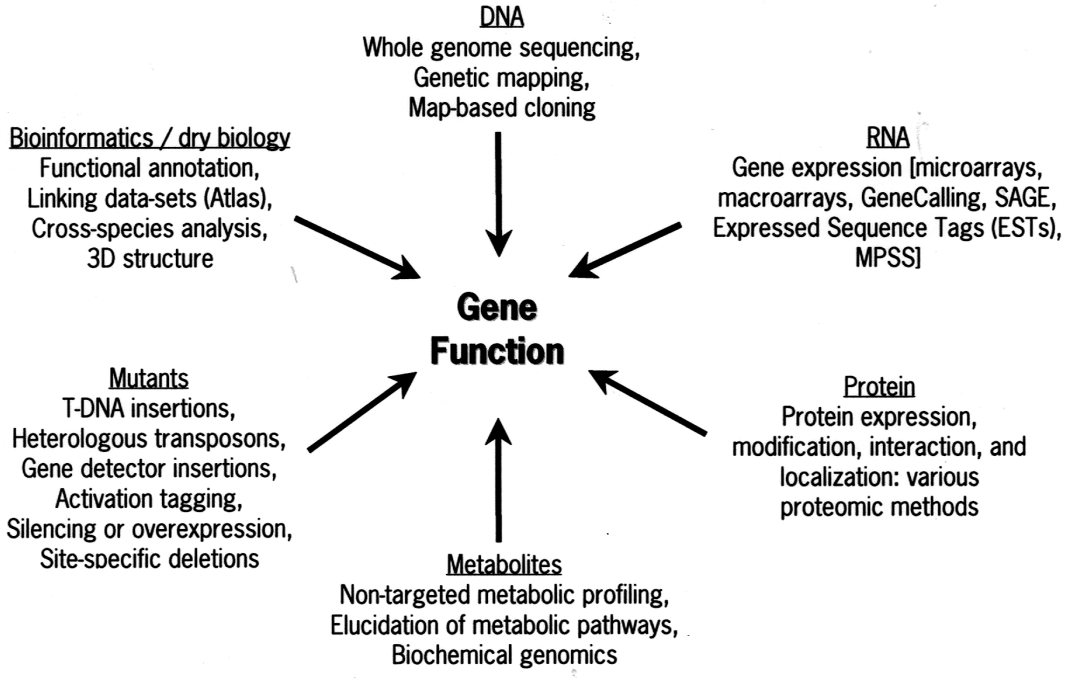


Figure 1 Current functional genomics approaches and methods to unravel gene function.

Insertion of foreign DNA by both transposon and T-DNA might be used not only for a loss of function approach but also for detection of specific and localised gene expression, which might provide evidence for gene function. Gene detector insertions lines may for example be produced by the insertion of promoterless marker genes, which will be expressed when inserted down stream of a promoter region (Bouchez and Hofte, 1998). It may also be used for gain of function approaches in which gene expression will be activated in an abnormal location causing mis-expression and a mutant phenotype. Inserting fragments that contain a strong enhancer element and thus "activating" the expression of a gene located nearby most often generates mis expression mutants (Weigel et al., 2000). The applicability of all of the above described insertional mutagenesis methods are largely dependent on the availability of an efficient gene transfer procedures such as the floral dip method applied to *Arabidopsis* (Clough and Bent, 1998).

Expressed Sequence Tags (ESTs) and Differential Gene Expression Analyses

The production of large number of Expressed Sequence Tags (ESTs), which are single pass cDNA sequences, was one of the first approaches used to obtain valuable and extensive insight into genomic information (Ohlrogge and Benning, 2000). Single cDNA clones to be sequenced can be randomly picked out of different cDNA libraries made from a variety of tissue types. In human and mouse alone the number of ESTs present in the public databases ranges between 3 to 4 million (www.ncbi.nlm.nih.gov). Although they cover less than 50% of genes identified by genomic sequencing (even in organisms with large collections), ESTs are deemed valuable as a tool for several applications amongst them gene discovery, genome annotation, identifying alternative splicing, and analyzing gene expression levels (so called "electronic Northerns"). ESTs collections provide relatively extensive and immediate access to the transcriptome of organisms where sequence information is non-existent or limited.

Several novel methods for transcriptome analysis either sequence-based (Brenner *et al*., 2000; Velculescu *et al*., 1995), fragment-based (Bachem *et al*., 1996; Shimkets *et al*., 1999) or hybridisation-based such as macro- and microarrays (Desprez *et al*., 1998; Lockhart *et al*., 1996; Schena *et al*., 1995) are currently available. Reports concerning the use of microarrays for large-scale monitoring of gene expression are constantly on the increase (Aharoni and Vorst, 2002). Two main types of microarrays namely "spotted arrays" and oligonucleotide arrays (GeneChips) are often employed for the hybridisation of flourecently labelled reference and test RNA populations, allowing parallel and quantitative comparison of transcript levels in samples under investigation. Oligonucleotide arrays are most suited to the multiplexed discovery and detection of DNA variation such as single nucleotide polymorphisms (SNPs). Whole genomes microarrays exist for some organisms and are expected to be applied to *Arabidopsis* and rice in the coming year (Zho and Wang, 2000). A detailed review of the principles and current applications of microarrays is provided in Chapter 2 of this thesis.

Proteomic Approaches

Gene products, namely the proteins are more difficult to decipher on a large-scale compared to nucleic acid. This is mainly due to their heterogeneity, their relatively low stability and their modification *in-vivo* , combined with the lack of tools for efficient protein sequence identification. The large-scale study of proteins (proteomics) usually by biochemical methods is complementary to genomics for various reasons. It is clear that gene sequence information is not sufficient to determine function. Annotation of genes in genomes using bioinformatics tools is far from optimal, and is estimated to contain an error of nearly 10% (Pandey and Mann, 2000). Therefore, verification of a gene product by proteomic methods is important for correct annotations of genomes. Apart from this, the existence of an Open Reading Frame (ORF) does not directly imply the existence of a functional gene. Proteomic approaches are able to detect and confirm post-translational modifications, protein expression levels (which do not always correlate with mRNA

levels), localisation, protein-protein interactions and mechanisms such as regulation of protein function such as for example by proteolysis and sequestration (Pandey and Mann, 2000; Dutt and Lee, 2000; Zhu and Snyder, 2001).

The use of mass spectrometry (MS) for the identification of proteins rather then there mere display was a significant breakthrough for proteomic analyses (Andersen and Mann, 2000). The use of MS mainly relies on the characterisation of peptides resulting from protein digestion by proteases. Two main approaches for MS protein identification exist, namely the acquisition of a "peptide-mass fingerprint" and the fragmentation of individual peptides in the mixture to gain sequence information (Gygi and Aebersold, 2000). In the first approach the mass spectrum of the peptides is obtained by an MS method termed matrix-assisted laser desorption/ionization (MALDI). The mass spectrum of the peptides is searched against the theoretical masses of tryptic peptides in a protein sequence database. This method is more successful when one is dealing with organisms where the complete genome has been sequenced. The second approach results in sequence of tryptic peptides of a protein rather than peptide masses. Peptides are ionised by electrospray ionization (ESI) directly from a liquid phase and the ions are sprayed into a tandem MS which, first isolates and then fragments them further for sequencing. One of the early tools in proteomics was the use of 2-dimensional (2D) gel electrophoresis. Isolation of proteins separated on 2D gels followed by MS for the identification of differential expression is a common application, although abundant proteins are often the ones separated and detected (Gygi and Aebersold, 2000). Novel methods are currently under development such as protein chips and quantified MS allowing for comparison of proteins levels from cell lysates without prior gel separation.

Protein-protein interactions can also be studied on a large scale as has been shown for the entire proteome of yeast using the two-hybrid system (Uetz et al., 2000). The two-hybrid system is based on modular structure of transcription factors wherein close proximity of the DNA binding domain to the activation domain induces transcription of a reporter gene. In this way simply sequencing the relevant clones can identify interacting proteins. MS might also be used to study protein-protein interaction by analysing protein complexes purified by affinity-based methods (Gygi et al., 1999).

Metabolite analysis

In many cases metabolites are the ultimate gene products and therefore integrating metabolic analysis into functional genomics approaches is an absolute request. Examples of large-scale metabolic analysis combined with linking gene to function are scarce in all scientific disciplines. This is mainly due to the large amount of different metabolites, often (in complex organisms) much higher compared to the amount of different genes and proteins. In addition, the enormous variability in the metabolic profiles between organisms and the absence of a high-throughput tool

for the unambiguous identification of metabolites have hampered developments in metabolic analysis.

Two main approaches to the assignment of gene function *via* metabolic investigation are currently being developed and utilized. In the first approach referred to as "biochemical genomics" a large-scale analysis of recombinant proteins is performed. The method allows the identification of biochemical reactions catalyzed by enzymes with an unknown function. Martzen et al., (1999) used "biochemical genomics" to identify yeast proteins with specific enzyme activities by constructing an array of individual yeast strains (representing a genomic set of 6144 ORFs) expressing glutathione S-transferase (GST) - ORF fusion proteins. In order to link a gene to a specific biochemical activity, the strains were grown, expressed and purified in pools. Enzymatic assays were performed in the same format and the subsequent analyses of individual strains included in an "active" pool led to the identification of the corresponding gene. Such a method was successful for matching genes to enzyme activities although due to its "*in-vitro*" context the biological functions of such genes could not be unequivocally determined.

The second approach metabolically profiles different extracts in the same fashion as profiling gene and protein expression. Comparisons of metabolic profiles between samples during development of organisms and under different environmental conditions (including the analyses of loss and gain of function mutants) may provide clues to gene function. Such an approach has been applied to yeast and recently has also been reported for profiling *Arabidopsis* and potato extracts. Raamsdonk et al. (2001) used nuclear magnetic resonance (NMR) spectroscopy to analyze metabolic changes arising from gene deletions. The authors demonstrate that by clustering together metabolic profiles of yeast strains with deletions in genes related to similar metabolic activity and genes with unknown function, one could identify a metabolic phenotype and link the unknown gene to a functional domain. The method developed termed FANCY (functional analysis by co-responses in yeast) enabled one to assign a metabolic phenotype to "silent genes" whose deletion did not result in an effect on yeast growth rate, but effected overall metabolite concentration.

Plants contain a more complex and larger mix of metabolites mainly due to their extensive secondary metabolism. As a result, many phenotypes will remain undiscovered if the analysis of metabolite concentrations does not form part of the screening procedure. Presently the most common phenotypic screens are based on conditional lethality, fertility and on easily detected morphological phenotypes such as dwarfism. To date only a fraction of phenotypes have been identified by biochemical screens, mainly through targeted approaches in which analytical techniques dedicated to the analysis of a specific compound or a similar class of compounds were utilized. For metabolic profiling a non-targeted approach in which a large array of metabolites is screened is preferred and in the last two years a number of studies employing such methods have been reported. Gas chromatograph (GC) coupled to a mass spectrometer (GCMS) was used to quantify 150 compounds and to obtain biochemical profiles of potato tubers grown under different conditions including genetically manipulated ones (Roessner et al., 2000; Roessner et al., 2001a).

The same method was employed to profile 326 distinct metabolites in *Arabidopsis* and to make a comparison between the profiles of two ecotypes (Fiehn et al., 2000). GCMS is used to separate compounds relative to their vapor pressures and affinity for the material in the chromatographic column and therefore are limited to the analysis of volatile and heat stable compounds. Many biological compounds are not sufficiently volatile as such to be separated by GCMS. They must undergo prior chemical derivatization and this makes the analysis far more complicated. GCMS has been used to evaluate the biochemical composition of different apricot cultivars under different conditions by determining sugar, sugar alcohol, organic acid and amino acid composition (Katona et al., 1999).

Mass spectrometers are very sensitive and selective detectors and when coupled with the appropriate sample introduction and ionization techniques, they can selectively analyze most chemicals, both organic and inorganic. The measurements of a compounds mass or the distinctive masses of fragments that form during ionization can be used to derive the chemical identity (Glassbrook and Ryals, 2001). It is thus likely that mass spectrometry will play a key role in future metabolic profiling equal to its importance in proteomic analysis. Different types of mass analyzers and ion sources can be used such as quadropol, ion trap and time of flight. In chapter 3 of the thesis the use of another type of mass analyzer (Fourier Transform ion cyclotron Mass Spectrometry) for metabolic profiling is described.

Bioinformatics

The first step of data analysis is proper storage of information, allowing the integration, comparisons and presentation of experiments. Similar to the case of sequencing data, including results from profiling experiments (e.g. millions of data points from microarray hybridizations) in scientific publications is not common. In scientific publications complete data sets from profiling experiments are usually provided as a supplement to the printed paper on the web, whilst the manuscripts themselves present an overview or focus on certain aspects of the results. In the early days of sequencing, central data repositories (e.g. GenBank and SwissProt) were established to house data. Standardized databases to house results from different profiling (e.g. microarray) experiments are still in their infancy. An example of a microarray results database is the Stanford Microarray Database (SMD; Sherlock et al., 2001) which stores raw and normalised data from microarray experiments, as well as their corresponding image files. In addition, SMD provides interfaces for data retrieval, analysis and visualisation. It currently stores results of more than 9000 array experiments, from 9 different organisms. The amount of data is enormous, for example, a single microarray of 20,000 spots hybridised in an experimental series of three may generate more than 50 million data points.

Besides storing profiling data, the information needs to be analysed by bioinformatics in order to make biological sense out of it and to assign functions to yet uncharacterised coding regions. In the case of microarray data, one approach for assigning gene function is by clustering, which

groups genes with unknown and known functions with similar expression profiles, and in this manner a gene function may be elucidated through a "guilt by association" process. Statistical methods for clustering such as hierarchical clustering (Eisen et al., 1998), and self-organising maps (Tamayo et al., 1999) are often used for data interpretation and visualisation. Similar clustering methods can be used for other profiling methods such as for example grouping samples according to their metabolites concentrations (Roessner et al., 2001b).

One step further is to integrate the data from the different levels of functional genomics to an atlas of functional maps (Vidal., 2001). The atlas will integrate data-sets originating from sequencing, gene/protein/metabolite expression profiling, mutagenesis, protein localisation, structural genomics, biochemical genomics and this will allow the formulation of hypothesis which can then be tested experimentally. This approach was demonstrated by Marcotte et al, (1999) and provided a perfect example of the benefit of integrating data from the various levels of functional genomics. They grouped and determined functional relationships for all yeast proteins by correlating experimental data, related metabolic function, related phylogenetic profiles, "Rosetta Stone Method" (i.e. link proteins whose homologues are fused into a single gene in another organisms) and mRNA expression. Using this approach 93,000 pair-wise links were discovered. Links between characterized and uncharacterized proteins allowed a general function to be assigned to more than half of the 2,557 previously uncharacterized yeast proteins. Databases integrating genetic and phenotypic data are currently at the stage of conceptual modeling (Paton et al., 2000) and will be major information resources for biological sciences in the future.

Fruit Maturation and Ripening

General

Plants develop fruit in order to ensure seed dispersal and the establishment of a new generation. Real fruit are normally defined as those arising from the carpel. False are those developing from extracarpellary tissues, such as for example strawberry, which is actually a swollen receptacle (Giovannoni, 2001). After anthesis fruit passes through three main phases of development which include cell division, followed by cell expansion and ultimately ripening (Gillaspy et al., 1993). The process of ripening in fleshy fruit has attracted the most scientific attention for studying the genetic bases of fruit development, since ripe fruit serve a large portion of the human diet and the process by itself is a unique aspect of plant development (Giovannoni, 2001). Ripening is a highly complex process, which is characterised by a series of co-ordinated biochemical and physiological changes. These include major alterations to texture, pigmentation and soluble solids levels which are often accompanied by the biosynthesis of flavour and aroma compounds (Seymour, 1993). Apart from their classification according to the tissue of origin, fruit are often categorised as dry or

fleshy and/or climacteric or non-climacteric. The sharp increase in climacteric ethylene production is considered as controlling the initiation of the major biochemical and physiological ripening changes. Non-climacteric fruit such as strawberry and grape do not show the rise in ethylene and respiration, which is observed for climacteric fruit such as tomato and banana. Although it is clear that while such division is convenient in certain cases the ethylene based classification between climacteric and non climacteric fruit does not hold in all cases (Kuntz et al., 1998; Lelievre et al., 1997).

Most of the scientific research performed to date on fleshy fruit have used tomato as a model system. Apart from its economical importance as a commercial fleshy fruit crop (second after citrus), various other scientific reasons makes it the plant of choice (Giovannoni, 2001). Tomato is a diploid and has a relatively small genome (950 Mb). It is easy to propagate, cross and manipulate genetically, has relatively short lifetime, bears fruit year around if grown in the greenhouse and contains a wide germplasm resource. In recent years several genetic tools have been developed in tomato, which include the construction of a genetic map containing over a 1000 molecular markers, large insert genomic libraries (most useful for positional cloning), and the introduction of the Microtom tomato as a tool for functional genomics (Meissner et al., 2000). This in addition to the production of over 100,000 Expressed Sequence Tags (ESTs) from 20 different cDNA libraries, 30,000 of them derived from different fruit stages. *Arabidopsis thaliana* is the obvious choice as the model for dry fruit research and several laboratories aim at benefiting from the advantages of using *Arabidopsis* to unravel gene function in tomato (Mysore et al., 2001).

Several important fruit crops are defined as non-climacteric such as the grape, citrus and strawberry. In fruit such as grape and strawberry the phytohormone auxin has been shown to be of major importance in the regulation of ripening (Davies et al., 1997; Given et al., 1988). Although ethylene in the past has been generally considered not to regulate the ripening of non-climacteric fruit, evidence is accumulating that it might influence specific ripening related processes, such as de-greening in citrus (simultaneous degradation of chlorophyll and accumulation of carotenoids; Goldschmidt et al., 1993), and anthocyanin biosynthesis in grapes (Roubelakis-Angelakis and Kliewer 1986). Ethylene regulated genes were also isolated from the non-climacteric pineapple and orange fruit during maturation (Cazzonelli et al., 1998; Alonso et al., 1995a).

Main Fruit Ripening Processes

Although a wide range of fruit types exist which differ in their metabolism, the biochemical pathways involved are common to all fruit and even to other plant tissues (Seymour, 1993). Respiration for example, which is responsible for the supply of energy and carbon skeleton building blocks, shows different patterns and rates between different fruit and for other tissues. Certain types of sugars and organic acids, which accumulate in the cell vacuole play a dual role in ripening, by both providing substrates for respiration and for the biosynthesis of flavour and

fragrance compounds. Malate and citrate are the most common organic acids present in fruit (usually declining in levels during maturation), whilst the sugars glucose, fructose and sucrose (increase in levels during maturation) are also often present but in different ratios. Organic acids and sugars present in the fruit are normally derived from the remainder of the plant, although a large number of fruit may accumulate low levels of these compounds in early stages of development when the tissue is still performing photosynthesis.

The process of fruit softening and tissue deterioration is a major ripening event, occurring as a result of extensive cell wall modifications. Carbohydrates polymers make up to 90-95% of the structural components of the cell wall and they include cellulose, hemicelluloses and pectins. The remaining cell wall components are structural proteins such as hydroxyproline-rich glycoproteins. Chemical changes in fruit cell walls during ripening include solubilisation and degradation of pectin, loss of neutral sugar from pectin side chains, and a reduction in the molecular weight of xyloglucan (Redgwell et al., 1997). Throughout the years, studies have focused largely on cell wall break down as the primary reason for softening. Numerous studies on polysaccharide hydrolases such as polygalacturonases, pectin methylesterases, beta-galactosidases and pectate lyases have been conducted and their activities and expression of their corresponding genes correlated with fruit softening (Brownleader et al., 1999). It is however, clear that processes other than cell wall hydrolysis contribute to textural changes and the loss of fruit firmness during ripening (Brummell et al., 1999a). One such process is the loosening of cell wall components mediated by enzymes such as expansin, another major process is the depolymerization of high molecular mass hemicelluloses.

An important aspect of fruit ripening is the accumulation of a diverse array of secondary metabolites. The precursors for their metabolism will be produced early in fruit development and later in the ripe stage these compounds will provide the “make up” for the fruit to become and remain attractive. Colouring of berries for example will most often be associated with the accumulation of the flavonoids anthocyanins (Kahkonen et al., 2001). In other fruit such as tomato, pigmentation at the ripe stage is due to various carotenoids often synthesised parallel to chlorophyll degradation (Jacob-Wilk et al., 1999). Another set of flavonoids such as the flavonols quercitin and kaempferol will protect the fruit against UV damage (Winkel-Shirley, 2001). Astringency in early fruit development may be associated to the accumulation of the phenolic compounds like tannins and this will prevent the fruit being eaten before the seeds have matured (Cheng and Breen, 1991). The aroma of fruit is produced by a large mix of in some cases a few hundred molecules. These volatile compounds are derived from several main biosynthetic pathways such as the lypoxygenase pathway (fatty acid metabolism), amino acid degradation pathways, the phenylpropanoid pathway and the isoprenoid pathway (Hornstein, 1999). A more detailed description of fruit flavour and aroma components is provided below. The different flavour and aroma components play an additional role in ripening and that is the protection of the fruit against pathogens.

Ripening processes are well co-ordinated between themselves and degradation product from one metabolic pathway might serve as a substrate for another pathway involved in another ripening process. This co-ordination was nicely exemplified by Frenkel et al, (1998) showing that methanol released as a results of pectin methylesterase action on cell walls is correlated with methanol levels in the fruit which might directly effect fruit aroma and flavour. In order to execute successfully a complicated program such as ripening, fruit have to produce a new batch of "ripening associated proteins". The synthesis of these proteins starting at the levels of gene expression is under strict regulation, which includes the action of hormones and signal transduction cascades.

Molecular Studies in Fruit

Molecular work on fruit ripening has been performed mainly on tomato although in recent years there has been a dramatic increase in the investigations of other fruit species including melon (Hadfield et al., 2000), grape (Davies and Robinson, 2000), strawberry (Wilkinson et al., 1995a), citrus (Alonso and Garnel, 1995b), raspberry (Jones et al., 2000), pear (Itai et al., 2000) and banana (Clendennen and May, 1997). The majority of the studies identified genes with elevated expression during ripening and correlated their putative identity with a specific ripening process. Genes associated with an array of ripening related processes have been cloned amongst them cell wall metabolism, pigmentation (anthocyanin and carotenoid), stress and pathogen-related, sugar metabolism, chlorophyl degradation, hormonal regulation and signal transduction (mainly related to ethylene), fatty acid metabolism, water accumulation and regulatory genes (e.g. MYB and MADS box transcription factors).

Most important molecular insights into fruit ripening have been obtained by use of overexpression and antisense technology in transgenic fruit. Recently, for example Lu et al. (2001) cloned and characterised a ripening induced tomato Rab11 GTPase gene (*LeRab11a*). Antisense *LeRab11a* transformation of tomato plants resulted in fruit, which failed to soften, and had reduced expression of two cell wall hydrolyses. A range of additional abnormal developmental changes were observed, in addition to a reduction in ethylene levels in some of the lines. Another set of transgenic tomato plants with suppressed expression or overexpression of expansin (*Exp1*) provided strong evidence for the role of expansins in softening of ripe fruit (Brummell et al., 1999b). Fruit in which the Exp1 protein was suppressed were much firmer than the controls during ripening. They also showed dramatically inhibited polyuronoide depolymerization late in ripening but the breakdown of hemicelluloses was not prevented. The role of sucrose synthase (SuSy) was studied in transgenic tomato using antisense technology. The results suggested that SuSy participates in the control of sucrose import capacity of young tomato fruit, which is a requirement for fruit set and development (D'Aoust et al., 1999). Antisense suppression of ACC oxidase, the gene encoding the enzyme catalysing the ultimate step in ethylene biosynthesis, in another

climacteric fruit plant, melon revealed ethylene dependent and independent ripening processes (Ayub et al., 1996).

The role of ethylene in fruit ripening and especially the signal transduction cascades associated with it has been a subject of detailed investigation in recent years (e.g. Tiemen et al., 2001; Tiemen et al., 2000; Hacket et al., 2000). Research has been supported by the cloning and characterisation of *Arabidopsis* ethylene signal transduction cascade genes including those encoding the ethylene receptor (ETR) protein family (Giovannoni, 2001). Studies on ethylene signal transduction in tomato fruit have been further assisted by the use of the *Never ripe* (*Nr*) mutant which is insensitive to ethylene (Lanahan et al., 1994; Wilkinson et al., 1995b). Genes associated with important fruit characteristics mainly those traits influenced by quantitative trait loci (QTLs) were also discovered by map based cloning approaches. Frary et al, (2000) cloned a human oncogene RAS protein homolog (*fw2.2*) which influences fruit mass. They suggested that the fw2.2 protein regulates fruit mass through modulation of pre-anthesis carpel cell number and that it acts as a negative regulator of cell division in wild tomato. Also suggested was that during domestication higher fruit mass is gained due to the introduction of weaker alleles at the *fw2.2* loci. Genes associated with other QTLs have been reported. These include the *B* gene which forms part of an alternative pathway to β-carotene formation in tomato chromoplasts, influencing fruit color and another QTL (*Brix9-2-5*) influencing the total soluble content of tomato fruit (Ronen et al., 2000; Fridman et al., 2000). It is anticipated that gene discovery in ripening fruit will be translated into many fruitful applications using genetic engineering approaches.

Strawberry Fruit

Structure, Growth and Development

Strawberry belongs to the rose family *(Rosaceae,* subfamily Rosoideae, tribe Potentilleae) in the genus *Fragaria*. There are four basic fertility groups in *Fragaria* divided according to their ploidy level or chromosome number. *F. vesca* is the most common native species, which contains 14 chromosomes and is a diploid (Hancock, 1999). The cultivated varieties of commercial strawberries usually recognised as *F. x ananassa*, are almost all octoploids, containing 56 chromosomes and are derived chiefly from the octoploids *F. chiloensis* (native to South America) and *F. virginiana*, (native to the eastern United States). Most other evolutionary relationships within the genus are not clear. *F. vesca* may be the ancestor of *Fragaria* species since it will grow in most areas of other species and its chromosomes will pair with many of them including the octaploids. The first strawberry species were domesticated 2000 years ago, and the first commercial strawberry was introduced only 250 years ago.

The strawberry plant is a herbaceous perennial, its fruit initiate from an inflorescence and are actually an aggregate, composed of many ovaries, each with a single ovule (Perkins-Veazie, 1995a). The seeds (achenes) embedded in the epidermis of the swollen receptacle tissue are the

true fruit. The receptacle is composed of an internal pith, a cortex layer and an epidermal layer (Suutarinen et al., 1998). Fibrovascular strands connect the achenes to the interior of the receptacle and they supply nutrients to the achenes and the surrounding parenchyma cells. The achenes are about 1 mm in length and each receptacle can contain a few hundred of them. The mature achene contains a hard and relatively thick pericarp, a thin testa, a one layer endosperm and a small embryo. Embryo formation is completed 10 days after anthesis and its stores protein and fat but no starch. Receptacle growth fits a single or double sigmoid curve depending on the cultivar. The cortex cells are primarily responsible for receptacle growth. Growth is mainly as a result of cell enlargement, with cell size increasing towards the pith (Havis, 1943). Cell division accounts only for only 15-20% of total growth, occurring early in fruit development (ceasing 15 days following anthesis)(Cheng and Breen, 1992). Fruit development may be divided to 5 different stages including small green, large green, white, turning and red (ripe) and ripe fruit will appear approximately 30 to 40 days post anthesis. The ripening process is relatively rapid and occurs 5 to 10 days following the white stage.

Hormonal Influence

It is generally understood that fruit do not develop properly if seeds have not been formed, and this points to a clear interaction between these two entities. Already by 1950 Nitsch recognised the fact that strawberry could serve as perfect model for the study of this interaction and the genetic messages associated with ripening. The phytohormone auxin is the main signal molecule co-ordinating the growth and initiation of ripening in strawberry fruit (Nitsch, 1950; Given et al., 1988). Achenes are the source of auxin where it is synthesised early in fruit development. During the early stages of fruit development auxin promotes fruit growth. Removing part of the achenes in early developmental stages inhibits auxin transport to the receptacle causing distorted berry formation, the receptacle expanding only in proximity to the undisturbed achenes. If auxin is then applied to the receptacle in areas were achenes were previously located, fruit expansion continues up to 21 days post-anthesis. On the other hand strawberry ripening is triggered by the decline in the levels of auxin in the receptacle, probably due to the cessation of the auxin synthesis and transport from the maturing achenes (Figure 2). This is exemplified by the fact that removing the achenes at the large green stage will provoke ripening, as measured by anthocyanin accumulation, whilst removal of achenes and application of synthetic auxin at this stage will delay ripening (Given et al., 1988). Free auxin levels peak in the receptacle and achenes prior to the white stage and then they decline as the fruit matures. Auxin is also found in its conjugated form either ester or amide linked and this might be part of a mechanism to determine its levels in the fruit and hence to control fruit development and ripening (Archbold and Dennis, 1984). This co-ordinated action between achenes and receptacle is part of the mechanism ensuring achene maturation prior to fruit ripening. It has been clearly demonstrated that the declining auxin levels trigger the ripening by

inducing de-novo synthesis of specific ripening associated mRNAs (Manning, 1994; Manning, 1998).

Due to the prominent role ethylene plays in the ripening of climacteric fruit such as tomato, banana and melon, researchers have turned their attention to elucidate the function that ethylene plays in the development and ripening of non-climacteric fruit, including strawberry (e.g. Perkins-Veazie et al., 1995b; Perkins-Veazie et al., 1996; Luo and Liu, 1994; Tian et al., 2000). In strawberry endogenous ethylene production is extremely low, decreasing between green and white stages of development and then increasing in the red ripe stage (Knee et al., 1977; Abeles and Takeda, 1990). Inhibiting ethylene production or ethylene action in the large green stage did not influence the accumulation of strawberry fruit pigmentation, a clear indicator of fruit ripening (Given et al., 1988). Activity of the key enzyme in ethylene biosynthesis, ACC (1-aminocyclopropane-1-carboxylate) synthase could not be demonstrated in any stage or tissue of strawberry fruit. Relatively low ethylene production of strawberry fruit in the presence of excess ACC indicates that although the enzyme using ACC to perform the last step in the biosynthesis of ethylene, (ACC oxidase) is indeed active especially at early stages of development it might be another step of regulation in the pathway which prevents the fruit from producing high ethylene levels (Perkins-Veazie et al., 1995b).

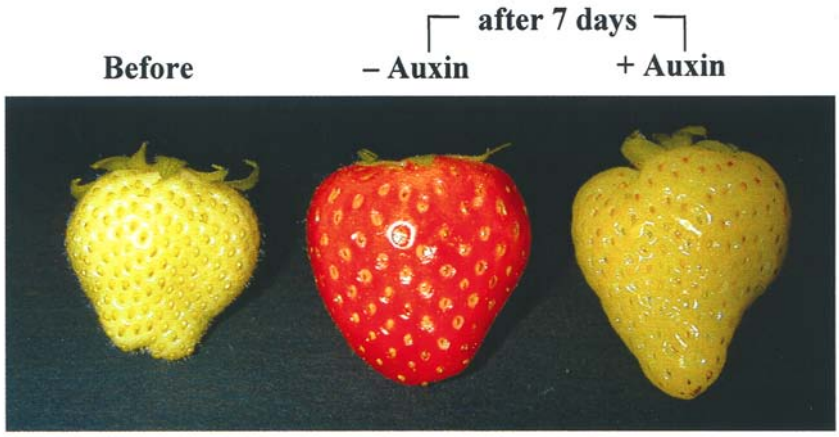


Figure 2 Effect of auxin treatment on strawberry fruit. White stage fruit treated with NAA at a concentration of 0.5 mM in a lanolin paste (+ Auxin) and control fruit treated with lanolin paste only (– Auxin).

Ripening-Biochemical and Molecular Studies

In recent years strawberry has emerged as a suitable model plant for studying non-climacteric fruit ripening (Giovannoni, 2001). Although strawberry fruit is derived from the flower receptacle it shares both in development and in ripening the characteristics of real fruit. This include for example the degradation of chlorophyll, the accumulation of anthocyanin, softening partially mediated by cell wall hydrolysing enzymes, the metabolism of sugars and organic acids and the production of flavour compounds. Strawberry despite its less economic value as a crop when compared to other non-climacteric fruit such as citrus and grape, has a relatively short reproductive

cycle (approximately half a year for the wild and a year for the cultivated), which makes genetic studies easier. In addition both *F. vesca* and *F. annanasa* are appropriate for efficient genetic transformation by the simple *Agrobacterium tumefaciens* mediated leaf disk transformation procedure (El –Mansouri et al., 1996; James et al., 1990). In many cases the choice to use *F. vesca* for genetic studies (e.g. mapping) might be a clear advantage on using the cultivated varieties due to the simplicity of dealing with a diploid plant and the shorter reproductive cycle (Nam et al., 1999). Problems associated with isolation of nucleic acid from plants such as strawberry (Manning, 1991), "contaminated" with problematic components such as carbohydrates and phenolics have been largely solved. This has enabled the application of more molecular biology methods such as for example construction of high quality cDNA and genomic libraries and microarray hybridisations for the study of strawberry .

Early investigation of the various ripening processes in strawberry mainly used biochemical techniques including enzymatic activity assays, to follow changes in specific metabolic profiles during the development of the fruit from green to red ripe. More recently, researchers have been investigating on a larger scale the molecular aspect of strawberry ripening, mainly at the level of gene expression. As the strawberry fruit ripens, cells expand and many changes occur which include tubular proliferation of the tonoplasts, cells contain a large vacuole, plastid degenerate and most of the starch present in them at early stages disappears (Knee et al., 1977). Due to the increase in cell size the cells are connected only by small parts at the tips. Hydration of the middle lamella and wall matrix material become extreme at the ripening phase. Nearly three quarters of the polyuronide in the hydrated cell-wall become freely soluble, including arabinose and galactose residues (Knee et al., 1977).

In similarity to other fruit, softening of strawberry during ripening might be attributed to the enzymatic degradation of cell wall material by cell wall hydrolases such as polygalcturonases (PGs), pectinmethylesterases (PMEs) and cellulases (CELSs). While PMEs and CELs enzyme activities have been clearly demonstrated the evidence for PGs activity has been somewhat contradicting. Barnes and Patchet, (1976) could not identify PG activity in crude extracts of the different fruit developmental stages. Later Nogata et al. (1993) reported on 3 different PG activities, which were purified from green stage fruit, and declined during fruit ripening. Recently the isolation of a cDNA showing homology to PG from the cultivated strawberry has been reported (*spG* gene, Redondo-Nevado et al., 2001). RNA gel blot analysis showed that *spG* gene expression is restricted to the receptacle tissue and the white stage of fruit development. Two other cDNAs putatively encoding PGs have been isolated by our group, showing different gene expression profiles during fruit development (Aharoni and O'Connell, *In press*). Barnes and Patchet, (1976) demonstrated pectinmethylesterase activity, which reached a maximum between the white and the red stages of development prior to its decline in the red stage. A strawberry pectinmethylesterase homolog was identified by our group and its expression matches the enzymatic activities described earlier (Aharoni and O'Connell, *In press*). From amongst the

different hydrolazes, genes encoding CELs have been the most investigated. A few reports demonstrate the existence of at least two CEL genes which showed receptacle specific and ripening regulated expression profile (Manning, 1998; Harpster et al., 1998; Trainotti et al., 1999a; Trainotti et al., 1999b; Llop-Tous et al., 1999). Ripening regulated CELs enzyme activity was detected earlier by Barnes and Patchet (1976), Abeles and Takeda. (1990) and later by Trainotti et al. (1999b). The cloning and characterisation of other genes encoding enzymes related to cell-wall metabolism was also reported and they include 3 different beta-galactosidases genes (Trainotti et al., 2001) a pectate lyase homolog (Medina-Escobar et al., 1997) and a family of six expansin genes (Civello et al., 1999; Harrison et al., 2001). Expansins do not hydrolyse cell wall components but promote cell wall loosening and catalyse wall extension by binding to the surface of cellulose microfibriles thereby disrupting the hydrogen bonds formed with xyloglucan molecules (McQueen-Mason et al., 1992).

Apart from changes in texture strawberry fruit undergo dramatic changes in the levels of different types of phenolic compounds, which provide it with colour, flavour and resistance to pathogenic attack and environmental conditions such as UV exposure. During early stages, non-tannin flavonoids and mainly condensed tannins accumulate to high levels and give strawberry an astringent flavour (Cheng and Breen, 1991). Later in development when fruit start to ripen other flavonoids like anthocyanin (mainly pelargonidin glucoside) and flavonols accumulate to high levels. The accumulation of anthocyanines parallels the rise in activities of both phenylalanine ammonia-lyase (the initial step of the phenylpropanoid pathway) and UDP-glucose:flavonoid-3-O-glucosyltransferase (adding the sugar molecule to the anthocyanines). Over the past years the majority of genes encoding flavonoid biosynthesis enzymes catalysing steps in the production of anthocyanines in strawberry have been isolated (Moyano et al., 1998; Manning et al., 1998; Aharoni et al., 2000; Deng and Davis, 2001). A subset of the genes have been recently used in a candidate gene approach to identify the wild type allele (*C*) of the *c* (yellow fruit color) locus, which possibly encodes the necessary protein for red fruit color in the wild strawberry fruit (Deng and Davis, 2001). Ellagic acid, a phenolic constituent of many plants derived as well from the phenylpropanoid pathway is present in the various strawberry plant parts including the achenes and receptacle and its levels dramatically decrease during ripening (Maas et al., 1991).

The levels of sugars and organic acids and the ratios between them play a significant role in the overall flavour of fruit. The major soluble sugars in strawberry are glucose, fructose and sucrose, and there levels rise during fruit maturation. Sucrose is present in lower concentrations and starts to accumulate later in fruit development (Hancock, 1999). In contrast, concentrations of minor soluble sugars such as inositol, xylose and galactose decrease during maturation (Moing et al., 2001). Invertases which irreversibly catalyse the hydrolyses of sucrose are present in multiple forms in plants mainly as wall-bound and soluble acid types. It is believed that they may play an important role in the determination of fruit sweetness. Both types of activities (i.e. wall bound and soluble acid) were characterised in strawberry and the isolation of a fruit cell-wall invertase was

reported by Manning (GenBank accession no. AF000520). In contrast to soluble sugars (major or minor) the levels of total organic acids in strawberry, rise until the turning stage and then they decrease during ripening. The three main organic acids in the fruit are citrate, malate and quinate, and minor organic acids include acetate, oxalate, succinate, isocitrate, fumarate, and aconitate (Moing et al., 2001). Citrate is the main organic acid present in strawberry and it influences the decline in fruit titratable acidity during the red ripe stage. The PH of the fruit changes slightly from 4.6 in early stages to 3.3 at the turning stage, rising to 3.7 at the ripe stage (Woodward, 1972). Amino acids are another soluble component which significantly contribute to fruit flavour either imparting flavour by themselves or through the supply of precursors for the biosynthesis of aroma components. In addition amino acid in fruit serve as building blocks for the synthesis of proteins and as a source of nitrogen. Asparagine is the dominant amino acid in all strawberry developmental stages and together with glutamine serves as the major nitrogen–transport compounds (Perkins-Veazie, 1995a). The total concentration of free amino acids in strawberries is much higher during maturation, and this might be a result of extensive protein metabolism occurring during ripening (Moing et al., 1991).

The biosynthesis of flavour/aroma compounds, which contribute to the characteristic aroma of strawberry, is an important process activated upon fruit ripening. Most of the research to date has been focused on identifying the compounds involved and activities of enzymes contributing to strawberry flavour/aroma. For example, lipoxygenase and hydroperoxide lyase activities were studied which are part of the biosynthesis of aldehydes, alcohols and esters which are important contributors to strawberry aroma (Perez et al., 1999). Both activities were shown to increase co-ordinately during ripening. Only a minor set of genes were associated with the biosynthesis of aroma and flavour compounds and they include for example the *SAAT* gene which encodes the ester forming enzyme in strawberry, and a putative pyruvate decarboxylase showing elevated expression during ripening and probably contributes to the levels of acetaldehyde in the ripening fruit (Aharoni et al., 2000). A more detailed review on strawberry fruit flavour components and findings related to their metabolism is provided below.

Aroma of Fruit

General

The flavours of foods including those of fruit are a combination of taste (e.g. sweetness and acidity) and odor, and are influenced by sensations of color, heat, cold and texture. Volatile components are crucial in determining the typical odor characteristics of individual flavours. Plant volatiles including those emitted by fruit may be divided to two classes. The first class encompasses those volatiles that are direct metabolites produced in plant organs by intracellular

biogenetic pathways. These type of volatiles might be regarded as “natural” or “constitutive” whilst a second class often termed “secondary” volatiles are often produced very quickly as a response to cell disruption (wounding), pathogen attack etc.. An excellent example of the production of “secondary” volatiles is their enzymatic synthesis via the lipoxygenase pathway of aliphatic C6 compounds (“green-grassy” notes) in disrupted tomato fruit tissue (Riley and Thompson, 1998).

Early research on fruit odor focused on identifying odor components present in the different fruit species. A second objective was to characterise which volatiles were key volatiles conveying the characteristic odor unique to a particular fruit. Later, researchers began investigating their biogenesis and the effect that processing and storage imposed on them. In general the odor or aroma components of a fruit comprise more than a hundred constituents, belonging to several different chemical classes. Their concentration will often vary between the different fruit tissue and will represent 10 to 100 ppm of the total fruit material. They are generally formed from non-volatile precursors, and in some fruit, such as citrus will accumulate in specialized structures adapted to accumulate high levels of them (Turner et al., 1998).

In order to study flavour volatiles one should first isolate the volatiles from a mixture with other non-volatile components. The volatiles are present in very low amounts, some of them might be even below the detection limit, therefore it is preferable that they be concentrated in order to obtain sufficient amount for analysis. For these reasons, the isolation of specific volatiles that contribute to the characteristic flavour and odour of fruit and their characterisation is rather difficult. Distillation (e.g. flash, steam, vacuum, CO_2), extraction (e.g. solvent, CO_2), adsorption (e.g. charcol), gas entrainment (open and closed systems) and freezing (e.g. lyophilization) are the main methods employed today for the extraction and concentration of volatiles (Schreier, 1984). Obviously, sample composition and compounds of interest will determine the procedure of choice. For plant volatiles, gas chromatography coupled to mass spectrometry (GC-MS) is the technique of choice for their separation and identification.

Biogenesis of Strawberry and Other Fruit Volatiles

Combined biochemical and molecular analyses of volatile components released by strawberry has demonstrated that their biogenesis forms an integral part of the ripening program (Maarse, 1991). In melon, blocking ethylene production using an antisense approach, resulted in a decrease of up to 10% in the production of the most potent volatile esters in the fruit (Bauchot et al., 1998). More than 15 compounds, including aldehydes, alcohols, ketones, sulfur containing compounds and a phenol were deficient in fruit of two tomato mutants, *rin* and *nor* strongly ripening inhibited (McGlasson et al., 1987). The biogenetic pathways of fruit volatiles in plant can be derived from enzymatically controlled lipid, terpene, amino acid, carbohydrate and phenylpropane metabolism. Production of fruit volatiles occurs mainly during the ripening phase although several volatiles

mainly the C6 components providing green notes often accumulate earlier in development. During the course of ripening the metabolism of the fruit mainly changes to catabolism in which high molecular weight structures such as proteins, polysaccharides and lipids are converted to ripening metabolites amongst them volatiles (Figure 3; Leahy and Roderick, 1999). In parallel there is active synthesis of a set of enzymes, which will catalyse these reactions. Another important factor which influences fruit volatiles biogenesis is fruit structure including cell wall hydrolysis and decomposition of cell membranes during ripening.

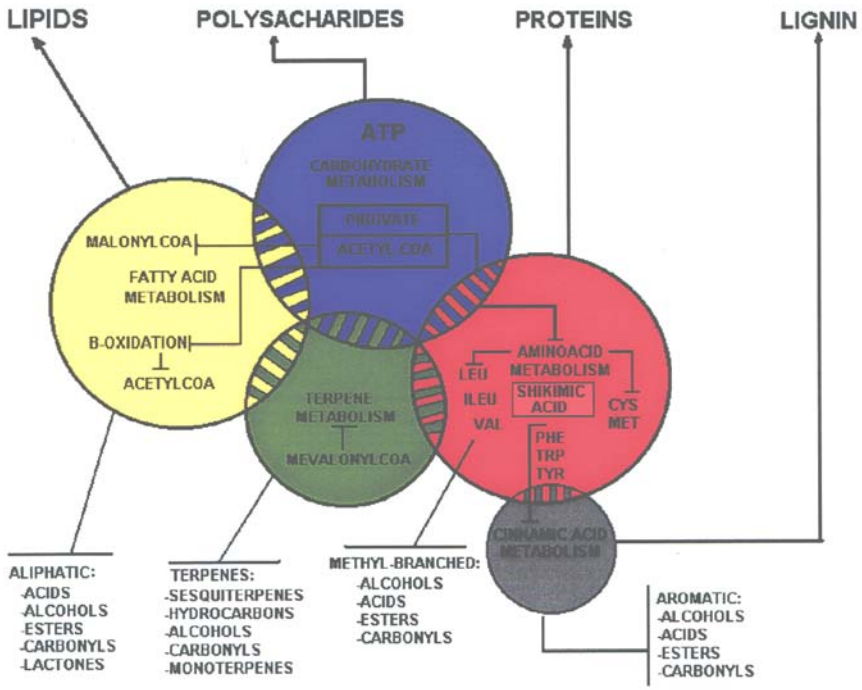


Figure 3 Schematic of the metabolic routes leading to the formation of different groups of volatile flavour compounds in fruits

This causes the release of bound enzymes, contributing precursors to the biosynthesis of volatiles and also their storage, such as in the case of citrus (in glandular cells of the exocarp). The formation of fruit volatiles is also influenced by external factors such as temperature, humidity and light (Leahy and Roderick, 1999).

In similarity with other fruit, a complex mixture of hundreds of compounds (more than 300) determines strawberry aroma (Zabetakis and Holden, 1997). The components identified may be grouped into several chemical classes, which include acids, aldehydes, ketones, alcohols, esters, and lactones (Figure 4). Other contributing groups are sulphur compounds, acetals, furans, phenols, terpenes and epoxides. Members of these groups, whilst often present at low levels, may have a significant impact on the overall aroma of strawberry. The compound 2,5-dimethyl-4-hydroxy-3(2H)-furanone (furaneol) and its methyl ether (2,5-dimethyl-4-methoxy-3(2H)-furanone) were the only compounds identified in diluted solutions, as exhibiting typical flavour/aroma

associated with strawberry (Roscher et al., 1997). The biochemical pathway leading to the synthesis of furaneol and its methyl ether is unknown although there is strong evidence that sugars (most likely fructose-1, 6-bisphosphate) supply precursors for furaneol biosynthesis (Schwab, 1998). Except carbohydrates the diverse volatile constituents of fruit like strawberry are derived from various other primary metabolism precursors such as fatty acids and amino acids.

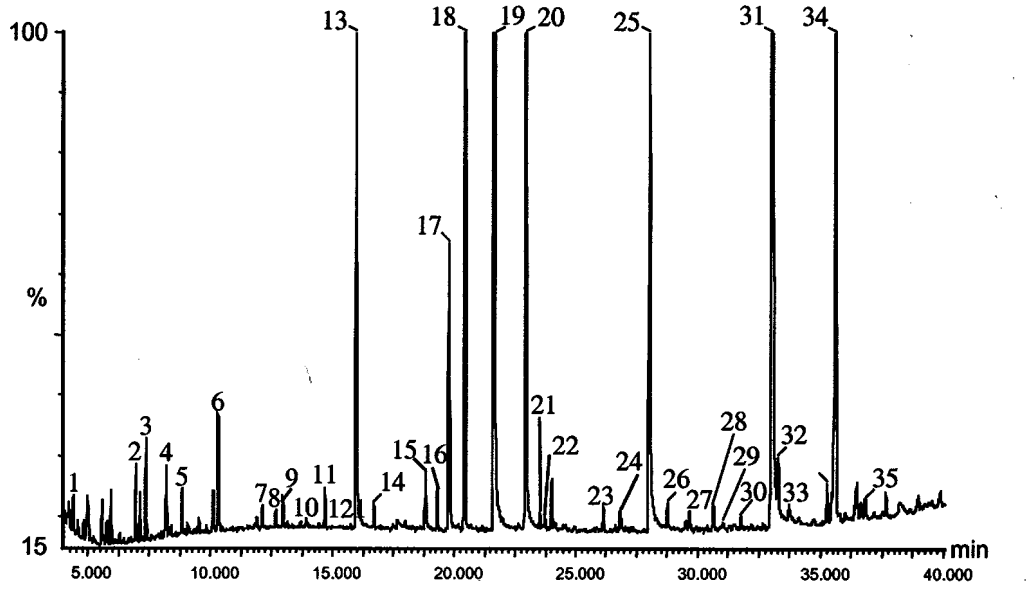


Figure 4 Volatile constituents of ripe strawberry fruit (cv. Elsanta) detected by GC-MS after XAD solid phase extraction. Peak numbers correspond to: 1- butyl acetate; 2- 4 methyl-2-butenal; 3- methyl hexanoate; 4- *E*-2-hexanal; 5- ethyl hexanoate; 6- 3-hydroxy-2-butanone; 7- *E*-2-hexenyl acetate; 8- *E*-rose oxide; 9- hexanol; 10- *Z*-3-hexenol; 11- *E*-2-hexenol; 12- *E*-linalooloxid, furanoid; 13- acetic acid; 14- *Z*-linalooloxid, furanoid; 15- propanoic acid; 16- linalool; 17- 2-methyl propionic acid; 18- methoxyfuraneol; 19- butanoic acid; 20- 2-methyl butanoic acid; 21- gamma-hexalactone; 22- 4-ethyl benzaldehyde; 23- delta-hexalactone; 24- 4-methyl pentanoic acid; 25- hexanoic acid; 26- benzylalcohol; 27- gamma octalactone; 28- benzothiazole; 29- delta-octalactone; 30- furaneol acetate; 31- furaneol; 32- *E*-nerolidol; 33- octanoic acid; 34- gamma-decalactone; 35- delta-decalactone. (The image is courtesy of W. Schwab).

Volatile esters are quantitatively and qualitatively the most important compounds providing fruity odors, and in strawberry alone more than a hundred ester types have been detected (Maarse, 1991). A pathway leading to the biosynthesis of volatile esters in fruits has been proposed previously (Olias et al., 1995). Fatty acids are most important precursors for the formation of volatile aroma compounds. The oxidative degradation of membrane lipids to linolenic and linoleic acid catalysed by lypoxygenase is the main source of precursors for the generation of alcohols, aldehydes, acids and esters found in fruits (Perez et al., 1999). Thiolase is the last enzyme in the β-oxidation of fatty acids. It catalyses the thiolytic cleavage involving another molecule of CoA. The product of this reaction is acetyl-CoA and acyl CoA derivatives containing two carbon atoms less than the original acyl-CoA molecule that underwent oxidation. The acyl CoA formed in the

cleavage reaction may be utilized at the final stage of the biosynthetic pathway for ester formation in fruit (Bojorquez and Gomez-Lim, 1995). Therefore the profile of fatty acid precursors found in each fruit, along with the specificity of the enzymes in the biosynthetic pathway leading to ester formation could have a key role in determining the type of esters formed.

In addition to fatty acids the transamination and oxidative decarboxylation of amino acids provides the precursors for volatile aroma compounds such as aldehydes, acids, alcohols, esters and thiols (Perez et al., 1992). The transamination of amino acids is catalyzed by aminotransferases. In strawberry alanine is proposed to be the main free amino acid metabolized to flavor compounds (Perez et al., 1992). This was based mainly on the dramatic decrease in its content just before the formation of volatile aroma compounds had commenced in the fruit. In addition feeding alanine to strawberry cultures resulted in the formation of several esters such as methyl and ethyl hexanoate which are important constituents of strawberry volatiles (Perez et al., 1992). Alcohols and aldehydes are derived from the metabolism of their corresponding amino acids and oxo-sugars. Ketoacids produced by transamination can be enzymatically degraded to the corresponding aldehydes or carboxylic acids. The enzyme which catalyses this reaction is pyruvate decarboxylase. Alcohol dehydrogenase enzymes have been implicated in the interconversion of the aldehyde and alcohol forms of flavor volatiles (Scharpf, 1989). Esterification is the result of transacylation from acyl-coenzyme A (CoA) to an alcohol (Perez et al., 1996). The enzyme catalyzing the reaction is termed an alcohol acyltransferase (AAT), and it plays a major role in the biosynthesis of volatile esters. Volatile components might be glyco-conjugated and thus “stored” in fruit as non-volatile components. Free volatile compounds may be released from this odourless glycosides by either acidic or enzyme mediated hydrolysis (Perez et al., 1997). A large number of glycosylated volatile components have been identified from both the wild and cultivated strawberries, amongst them furaneol, cinnamic acid, benzoic acid, benzyl alcohol and hexanoic acid (Wintoch et al., 1991).

Molecular Studies Related to Fruit Flavour and Aroma

To date only a few molecular studies in relation to fruit odor have been described. Alcohol dehydrogenase (ADH) implicated in the inter-conversion of the aldehydes and alcohol forms of flavor volatiles has been isolated from several fruit species. The tomato *ADH2* gene was shown to accumulate in the fruit during ripening (Longhurst et al., 1994) and was capable of performing the appropriate reactions *in-vitro* (Bicsak et al., 1982). Spiers et al. (1998) provided a direct proof for *ADH2 in- vivo* function by overexpressing it under the control of either a constitutive or a fruit specific promoter. They obtained tomato plants, which showed either high or low ADH activity (low due to suppression) and modified in the balance between odor related aldehydes and alcohols. The gene encoding the enzyme pyruvate decarboxylase catalyzing the conversion of pyruvate to acetaldehyde, which in turn may be converted to the alcohol ethanol and as a result contributes to

ethyl ester biosynthesis was reported in several fruit species including strawberry (Aharoni et al., 2000).

Esters contribute to the aroma of many fleshy fruit species and various investigators have studied the activity of ester-forming enzyme (an alcohol acyltransferase, AAT) in different fruits such as banana, melon and strawberry (e.g. Harada et al., 1985; Shalit et al., 2001; Olias et al., 1995; Perez et al., 1993). Recently the isolation and characterization of the gene encoding the strawberry AAT (*SAAT*) was described (Aharoni et al., 2000). *SAAT* showed a fruit specific gene expression pattern and enzymatic activity studies with the recombinant protein demonstrated its ability to utilize a large set of alcohols and acyl CoAs. The broad substrate specificity might account for the formation of more than a hundred different ester types in a single fruit such as strawberry during ripening. A more extensive description of the *SAAT* gene and its corresponding protein is provided in chapter 7 of this thesis.

Several other genes, which might encode enzymes, associated with the biogenesis of fruit volatiles; mainly the production of fatty acids precursors were also isolated and partially characterized from different fruit. These include a strawberry Δ9 desaturase, an acyl carrier protein (ACP) and a malonyl-CoA decarboxylase (Manning, 1998; Aharoni et al., 2000), and a raspberry Δ9 desaturase (Jones et al., 2000). However, the *in-vivo* function of these genes remains to be determined.

Scope of the Thesis

The research presented in this thesis describes a comprehensive study of strawberry fruit development and ripening and the molecular genetic, biochemical and physiological processes associated with it (Figure 5). This study was from the first to apply microarray technology to the global analysis of gene expression in plants. The methods, principals, and applications of microarray technology in functional genomics approaches are reviewed in *Chapter 2*. Analysis of gene expression using cDNA microarrays, during fruit development , under stress and hormonal treatment (*Chapter 3*), and in achene and receptacle tissues (*Chapter 4*), resulted in the identification of ripening associated, tissue specific, auxin dependent and independent genes and processes in the ripening strawberry fruit.

Chapter 5 and *Chapter 6* describe the isolation and characterization of genes encoding the last committed steps in the biosynthesis of volatile flavor/aroma components in the ripe strawberry fruit. Various flavor/aroma candidate genes, including those involved in the production of volatile esters and terpenoids (mono- and sesquiterpene), were identified either by microarray analysis or by other established cloning methods. The *SAAT* gene described in *Chapter 5*, encodes the ester-forming enzyme from strawberry and its characterization provided us with valuable insight into how more than a hundred types of esters can be simultaneously produced in the ripening tissue of a single fruit. *Chapter 6* describes our investigation of another group of volatile flavors, the terpenoids. Gene and protein structure and. The study shed light on the evolutionary events that lead to the cessation of the production of specific monoterpene molecules in the octaploid cultivated varieties, and on the other hand the acquisition of novel compounds during domestication, not present in the diploid wild species.

Chapter 7 of the thesis reports the first study of the regulation of a ripening related process and metabolic pathway in the ripe strawberry fruit. The strawberry *FaMYB1* gene encoding a member of the MYB family of transcription factors, showing a ripening regulated expression pattern, was characterized by overexpression studies in transgenic tobacco. The results suggest that FaMYB1 is involved in the control of flavonoid biosynthesis in late strawberry ripening, possibly acting as a repressor.

Chapter 8 describes the application of another emerging technology, Fourier Transform Ion Cyclotron Mass-Spectrometry (FTMS) to the study of metabolite changes occurring in fruit during the transition from immaturity to ripening. In the same chapter the utility of the method for functional genomic approaches is further demonstrated by the non-targeted metabolic analysis of a transgenic tobacco line overexpressing a strawberry transcription factor.

In the summary, at the end of this thesis, the results and conclusions from all chapters of the thesis are presented and collectively discussed. Prospects for future research based on our findings are also disclosed.

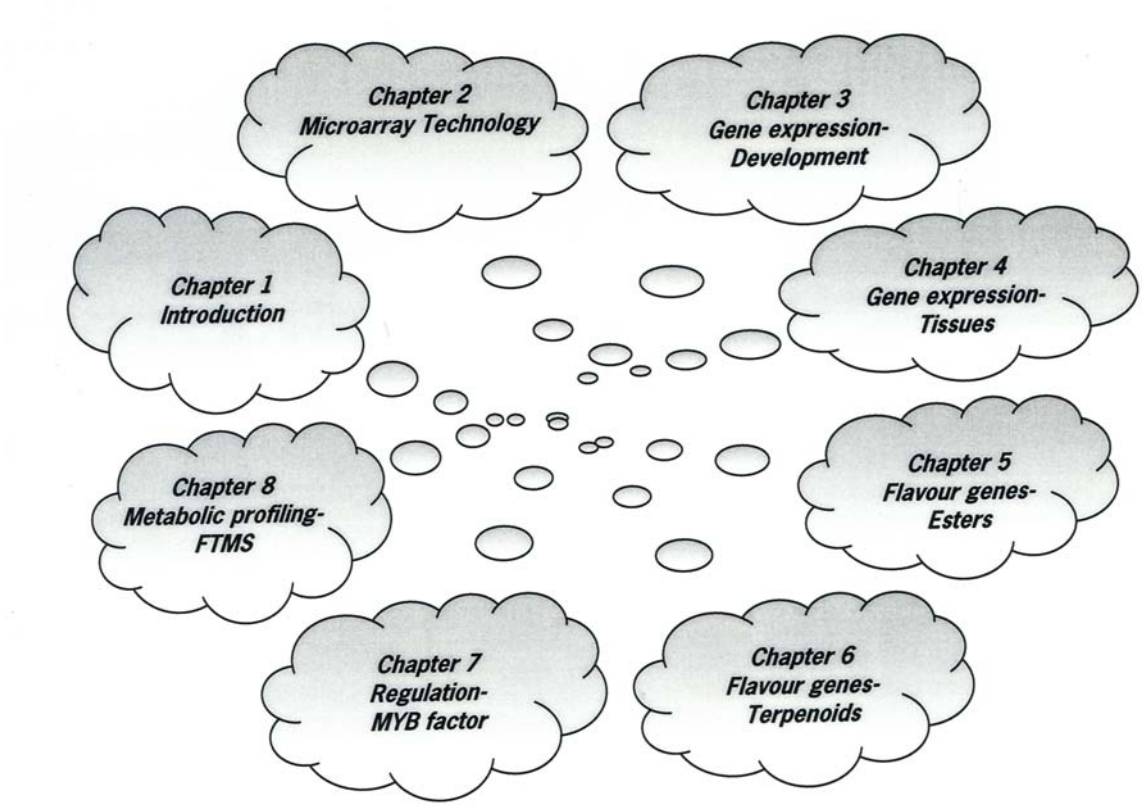


Figure 5 Schematic outline of the thesis "Strawberry and Beyond: A Novel and Comprehensive Investigation of Fruit Maturation and Ripening".

REFERENCES

Abeles FB, Takeda F (1990) Cellulase activity and ethylene in ripening strawberry and apple fruits. Sci Hort **42:** 269-275

Aharoni A, Keizer LCP, Bouwmeester HJ, Sun ZK, Alvarez Huerta M, Verhoeven HA, Blaas J, van Houwelingen A, De Vos RCH, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O'Connell AP (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12: 647-661

Aharoni A, Vorst F (2002) DNA microarrays for functional plant genomics. Plant Mol Biol **48:** 99-118

Alonso JM, Chamaro J, Granell A (1995a) Evidence for the involvement of ethylene in the expression of specific RNAs during maturation of the orange, a non-climacteric fruit. Plant Mol Biol **29:** 385-390

Alonso JM, Granell A (1995b) A putative vacuolar processing protease is regulated by ethylene and also during fruit ripening in Citrus fruit. Plant Physiol **109:** 541-547

Andersen JS, Mann M (2000) Functional genomics by mass spectrometry. FEBS Lett **480:** 25-31

Archbold DD, Dennis FG, Jr. (1984) Quantification of free ABA and free and conjugated IAA in strawberry achene and receptacle tissue during fruit development. J. Am Soc Hortic Sci **109:** 330-335

Ayub R, Guis M, BenAmor M, Gillot L, Roustan JP, Latche A, Bouzayen M, Pech JC (1996) Expression of ACC oxidase antisense gene inhibits ripening of cantaloupe melon fruits. Nat Biotechnol **14:** 862-866

Bachem CWB, Van Der Hoeven RS, De Bruijn SM, Vreugdenhil D, Zabeau M, Visser, RGF (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. Plant J **9:** 745-753

Barnes MF, Patchett BJ (1976) Cell wall degrading enzymes and the softening of senescent strawberry fruit. J Food Sci **41:** 1392-1395

Bauchot AD, Mottram DS, Dodson AT, John P (1998) Effect of aminocyclopropane-1-carboxylic acid oxidase antisense gene on the formation of volatile esters in cantaloupe charentais melon (Cv. Vedrandais). J Agric Food Chem **46:** 4787-4792

Bicsak TA, Kann LR, Reiter A, Chase T, Jr. (1982) Tomato alcohol dehydrogenase: purification and substrate specificity. Arch Biochem Biophys **216:** 605-615

Bojorquez G, Gomez-Lim MA (1995) Peroxisomal thiolase mRNA is induced during mango fruit ripening. Plant Mol Biol **28:** 811-820

Bouchez D, Hofte H (1998) Functional genomics in plants. Plant Physiol **118:** 725-732

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo SJ, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol **18:** 630-634

Brownleader MD, Jackson P, Mobasheri A, Pantelides AT, Sumar S, Trevan M, Dey PM (1999) Molecular aspects of cell wall modifications during fruit ripening. Crit Rev Food Sci Nutr **39:** 149-164

Brummell DA, Harpster MH, Dunsmuir P (1999a) Differential expression of expansin gene family members during growth and ripening of tomato fruit. Plant Mol Biol **39:** 161-169

Brummell DA, Harpster MH, Civello PM, Palys JM, Bennett AB, Dunsmuir P (1999b) Modification of expansin protein abundance in tomato fruit alters softening and cell wall polymer metabolism during ripening. Plant Cell **11:** 2203-2216

Cazzonelli CI, Cavallaro AS, Botella JR (1998) Cloning and characterization of ripening-induced ethylene biosynthetic genes from non-climacteric pineapple (Ananas comosus) fruits. Aust J Plant Physiol **25:** 513-518

Cheng GW, Breen PJ (1991) Activity of phenylalanine ammonia-lyase (PAL) and concentrations of anthocyanins and phenolics in developing strawberry fruit. J Am Soc Hortic Sci **116:** 865-869

Cheng GW, Breen PJ (1992) Cell count and size in relation to fruit size among strawberry cultivars. J Am Soc Hortic Sci **117:** 946-950

Civello PM, Powell ALT, Sabehat A, Bennett AB (1999) An expansin gene expressed in ripening strawberry fruit. Plant Physiol **121:** 1273-1279

Clendennen SK, May GD (1997) Differential gene expression in ripening banana fruit. Plant Physiol **115:** 463-469

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16:** 735-743

D' Aoust M, Yelle S, Binh NQ (1999) Antisense inhibition of tomato fruit sucrose synthase decreases fruit setting and the sucrose unloading capacity of young fruit. Plant Cell **11:** 2407-2418

Davies C, Boss PK, Robinson SP (1997) Treatment of grape berries, a nonclimacteric fruit with a synthetic auxin, retards ripening and alters the expression of developmentally regulated genes. Plant Physiol **115:** 1155-1161

Davies C, Robinson SP (2000) Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins. Plant Physiol **122:** 803-812

Delneri D, Brancia FL, Oliver SG (2001) Towards a truly integrative biology through the functional genomics of yeast. Curr Opin Biotech **12:** 87-91

Deng C, Davis TM (2001) Molecular identification of the yellow fruit color (c) locus in diploid strawberry: a candidate gene approach. Theor Appl Genet **103:** 2-3

Desprez T, Amselem J, Caboche M, Hofte H (1998) Differential gene expression in Arabidopsis monitored using cDNA arrays. Plant J **14:** 643-652

Dutt MJ, Lee KH (2000) Proteomic analysis. Curr Opin Biotechnol **11:** 176-179

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA **95:** 14863-14868

El Mansouri I, Mercado JA, Valpuesta V, Lopez Aranda JM, Pliego Alfaro F, Quesada MA (1996) Shoot regeneration and Agrobacterium-mediated transformation of Fragaria vesca L. Plant Cell Rep **15:** 642-646

Fiehn O, Kopka J, Doermann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol **18:** 1157-1161.

Frary A, Nesbitt TC, Grandillo S, van der Knaap E, Cong B, Liu JP, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. Science **289:** 85-88

Frenkel C, Peters JS, Tieman DM, Tiznado ME, Handa AK (1998) Pectin methylesterase regulates methanol and ethanol accumulation in ripening tomato (Lycopersicon esculentum) fruit. J Biol Chem **273:** 4293-4296

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci USA **97:** 4718-4723

Gillaspy G, Ben David H, Gruissem W (1993) Fruits: a developmental perspective. Plant Cell **5:** 1439-1451

Giovannoni J (2001) Molecular biology of fruit maturation and ripening. Ann Rev Plant Physiol Plant Mol Biol **52:** 725-749

Given NK, Venis MA, Grierson D (1988) Hormonal regulation of ripening in the strawberry, a non-climacteric fruit. Planta **174:** 402-406

Glassbrook N, Ryals J (2001) A systematic approach to biochemical profiling. Curr Opin Plant Biol **4:** 186-190

Goldschmidt EE, Huberman M, Goren R (1993) Probing the role of endogenous ethylene in the degreening of citrus fruit with ethylene antagonists. Plant Growth Regul **12:** 325-329

Gygi SP, Aebersold R (2000) Mass spectrometry and proteomics. Cur Opin Chem Biol **4:** 489-494

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol **17:** 994-999

Hackett RM, Ho CW, Lin ZF, Foote HCC, Fray RG, Grierson D (2000) Antisense inhibition of the Nr gene restores normal ripening to the tomato Never-ripe mutant, consistent with the ethylene receptor-inhibition model. Plant Physiol **124:** 1079-1085

Hadfield KA, Dang T, Guis M, Pech JC, Bouzayen M, Bennett AB (2000) Characterization of ripening-regulated cDNAs and their expression in ethylene-suppressed Charentais melon fruit. Plant Physiol **122:** 977-983

Hancock JF (1999) Strawberries. CABI Publishing, Wallingford

Harada M, Ueda Y, Iwata T (1985) Purification and some properties of alcohol acetyltransferase from banana fruit. Plant and Cell Physiol **26:** 1067-1074

Harpster MH, Brummell DA, Dunsmuir P (1998) Expression analysis of a ripening-specific, auxin-repressed endo-1,4-beta-glucanase gene in strawberry. Plant Physiol **118:** 1307-1316

Harrison EP, McQueen-Mason SJ, Manning K (2001) Expression of six expansin genes in relation to extension activity in developing strawberry fruit. J Exp Bot **52:** 1437-1446

Havis AL (1943) A developmental analysis of strawberry fruit. Am J Bot **30:** 311-314

Hornstein I (1999) Fruit flavor biogenesis. *In* MM Leahy, RG Roderick, R Teranishi, EL Wick eds, Flavor chemistry: thirty years of progress. Kluwer Academic / Plenum Publishing Corporation, New York, pp 275-285

Itai A, Tanabe K, Tamura F, Tanaka T (2000) Isolation of cDNA clones corresponding to genes expressed during fruit ripening in Japanese pear (Pyrus pyrifolia Nakai): involvement of the ethylene signal transduction pathway in their expression. J Exp Bot **51:** 1163-1166

Jacob Wilk D, Holland D, Goldschmidt EE, Riov J, Eyal Y (1999) Chlorophyll breakdown by chlorophyllase: isolation and functional expression of the Chlase1 gene from ethylene-treated Citrus fruit and its regulation during development. Plant J **20:** 653-661

James DJ, Passey AJ, Barbara DJ (1990) Agrobacterium-mediated transformation of the cultivated strawberry (Fragaria X ananassa Duch.) using disarmed binary vectors. Plant Sci Lim **69:** 79-94

Jones CS, Davies HV, Taylor MA (2000) Profiling of changes in gene expression during raspberry (Rubus idaeus) fruit ripening by application of RNA fingerprinting techniques. Planta **211:** 708-714

Kahkonen MP, Hopia AI, Heinonen M (2001) Berry phenolics and their antioxidant activity. J Agric Food Chem **49:** 4076-4082

Katona ZF, Sass P, Molnar-Perl I (1999) Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography-mass spectrometry. J Chrom **847:** 91-102.

Kim SK (2001) Functional genomics: The worm scores a knockout. Curr Biol **11:** R85-R87

Knee M, Sargent JA, Osborne DJ (1977) Cell wall metabolism in developing strawberry fruits. J Exp Bot **28:** 377-396

Kuntz M, Chen HC, Simkin AJ, Romer S, Shipton CA, Drake R, Schuch W, Bramley PM (1998) Upregulation of two ripening-related genes from a non-climacteric plant (pepper) in a transgenic climacteric plant (tomato). Plant J **13:** 351-361

Lanahan MB, Yen HC, Giovannoni JJ, Klee HJ (1994) The Never ripe mutation blocks ethylene perception in tomato. Plant Cell **6:** 521-530

Leahy MM, Roderick RG (1999) Fruit flavor biogenesis, Flavor chemistry, pp 275-285

Lelievre Jean M, Latche A, Jones B, Bouzayen M, Pech JC (1997) Ethylene and fruit ripening. Physiol Plant **101:** 727-739

Llop-Tous I, Dominguez Puigjaner E, Palomer X, Vendrell M (1999) Characterization of two divergent endo-beta-1,4-glucanase cDNA clones highly expressed in the nonclimacteric strawberry fruit. Plant Physiol **119:** 1415-1421

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol **14:** 1675-1680

Longhurst T, Lee E, Hinde R, Brady C, Speirs J (1994) Structure of the tomato Adh2 gene and Adh2 pseudogenes, and a study of Adh2 gene expression in fruit. Plant Mol Biol **26:** 1073-1084

Lu CG, Zainal Z, Tucker GA, Lycett GW (2001) Developmental abnormalities and reduced fruit softening in tomato plants expressing an antisense Rabl1 GTPase gene. Plant Cell **13:** 1819-1833

Lukowitz W, Gillmor CS, Scheible WR (2000) Positional cloning in arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiol **123:** 795-805

Luo Y, Liu X (1994) Effects of ethylene on RNA metabolism in strawberry fruit after harvest. J Hort Sci **69:** 137-139

Maarse H (1991) Volatile compounds in foods and beverages. Marcel Dekker, Inc., New York, pp 764

Maas JL, Wang SY, Galletta GJ (1991) Evaluation of strawberry cultivars for ellagic acid content. HortScience **26:** 66-68

Manning K (1991) Isolation of nucleic acids from plants by differential solvent precipitation. Anal Biochem **195:** 45-50

Manning K (1994) Changes in gene expression during strawberry fruit ripening and their regulation by auxin. Planta **194:** 62-68

Manning K (1998) Isolation of a set of ripening-related genes from strawberry: their identification and possible relationship to fruit quality traits. Planta **205:** 622-631

Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D (1999) A combined algorithm for genome-wide prediction of protein function. Nature **402:** 83-86

Martzen MR, McCraith SM, Spinelli SL, Torres FM, Fields S, Grayhack EJ, Phizicky EM (1999) A biochemical genomics approach for identifying genes by the activity of their products. Science **286:** 1153-1155

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol **123:** 439-442

McGlasson WB, Last JH, Shaw KJ, Meldrum SK (1987) Influence of the non-ripening mutants rin and nor on the aroma of tomato fruit. HortScience **22:** 632-634

McQueen-Mason S, Durachko DM, Cosgrove DJ (1992) Two endogenous proteins that induce cell wall extension in plants. Plant Cell **4:** 1425-1433

Medina-Escobar N, Cardenas J, Moyano E, Caballero JL, Munoz-Blanco J (1997) Cloning, molecular characterization and expression pattern of a strawberry ripening-specific cDNA with sequence homology to pectate lyase from higher plants. Plant Mol Biol **34:** 867-877

Meissner R, Chague V, Zhu Q, Emmanuel E, Elkind Y, Levy A (2000) A high throughput system for transposon tagging and promoter trapping in tomato. Plant J **22:** 265-274

Moing A, Renaud C, Gaudillere M, Raymond P, Roudeillac P, Denoyes-Rothan B (2001) Biochemical changes during fruit development of four strawberry cultivars. J Am Soc Hortic Sci **126:** 394-403

Moyano E, Portero-Robles I, Medina Escobar N, Valpuesta V, Munoz-Blanco J, Caballero JL (1998) A fruit-specific putative dihydroflavonol 4-reductase gene is differentially expressed in strawberry during the ripening process. Plant Physiol **117:** 711-716

Mysore KS, Touri RP, Martin GB (2001) Arabidopsis genome sequence as a tool for functional genomics in tomato. Genome Biol **2**

Nam YW, Tichit L, Leperlier M, Cuerq B, Marty I, Lelievre JM (1999) Isolation and characterization of mRNAs differentially expressed during ripening of wild strawberry (Fragaria vesca L.) fruits. Plant Mol Biol **39:** 629-636

Nitsch JP (1950) Growth and morphogenesis of the strawberry as related to auxin. Am J Bot **37:** 211-215

Nogata Y, Ohta H, Voragen AGJ (1993) Polygalacturonase in strawberry fruit. Phytochemistry **34:** 617-620

Ohlrogge J, Benning C (2000) Unraveling plant metabolism by EST analysis. Curr Opin Plant Biol **3:** 224-228.

Olias JM, Sanz C, Rios JJ, Perez AG (1995) Substrate specificity of alcohol acyltransferase from strawberry and banana fruits. ACS symposium series**:** 134-141

Pandey A, Mann M (2000) Proteomics to study genes and genomes. Nature **405:** 837-846

Parinov S, Sundaresan V (2000) Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. Cur Opin Biotech **11:** 157-161

Paton NW, Khan-Shakeel A, Hayes A, Moussouni F, Brass A, Eilbeck K, Goble CA, Hubbard SJ, Oliver SG (2000) Conceptual modelling of genomic information. Bioinformatics **16:** 548-557.

Pereira A (2000) A transgenic perspective on plant functional genomics. Transgenic Res **9:** 245-260

Perez AG, Cert A, Rios JJ, Olias JM (1997) Free and glycosidically bound volatile compounds from two banana cultivars: Valery and Pequena Enana. J Agric Food Chem **45:** 4393-4397

Perez AG, Rios JJ, Sanz C, Olias JM (1992) Aroma components and free amino acids in strawberry variety Chandler during ripening. J Agric Food Chem **40:** 2232-2235

Perez AG, Sanz C, Olias JM (1993) Partial purification and some properties of alcohol acyltransferase from strawberry fruits. J Agric Food Chem **41:** 1462-1466

Perez AG, Sanz C, Olias R, Rios JJ, Olias JM (1996) Evolution of strawberry alcohol acyltransferase activity during fruit development and storage. J Agric Food Chem **44:** 3286-3290

Perez AG, Sanz C, Olias R, Olias JM (1999) Lipoxygenase and hydroperoxide lyase activities in ripening strawberry fruits. J Agric Food Chem **47:** 249-253

Perkins-Veazie P (1995a) Growth and ripening of strawberry fruit. Hort Rev **17:** 267-297

Perkins-Veazie PM, Huber DJ, Brecht JK (1995b) Characterization of ethylene production in developing strawberry fruit. Plant Growth Regul **17:** 33-39

Perkins-Veazie PM, Huber DJ, Brecht JK (1996) In vitro growth and ripening of strawberry fruit in the presence of ACC, STS or propylene. Ann Appl Biol **128:** 105-116

Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol **19:** 45-50

Redgwell RJ, MacRae E, Hallett I, Fischer M, Perry J, Harker R (1997) In vivo and in vitro swelling of cell walls during fruit ripening. Planta **203:** 162-173

Redondo-Nevado J, Moyano E, Medina Escobar N, Caballero JL, Munoz-Blanco J (2001) A fruit-specific and developmentally regulated endopolygalacturonase gene from strawberry (Fragaria x ananassa cv. Chandler). J Exp Bot **52:** 1941-1945

Riley JCM, Thompson JE (1998) Ripening-induced acceleration of volatile aldehyde generation following tissue disruption in tomato fruit. Physiol Plant **104:** 571-576

Roessner U, Wagner C, Kopka J, Trethewey Richard N, Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J **23:** 131-142

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001a) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell **13:** 11-29

Roessner U, Willmitzer L, Fernie AR (2001b) High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies. Plant Physiol **127:** 749-764

Ronen G, Carmel Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to beta-carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold color mutations in tomato. Proc Natl Acad Sci USA **97:** 11102-11107

Roscher R, Schreier P, Schwab W (1997) Metabolism of 2,5-dimethyl-4-hydroxy-3(2H)-furanone in detached ripening strawberry fruits. J Agric Food Chem **45:** 3202-3205

Roubelakis-Angelakis KA, Kliewer WM (1986) Effects of exogenous factors on phenylalanine ammonia-lyase activity and accumulation of anthocyanins and total phenolics in grape berries. Am J Enol Vit **37:** 275-280

Scharpf LG, Jr., Chandan RC (1989) Biotechnology and natural flavors for yogurt. National Yogurt Association; Mclean, Virginia 22102; USA

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:** 467-470

Schreier P (1984) Chromatographic studies of biogenesis of plant volatiles. Huthig Verlag, Heidelberg, pp 171

Schwab W (1998) Application of stable isotope ratio analysis explaining the bioformation of 2,5-dimethyl-4-hydroxy-3(2H)-furanone in plants by a biological Maillard reaction. J Agric Food Chem **46:** 2266-2269

Seymour GB, Taylor JE, Tucker GA (1993) Biochemistry of fruit ripening. Chapman & Hall, London, pp 454

Shalit M, Katzir N, Tadmor Y, Larkov O, Burger Y, Shalekhet F, Lastochkin E, Ravid U, Amar O, Edelstein M, Karchi Z, Lewinsohn E (2001) Acetyl-CoA: alcohol acetyltransferase activity and aroma formation in ripening melon fruits. J Agric Food Chem **49:** 794-799

Sherlock G, Hernandez Boussard T, Kasarskis A, Binkley G, Matese John C, Dwight Selina S, Kaloper M, Weng S, Jin H, Ball Catherine A, Eisen Michael B, Spellman Paul T, Brown Patrick O, Botstein D, Cherry JM (2001) The Stanford Microarray Database. Nucleic Acids Res **29:** 152-155.

Shimkets RA, Lowe DG, J.T.N. T, Sehl P, Jin H, Yang R, Predki PF, Rothberg BEG, Murtha MT, Roth ME, Shenoy SG, Windemuth A, Simpson JW, Simons JF, Daley MP, Gold SA, McKenna MP, Hillan K, Went GT, Rothberg JM (1999) Gene expression analysis by transcript profiling coupled to a gene database query. Nat Biotechnol **17:** 798-803.

Speirs J, Lee E, Holt K, Kim Y, Scott NS, Loveys B, Schuch W, Kim YD (1998) Genetic manipulation of alcohol dehydrogenase levels in ripening tomato fruit affects the balance of some flavor aldehydes and alcohols. Plant Physiol **117:** 1047-1058

Speulman E, Metz PLJ, van Arkel G, Hekkert PTL, Stiekema WJ, Pereira A (1999) A two-component Enhancer-Inhibitor transposon mutagenesis system for functional analysis of the Arabidopsis genome. Plant Cell **11:** 1853-1866

Suutarinen J, Anakainen L, Autio K (1998) Comparison of light microscopy and spatially resolved Fourier transform infrared (FT-IR) microscopy in the examination of cell wall components of strawberries. Lebensm. Wiss. u.-Technol. **31:** 595-601

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR (1999) Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA **96:** 2907-2912

Tian MS, Prakash S, Elgar HJ, Young H, Burmeister DM, Ross GS (2000) Responses of strawberry fruit to 1-Methylcyclopropene (1-MCP) and ethylene. Plant Growth Regul **32:** 83-90

Tieman DM, Ciardi JA, Taylor MG, Klee HJ (2001) Members of the tomato LeEIL (EIN3-like) gene family are functionally redundant and regulate ethylene responses throughout plant development. Plant J **26:** 47-58

Tieman DV, Taylor MG, Ciardi JA, Klee HJ (2000) The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. Proc Natl Acad Sci USA **97:** 5663-5668

Trainotti L, Spolaore S, Pavanello A, Baldan B, Casadoro G (1999a) A novel E-type endo-beta-1,4-glucanase with a putative cellulose-binding domain is highly expressed in ripening strawberry fruits. Plant Mol Biol **40:** 323-332

Trainotti L, Ferrarese L, Vecchia Fd, Rascio N, Casadoro G, dalla Vecchia F (1999b) Two different endo-beta-1,4-glucanases contribute to the softening of strawberry fruits. J Plant Physiol **154:** 355-362

Trainotti L, Spinello R, Piovan A, Spolaore S, Casadoro G (2001) beta-galactosidases with a lectin-like domain are expressed in strawberry. J Exp Bot **52:** 1635-1645

Turner GW, Berry AM, Gifford EM (1998) Schizogenous secretory cavities of *Citrus limon* (L.) Burm. f. and a reevaluation of the lysigenous gland concept. Int J Plant Sci **159:** 75-88

Uetz P, Giot L, Cagney G, Mansfield-Traci A, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi EA, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403:** 623-627.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science **270:** 484-487

Vidal M (2001) A biological atlas of functional maps. Cell **104:** 333-339

Weigel D, Ahn J, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang Z, Xia Y, Dixon RA, Harrison MJ, Lamb CJ, Yanofsky MF, Chory J, Ahn JH, Wang ZY, Xia YJ (2000) Activation tagging in Arabidopsis. Plant Physiol **122:** 1003-1013

Wilkinson JQ, Lanahan MB, Conner TW, Klee HJ (1995a) Identification of mRNAs with enhanced expression in ripening strawberry fruit using polymerase chain reaction differential display. Plant Mol Biol **27:** 1097-1108

Wilkinson JQ, Lanahan MB, Yen H, Giovannoni JJ, Klee HJ, Yen HC (1995b) An ethylene-inducible component of signal transduction encoded by Never-ripe. Science **270:** 1807-1809

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol **126:** 485-493

Wintoch H, Krammer G, Schreier P (1991) Glycosidically bound aroma compounds from two strawberry fruit species, Fragaria vesca f. semperflorens and Fragaria X ananassa, cv. Korona. Flav Frag J **6:** 209-215

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, M ELB, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, N MR, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross MacDonald P, Scherens B, Snyder M, Sookhai Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu KX, Zimmermann K, Philippsen P, Johnston M, Davis RW (1999) Functional characterization of the S-cerevisiae genome by gene deletion and parallel analysis. Science **285:** 901-906

Woodward JR (1972) Physical and chemical changes in developing strawberry fruits. J Sci Food Agric **23:** 465-473

Zabetakis I, Holden MA (1997) Strawberry flavour: analysis and biosynthesis. J Sci Food Agric **74:** 421-434

Zhu H, Snyder M (2001) Protein arrays and microarrays. Cur Opin Chem Biol **5:** 40-45

Zhu T, Wang X (2000) Large-scale profiling of the Arabidopsis transcriptome. Plant Physiol **124:** 1472-1476

CHAPTER 2

Asaph Aharoni and Oscar Vorst

DNA Microarrays for Functional Plant Genomics

Plant Molecular Biology 48 (1-2): 99-118 (2001)

DNA MICROARRAYS FOR FUNCTIONAL PLANT GENOMICS

DNA microarray technology is a key element in today's functional genomics toolbox. The power of the method lies in miniaturisation, automation and parallelism permitting large-scale and genome-wide acquisition of quantitative biological information from multiple samples. DNA microarrays are currently fabricated and assayed by two main approaches involving either *in situ* synthesis of oligonucleotides ("oligonucleotide microarrays") or deposition of pre-synthesised DNA fragments ("cDNA microarrays") on solid surfaces. To date, the main applications of microarrays are in comprehensive, simultaneous gene expression monitoring and in DNA variation analyses for the identification and genotyping of mutations and polymorphisms. Already at a relatively early stage of its application in plant science, microarrays are being utilised to examine a range of biological issues, such as, for example, the circadian clock, plant defence, environmental stress responses, fruit ripening, phytochrome A signalling, seed development and nitrate assimilation. Novel insights are obtained into the molecular mechanisms co-ordinating metabolic pathways, regulatory and signalling networks. Exciting new information will be gained in the coming years not only from genome-wide expression analyses on a few model plant species, but also from extensive studies of less well studied species on a more limited scale. The value of microarray technology to our understanding of living processes will depend both on the amount of data to be generated and on its clever exploration and integration with other biological knowledge arising from complementary functional genomics tools for "profiling" the genome, proteome, metabolome and phenome.

Introduction

Controlling gene expression is one of the key regulatory mechanisms used by living cells to sustain and execute their function. Although the final activity of a gene is determined by the encoded protein, measurement of mRNA levels has proven to be a valuable molecular tool. For a few decades RNA-based assessment of gene expression has provided clues to gene function. Recently the availability of complete genome sequences and of large sets of Expressed Sequence Tags (ESTs) from numerous organisms triggered the development of efficient and accurate methods for large-scale and genome-wide analyses of genetic variation and gene expression patterns. As a result, several novel methods either sequence-based (Brenner *et al*., 2000; Velculescu *et al*., 1995), fragment-based (Bachem *et al*., 1996; Shimkets *et al*., 1999) or hybridisation-based such as macro- and microarrays (Desprez *et al*., 1998; Lockhart *et al*., 1996; Schena *et al*., 1995) are currently available (see also Breyene and Zabeau, 2001).

Microarray technology is a hybridisation-based method combining miniaturisation and the application of fluorescent dyes for labelling. The latter facilitates the combination of two differently labelled samples in a single hybridisation experiment and thus the use of competitive hybridisation to reduce experimental error. In this way relative expression levels of large numbers of genes can be determined simultaneously with a high degree of sensitivity. Today, two fundamentally different microarray-based technologies are available. Both are capable of large-scale expression analyses. A photolithographic method for high-density spatial synthesis of oligonucleotides was introduced by Fodor and colleagues. With this method arrays can be produced containing up to a few hundred thousand distinct elements (Fodor *et al*., 1991). As oligonucleotide arrays allow highly sensitive detection of DNA mismatches, they are well suited for DNA variation analysis as well. Manufacturing such arrays requires, however, prior sequence knowledge as well as complicated design and production methodologies (Lipshutz *et al*., 1999). The alternative method, in which pre-synthesised nucleic acids are mechanically deposited onto a solid surface, allows a more flexible design for the fabrication of microarrays (Duggan *et al*., 1999). In most cases PCR-amplified cDNA clones are used and the resulting arrays are referred to as cDNA microarrays. However, this technology can also be used to manufacture oligonucleotide arrays. In this article, we will focus on the use of both photolithographic oligonucleotide arrays and cDNA microarrays.

The basic microarray assay used with both types of array is similar and based on the specific hybridisation of a labelled sample to the immobilised nucleic acids (probe) on the array. As a result, the complex mixture of nucleic acids isolated from the biological sample under study is spatially separated into its constitutive components, the specific mRNAs. The physical separation on the array then enables the individual quantification of many specific mRNAs in a single hybridisation experiment. Furthermore, the independent detection of fluorescent signals at specific wavelengths allows simultaneous analysis of multiple dyes and thus mixed samples. Once data is

collected and normalised, expression ratios are obtained for each individual gene, representing relative expression levels for the samples investigated. Ultimately, biological meaning is inferred from data analyses of the comparison between samples and genes across one or multiple experiments and the combination with related biological knowledge.

Expression profiling using microarrays is currently being performed for numerous organisms, including several plant species, using an assortment of biological samples. The scale of these experiments ranges from a few hundred genes to genome-wide coverage (e.g. *Saccharomyces cerevisiae*, Lashkari *et al*., 1997; *Drosophila melanogaster*, Zou *et al*., 2000; *Caenorhabditis elegans,* Jiang *et al*., 2001). Next to its use for gene expression studies, microarrays are at present also being widely applied for DNA variation analyses (Lander, 1999). Variation in DNA sequence underlies much of the phenotypic differences that can be observed, not only between, but also within species and populations. Locating and identifying these genotypic differences allows linkage of genotypic and phenotypic variation. Point mutations, commonly referred to as Single Nucleotide Polymorphisms (SNPs), are the most frequent type of variation in genomes. Microarrays and in particular oligonucleotide arrays, may be used for large-scale SNP detection and discovery. In the case of organisms with small, sequenced genomes, such as yeast, microarrays may be used for the determination of genome-wide allelic variation (Lander, 1999).

This review describes various methodological aspects along with present and potential future applications of microarray technology for studying both whole genomes and gene function. We also summarise the recent exploitation of microarrays in plant science indicating a dramatic increase in the uses of the technique for surveying diverse biological events. The last section presents a perspective on how microarrays and other functional genomic tools can and ought to cross-fertilise each other in the years to come.

Principle of the Technology

Like other hybridisation-based analysis methods in molecular biology, the specificity of microarray technology relies on the selective and differential hybridisation of nucleic acids. Earlier methods, such as DNA and RNA gel blot analysis, use a unique, labelled nucleic acid molecule in solution. This so-called probe is hybridised to the complex mixture under study, such as a total RNA sample, that has been attached to a solid support. Information obtained from such experiments relates to the abundance of one single polynucleotide of interest. Array-based methods such as oligonucleotide arrays and cDNA arrays use the reverse strategy (Figure 1), where complex mixtures of labelled polynucleotides (such as cDNA derived from mRNA) are hybridised with large numbers of individual elements (e.g. unique PCR-products in cDNA microarrays), attached to a solid surface. In this way information on the abundance of many polynucleotide species is gained in parallel.

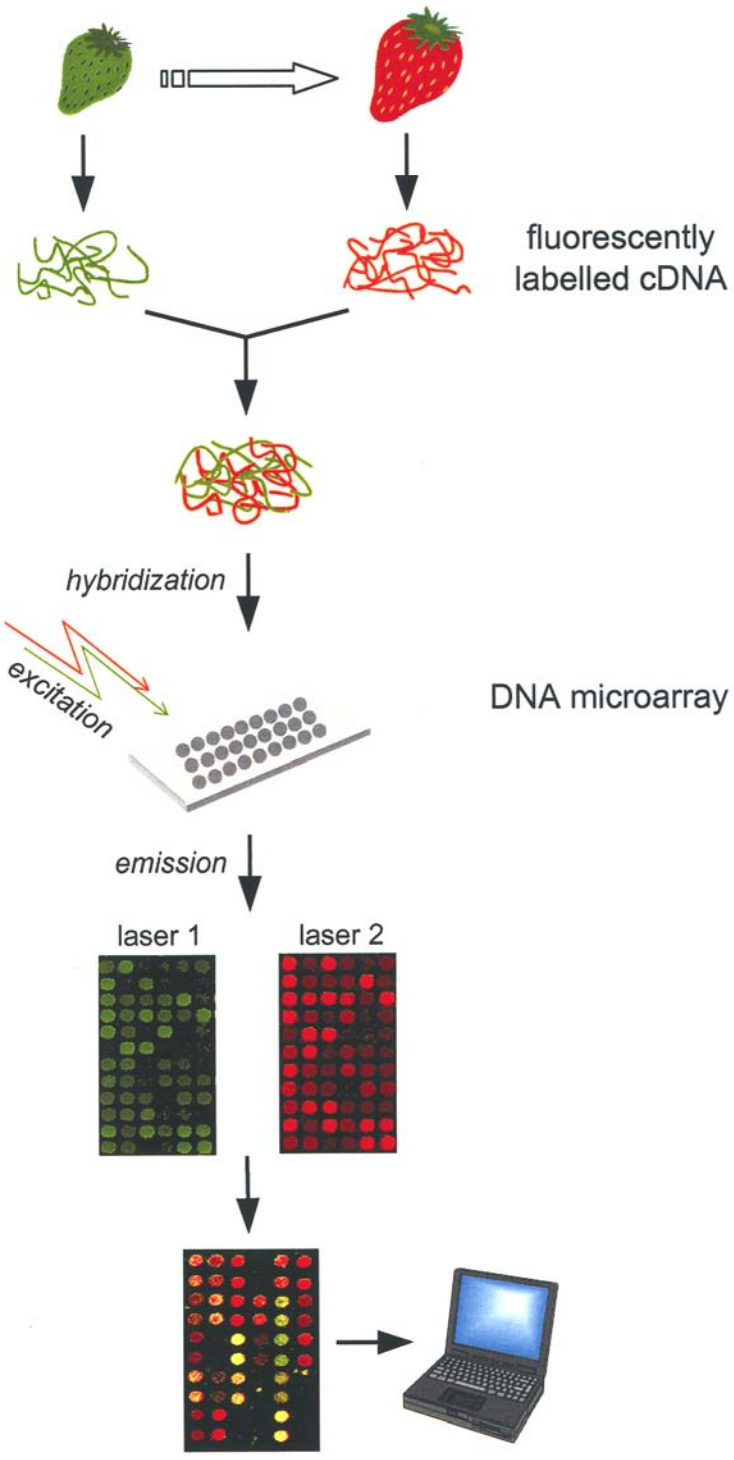

fluorescently
labelled cDNA
hybridization
excitation
DNA microarray
emission
laser 1
laser 2

Figure 1 (previous page) Scheme of a typical cDNA microarray assay for gene expression analysis. In this example mRNA levels are compared between the green and red stages of fruit development. First, mRNA is isolated from each tissue and reverse transcribed in the presence of different fluorescent dyes resulting in labelled cDNA. Next, the two cDNA populations are mixed and hybridised to a cDNA microarray. Each array element contains DNA representing a different gene. The specific cDNAs from both populations, representing individual transcripts, will hybridise specifically with the probe on the corresponding array element. After hybridisation, the microarray is scanned with a confocal laser devise for fluorescence emission at two wavelengths following independent excitation of the two dyes. The relative abundance of mRNA from each gene in green vs. red fruit is reflected by the ratio green / red as measured by the fluorescence emitted from the corresponding array element. Image analysis software is used to determine fluorescence intensities that allow the quantitative comparison between the two stages of fruit development for all genes on the array.

Labelling with fluorescent dyes possessing different excitation and emission characteristics allows the simultaneous hybridisation of two samples on a single array. The strength of fluorescence emission at the two wavelengths represents the amount of a specific poly-nucleotide from each sample bound to the array. In this manner a single experiment provides quantitative hybridisation data for hundreds to thousands of probes.

For expression studies using cDNA microarrays this approach of combining two differently labelled samples (reference and test sample) is common practice. For each gene the corresponding amount of signal in both samples can then be quantified in parallel and expression ratios obtained. This strategy, to use expression ratios instead of absolute expression levels, for the analysis of changes in gene expression, has been shown to be a very powerful one and has helped overcome a large source of experimental variation. Assuming the influence of the different dyes on the hybridisation characteristics of the labelled molecules to be identical, the initial ratios between specific, differently labelled mRNA molecules should be maintained upon hybridisation to the array. As a result, ratios between the two samples for each gene will then be independent of the amount of mRNA hybridised (Vorst *et al*., 2001).

Microarray Production

Two fundamentally different approaches are currently utilised in microarray fabrication. The printing-type technologies are based on the deposition of minute (sub-nanolitre) quantities of a DNA solution onto a solid surface (carrier). These fall into two distinct categories: contact printing (various methods for mechanical deposition) and non-contact printing (liquid delivery). Photolithographic techniques, on the other hand, can be used to synthesise oligonucleotides directly on the carrier.

Preparation of probes for cDNA microarrays

The first step of cDNA microarray fabrication requires the selection of probes to be used. For spotted arrays these can be PCR products resulting from the direct amplification of genomic DNA (by the use of gene specific primers) or amplified inserts from cDNA libraries (for instance ESTs) or any other library of interest. For example, the first microarray allowing genome-wide expression monitoring was generated by amplifying genomic DNA with specific primer pairs designed for 6200 open reading frames (ORFs) of yeast (DeRisi *et al*., 1997). Prior to spotting, probes are usually purified from unwanted PCR components and concentrated by precipitation or gel-filtration.

Instead of PCR products (ranging in size from approximately 0.2-2.5 kb), large synthetic oligonucleotides (50 to 80 base pairs) might also be spotted for the purpose of gene expression studies (Mir and Southern, 1999; Kane *et al*., 2000). The use of synthetic oligonucleotides may circumvent difficulties in distinguishing between expression of highly homologous transcripts (e.g. within gene families) or products of alternative splicing. Cross-hybridisation may cause a significant misinterpretation of microarray expression data and is of major concern since recent estimates from *Arabidopsis* predict 65% of the genes to occur in gene families (Arabidopsis Genome Initiative, 2000; Richmond and Somerville, 2000). Although, this approach is less time consuming than the use of PCR products, it does require prior knowledge of the sequence of the genes to be investigated.

As the production of whole-genome arrays is still very expensive and laborious, a subset of genes can alternatively be used to make a dedicated array. By selecting the appropriate tissue, developmental stage or treatment as source material, a cDNA library enriched for genes involved in the process under study can be obtained and used to pick clones. Another approach designed to enrich for targeted clones is to perform a pre-selection, identifying differentially expressed genes, such as suppression subtractive hybridisation (SSH; Diatchenko *et al*., 1996; Yang *et al*., 1999) or representational difference analysis (RDA; Welford *et al*., 1998).

Printing cDNA microarrays

In contact printing, an array of either solid or split pins are dipped into the DNA solution for loading. A micro-droplet is subsequently deposited upon direct contact with the solid surface of the array (Figure 2A and 2B). The method uses a motion control system that spots or prints a precise sample of each probe onto multiple surfaces (often 50 to 100 microscope slides) in a serial operation. Depending on the application, contact printing usually produces sub-nanolitre droplets at a pitch of 100 to 250 µm. An example of a split spotting pin is shown in Figure 2C. Mechanical spotting as described above is, at present, the most common way for the fabrication of cDNA microarrays.

Non-contact printing involves the controlled ejection of small (nano to picolitres) volumes of DNA solution from a dispenser onto the surface from a defined distance. In contrast to contact printing, this method allows flexibility in printing volume. The most common type of non-contact dispensing uses various types of ink-jet technology (e.g. thermal, solenoid, piezoelectric) for droplet generation and delivery (Okamoto *et al*., 2000). As for contact printing, both oligonucleotides and cDNAs can be handled. cDNA microarrays produced by ink jet printing may contain thousands of array elements.

cDNA microarrays are often fabricated on glass surfaces such as microscope slides. In order to enhance the adhesion of probes, to lower the background and to restrict spreading of the droplets, the slides are pre-coated with e.g. poly-lysine or amino silanes. After spotting, the DNA is immobilised (either by UV cross-linking or baking), the unused surface is blocked (by succinic anhydride or sodium borate) and, as a final step, the DNA on the slide is denatured (by heat or alkali treatment). Processed slides may be stored dry for several months prior to hybridisation.

Photolithographic microarray production

Synthesis of oligonucleotides on a surface using photolithography is used to fabricate high-density oligonucleotide microarrays. The most widely used method is the Affymetrix GeneChip technology. Photolithographic microarray fabrication involves DNA synthesis directly on the solid carrier surface using combinatorial chemistry methodology (Figure 3A). This solid surface, derivatized with chemical linkers containing photolabile protective groups, is activated selectively by shining light through a photo mask. Subsequently, the surface is flooded with a modified nucleotide to be coupled to the activated region of the chip. A repeated series of steps involving selective activation of specific regions and nucleotide coupling allows parallel oligonucleotide synthesis at many locations (Lemieux *et al*., 1998). Microarrays presently produced by Affymetrix normally contain 25-mer oligonucleotides within 20 to 24 μm feature size.

Typically, for expression monitoring, 16 oligonucleotide probe pairs, (i.e. 16 perfect match (PM) and 16 additional mismatch oligonucleotides (MM) for increased sensitivity and specificity of detection) are designed on non-conserved regions of a gene (http://www.affymetrix.com). The MM oligonucleotide is identical to the PM except for a single base difference at the central position. A single microarray for expression analysis containing more the 400,000 features will therefore represent approximately 13,000 genes.

Compared to the production of cDNA microarrays, the production of oligonucleotide arrays has the advantage that only sequence information and oligonucleotide design are pre-requisites, while e.g. handling of clones, primers, PCR products etc. is avoided. Oligonucleotide probes complementary to the known reference sequence (non-overlapping if possible) are usually selected to cover regions of 200 to 300 bases of the gene, cDNA or EST. Apart from the empirical

composition parameters used for their design, other criteria such as uniqueness compared to family members and other genes, should also be taken into account during probe selection.

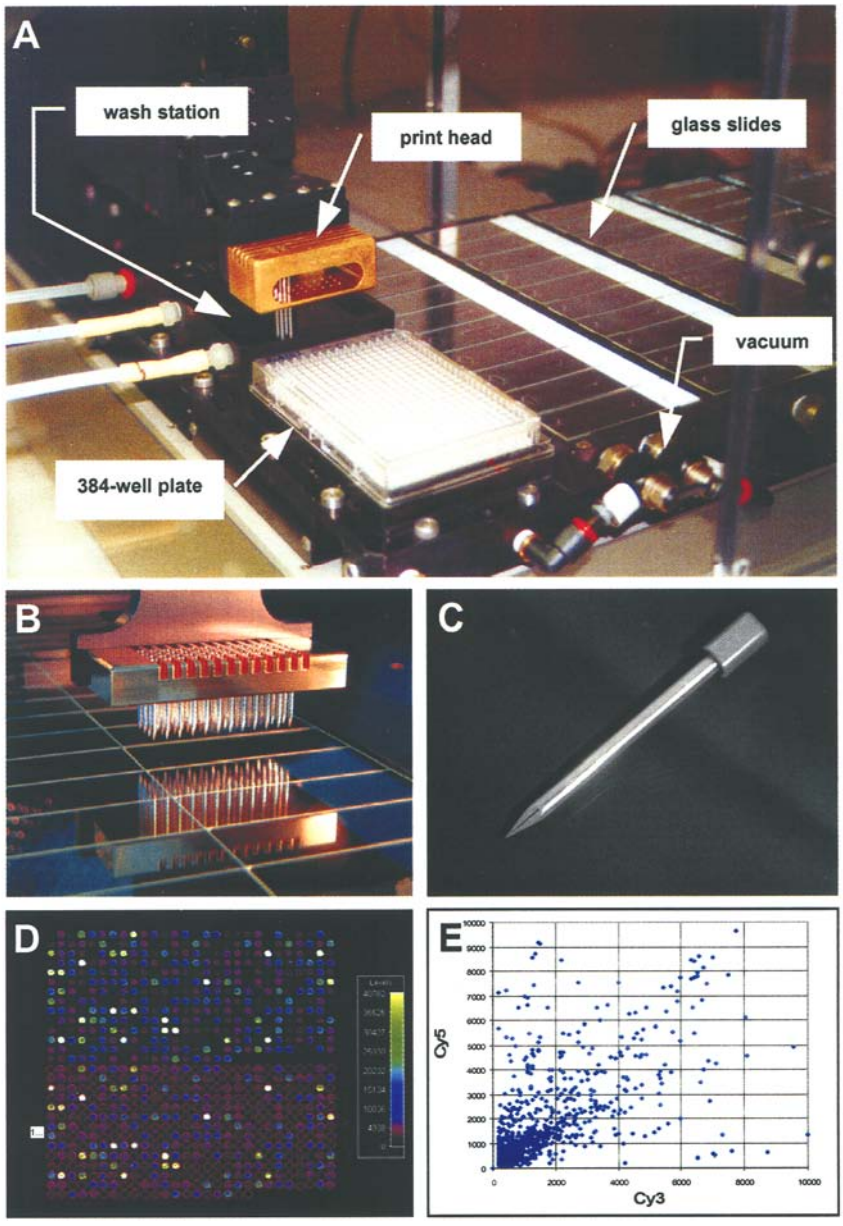


Figure 2 Fabrication of cDNA microarrays by contact printing. (A) Motion control system printing a sample of each probe onto multiple microscope slides in a serial operation [PixSys 7500 arrayer (Cartesian Technologies) provided with ChipMaker microspotting print head (TeleChem International, Inc)]. (B) Print head used for contact printing containing an array of 48 split pins (ChipMaker). (C) Detail of a split spotting pin. (D) Extracting the signal intensity of each individual array element. Image analysis software is used to superimpose a grid onto the image as shown in the lower half of the array. (E) Graphical representation of a typical raw data set obtained from a single microarray experiment. Each spot in the graph corresponds to an individual array element. Figures (B) and (C) are courtesy of T. Martinsky, TeleChem International, Inc.

Labelling and Hybridisation

In a cDNA microarray experiment, samples under study are typically RNA preparations from two or more biological sources. The fluorescent labelling of mRNA is commonly performed by first-strand cDNA synthesis in the presence of modified nucleotides using oligo-dT as a primer. Including fluorescently-labelled nucleotides during the reverse transcription reaction results in the direct synthesis of labelled cDNA. Alternatively, amine-modified nucleotides (e.g. 5-(3-aminoallyl)-2'-dUTP) can be incorporated that facilitate chemical linkage to the fluorescent dyes in a second reaction step involving monofunctional NHS-ester dyes**.**

The fluorescent labels, Cyanine-3 and Cyanine-5 are frequently paired, as they possess reasonably high incorporation efficiencies with reverse transcriptase, good photostability and yield, and absorb and emit light at distinct and separable wavelengths. Both total RNA or mRNA may be used as the starting material for labelling although the use of the latter provides the best hybridisation results. To obtain the desired fluorescent signal, 10 µg to 50 µg of total RNA or 0.5 µg to 2.5 µg of mRNA is used, per sample, per array. Improvements in labelling schemes such as by target amplification allows reduction of the amount of RNA required (0.1 µg total RNA) and the use of a minimum amount of tissue. This will facilitate studying gene expression samples derived from just a few cell layers (Hertzberg *et al*., 2001). A mixture of equal amounts of both labelled samples (between 5 µl to 50 µl total volume) are hybridised to the array usually under a coverslip. This is then placed in a specially designed reaction chamber to avoid evaporation. The hybridisation conditions, such as ionic strength, temperature and target concentration depend on the application. Hybridisation temperatures of 42 ºC (when hybridising in 50% formamide) to 70 ºC (when using SSC-based buffers) for several hours to overnight are typical.

When using photolithographic oligonucleotide arrays for gene expression studies, an alternative labelling procedure is used. Anti-sense copy RNA (cRNA) is made *in vitro* using T7-polymerase in the presence of biotinylated ribonucleotides (Bio-CTP or Bio-UTP). After sample amplification, it is fragmented to a length of 50 to 100 bases (Figure 3B). Hybridisation is thus RNA:DNA as opposed to DNA:DNA as in cDNA microarrays and only a single sample is usually hybridised per array. After hybridisation (30 min to 22 hours) the arrays are stained with streptavidin phycoerythrin conjugate and a confocal laser microscope is used for scanning.

Acquisition of Microarray Expression Data

Once the fluorescent sample is hybridised to a cDNA microarray, unbound material is washed away and the sample hybridised to each element is visualised by fluorescence detection. Both confocal scanning devices and CCD cameras are being used for this purpose. Fluorescence emission from the microarray is converted into a digital output for each dye, and is stored as separate image files. Next, image analysis software is used for quantification of individual array

elements. A grid is superimposed over the image and the average (or median) pixel intensities for each element is calculated for both dyes (Figure 2D). Background fluorescence is then subtracted from the raw data. Although fluorescent signals measured (directly) on areas between the array elements are often employed for background, it is more appropriate to use signals from foreign, non-plant array elements, e.g. yeast clones, that have been included on the array for this purpose (Vorst *et al*., 2001). As the mere presence of DNA on the array restricts the formation of background fluorescence, the use of non-elementary (between the spots) fluorescence will lead to an overestimation of the background signal.

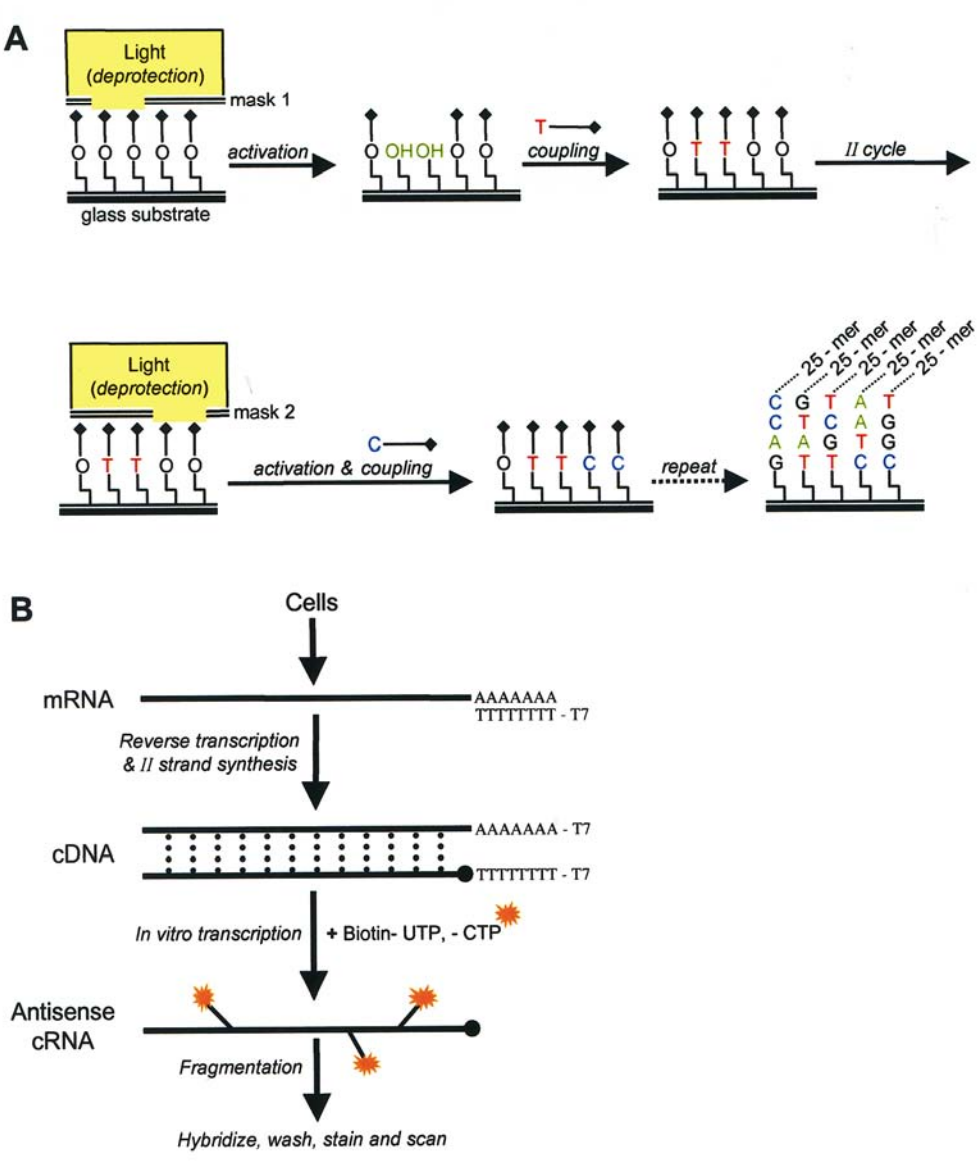


Figure 3 Photolithographic production of oligonucleotide microarrays and their application in expression studies. (A) Array Production. Glass substrate modified with a covalent linker molecule terminated with a photolabile protecting group (O ⟶) is selectively deprotected and activated (OH) for DNA synthesis by shining light through a photomask (mask 1). The surface is then flooded with protected nucleotides (A,T,C or G ⟶) which couple to the activated site. A similar process is repeated in a second cycle by using an additional mask (mask 2) which deprotects a new set of defined regions of the substrate. The process is repeated, activating different sets of sites by using different

masks and coupling different nucleotides. (B) Expression analysis. For oligonucleotide microarray hybridisation, labelled RNA samples are prepared by first converting extracted mRNA to double stranded cDNA. The cDNA is then copied to antisense RNA (cRNA) by an *in vitro* transcription reaction performed in the presence of biotin-labelled ribonucleotide triphosphates (UTP or CTP). Fragmented cRNA (50 to 100 nucleotides) is used for hybridisation. Following a brief washing step to remove non-hybridised cRNA, the microarrays are stained by streptavidin phycoerythrin and scanned.

Subsequently the figures are normalised to correct for channel specific effects such as differences in quantum yield of the dyes and unequal labelling efficiencies of the samples. Normalisation also corrects for any unwanted differences in the amount of sample used. Several ways of normalisation are being used: (i) Overall hybridisation signal; (ii) use of so-called housekeeping genes; and (iii) spiking with a foreign mRNA species that has been included on the array for this specific purpose.

For whole-genome arrays the use of the total fluorescence signal might be appropriate, as it corresponds to the total amount of RNA used to hybridise. However, when working with a subset of genes this approach is undesirable, especially when the genes included are 'related' to the process under study. When the represented genes are on average up- or down-regulated, this will be overlooked. Alternatively, when using a swapped dye approach, in which each hybridisation is repeated with both samples labelled reciprocally, an analysis of variance (ANOVA) can be applied on the log ratios. This allows the elimination of array effects, dye effects, sample effects, gene effects and gene x dye effects. Expression ratios are then produced as gene x sample effects (sample-specific effects for every gene) and will be centred on 1 (log ratios on 0). Again, when expression of the genes is biased toward one of the two samples, this bias will no longer be visible and valuable information will thus be lost.

The use of housekeeping genes, that are – by definition – equally expressed in each cell type under every condition, seems an attractive alternative. A set of these genes, included as probes on the array, could be used for channel normalisation. However, the correct means for selection of such genes remains elusive. The third strategy in which, before labelling, a known amount of a specific foreign mRNA is added to the sample is, in principal, the most elegant. On the other hand, this method might introduce a systematic error, when spiking is inaccurate. A graphical illustration of a typical data set obtained from a single microarray experiment is shown in Figure 2E.

In an oligonucleotide array experiment, following scanning of the arrays, intensities from each element are extracted and quantitative analysis of the hybridisation results is performed by analysing the hybridisation pattern of the set of PM and MM probes of every gene. In the presence of a specific RNA in the hybridisation solution the PM probes will hybridise more strongly on average than their MM partners (Figure 4). This assumption is used to determine the presence or absence of an RNA. To determine the quantitative RNA abundance, the average of the differences (PM minus MM) for each probe family is calculated. For evaluating gene expression between

multiple samples, the magnitude of the changes in the average of the differences (PM-MM) is directly compared after normalisation with known amount of spiked controls.

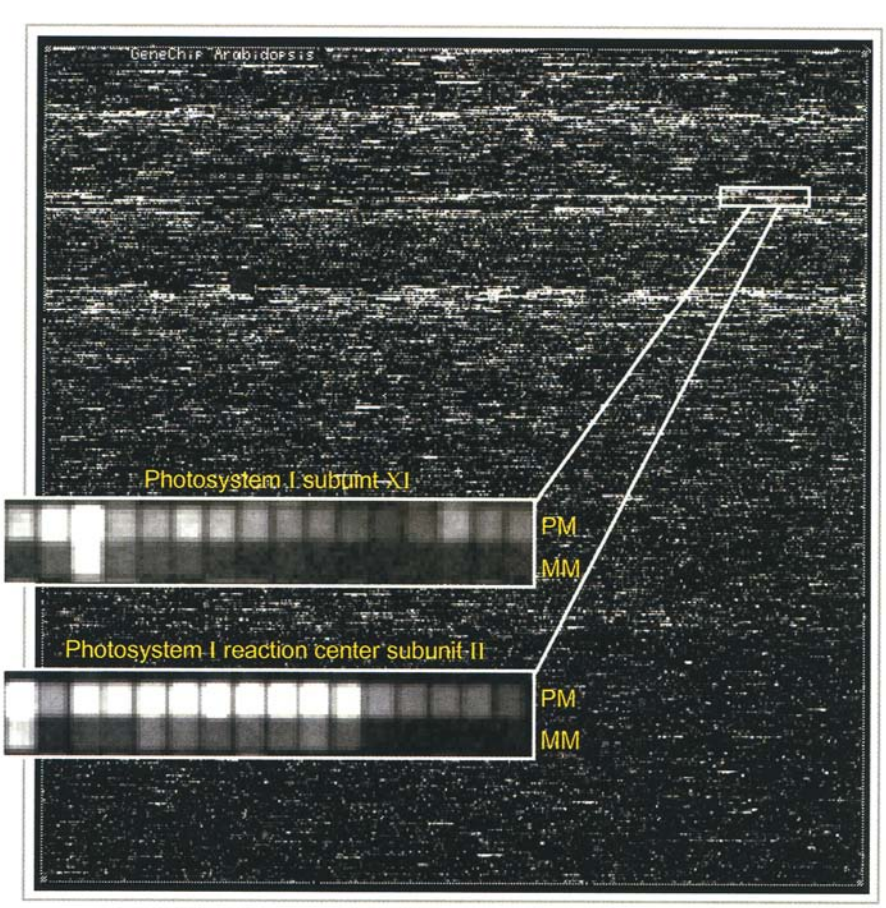


Figure 4 Fluorescent image of an oligonucleotide microarray (Affymetrix GeneChip) representing more than 8,000 *Arabidopsis* genes. The microarray was hybridised with RNA derived from four week old *Arabidopsis* leaf tissue. Every gene is represented by a set of 16 perfect match (PM) and mismatch (MM) probes facilitating quantitative interrogation. A small portion of the microarray containing PM and MM probe pairs for two photosynthesis related genes has been magnified. The MM probe of each pair serves as an internal control for hybridisation specificity. Image contributed by B. Han and T. Zhu (Torrey Mesa Research Institute, San Diego).

Data Exploration

The overwhelming amount of data generated by microarray experiments poses in itself problems to the average plant molecular biologist not yet acquainted with handling large data sets. These problems range from important but trivial subjects, such as clone tracking (associating the obtained hybridisation values with clone identities) and data storage, to more sophisticated topics such as the visualisation and analyses of multiple-sample experiments, often quickly accumulating tens of thousands of data points, and general data mining.

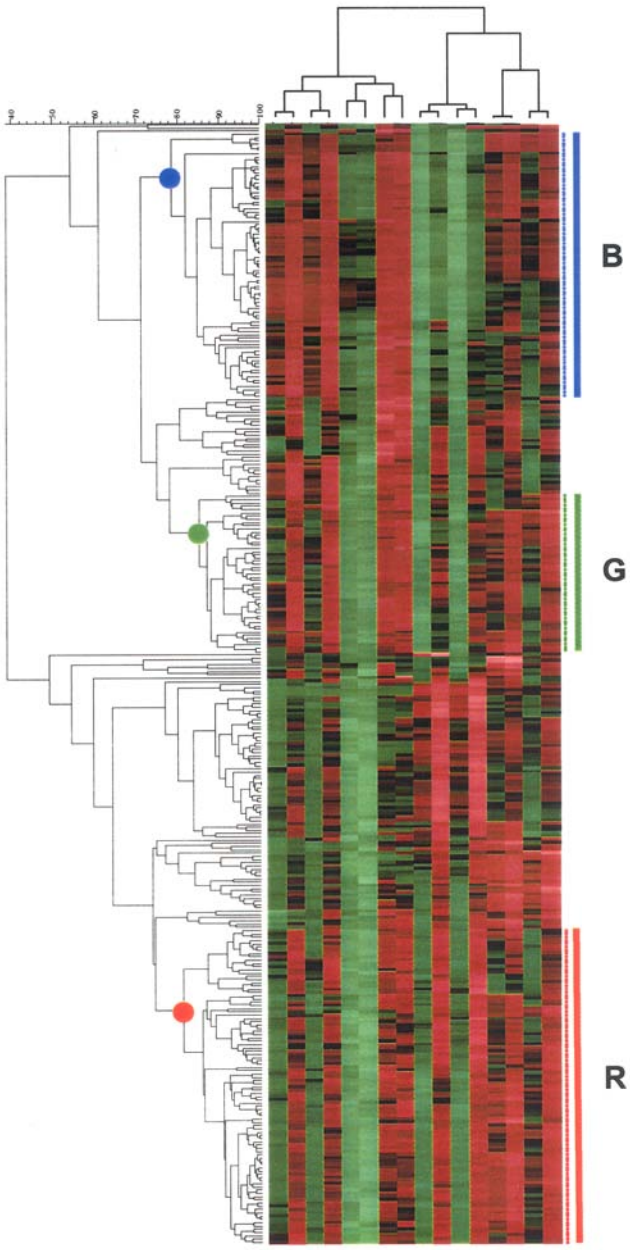


Figure 5 Hierarchical clustering of genes based on their expression profile. Each gene is represented by a row of coloured boxes, the columns correspond to the different experiments. Relative expression levels of each gene for each experiment is indicated on a colour scale ranging from green (low expression) to red (high expression). The tree to the left represents the clustering of the genes using the Pearson correlation as a distance measure, resulting in the genes showing a similarly 'shaped' expression profile clustering together. The lengths of the branches indicate the correlation with which genes were joined; long branches representing a low correlation. Three arbitrary gene clusters are indicated in blue (B), green (G) and red (R). The tree on top shows the experiments being clustered in a similar way. The closely linked pairs are the result from all experimental data being included twice using different normalisation strategies. The hierarchical clustering was performed using GenMaths (Applied Maths) software.

Several standard statistical techniques are currently being used to help interpret microarray data, including hierarchical clustering, principal component analysis (PCA) and self organising maps (SOM). These are all focused on grouping genes (or samples) together which show similar behaviour. This type of analysis, using large data sets, can provide novel perspectives on cellular regulatory mechanisms and can associate expression of unknown genes with a putative function (Figure 5). Hierarchical clustering of gene expression data in combination with false colour-coding of the expression levels has become a popular way of data analysis and presentation (Eisen *et al*., 1998). With this technique genes are grouped in clusters based on the similarity between their expression profiles. In a bottom up approach genes are joined to form nodes, which in turn are then further joined. Joining proceeds until all genes are combined in a single hierarchical tree.

Although nice figures seem to be guaranteed, one should be careful when drawing conclusions from them. As many similarity measures and clustering algorithms are available different outcomes are possible, depending on the method chosen. When using e.g. Euclidean distance or Pearson correlation as a similarity measure, different relationships between the studied genes are explored. In the latter case clusters are formed based on the 'shape' of the expression profiles and not on the absolute expression values. Also, the presence of genes showing no apparent behaviour might obscure the formation of discrete clusters; a problem that can be overcome by first filtering out the non-responsive genes.

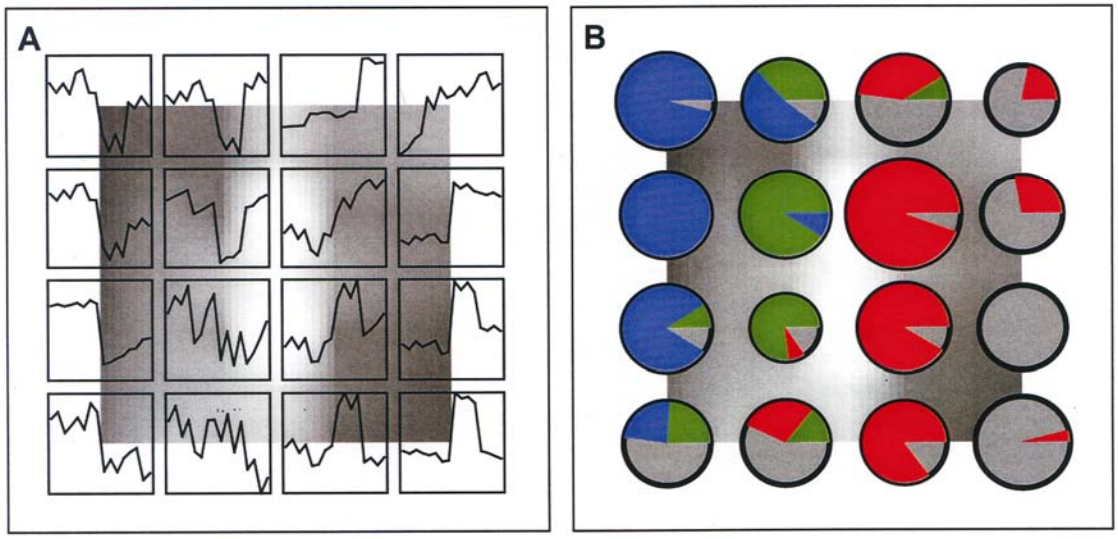


Figure 6 Visualisation of expression data using a self-organising map (SOM) of 4 x 4 partitions. (A) Representation of the reference expression profile for each partition as derived from the data set (identical to the one used for Figure 5). Typical for this type of analysis is the fact that partitions possessing a similar reference profile are neighbouring. (B) Composition of the partitions. The size of the pie diagrams correspond to the number of genes assigned to a partition. The presence of members of the three gene clusters defined in Figure 5 is indicated in colour (blue, green and red). There is a tendency of the members of each gene cluster to group together in the same or neighbouring partitions. Moreover, the green cluster occupies a intermediate position, as is the case in the PCA-analysis as well (Figure 7A). The gray shading in the back indicates the distances between the partitions in the SOM, darker shading corresponds with areas of high similarity. The SOM shown was prepared using GenMaths (Applied Maths) software.

A self-organising map has a pre-defined number of partitions in a two-dimensional grid. Via an iterative procedure the algorithm produces a reference expression profile for each partition in such a way that related profiles are neighbouring. In a final step each gene is assigned to a partition depending on which reference profile its expression most closely resembles. The advantages of this procedure are the absence of any hierarchy and also a clear delimitation of the clusters obtained. An apparent drawback is the need for the number of partitions to be defined beforehand (Figure 6).

Principal component analysis is a standard statistical method that helps to visualise multi-dimensional data sets in two- or three-dimensional space. When considering genes, the dimensions (experiments) are redefined as components in such a way that as much of the original variation as possible is retained. In this way a scatter plot of the genes in principal component space can be made with as little loss of information as possible (Figure 7). This is a valuable way of analysing data, especially in cases where biological meaning can be given to the principal component axes.

Current Applications of Microarrays

As in other biological disciplines, plant microarrays will prove an indispensable tool for research in the field of molecular plant science. The systematic, non-biased, accurate and large-scale acquisition of data using microarray technology enables new experimental approaches for plant molecular biologists. Microarray technology already provides, even at this early stage of its application in plant research, a global overview of biological mechanisms which until just recently were investigated in a "gene by gene" manner (Table 1). Microarrays are currently used for two main applications namely, gene expression studies and DNA variation analysis.

Expression profiling

At the moment gene expression monitoring is the most widespread application of microarrays. Microarray assays may be directly integrated into functional genomic approaches aimed both at assigning function to identified genes, and to studying the organisation and control of genetic pathways acting together to make up the functional organism. The rationale behind this approach is that genes showing similarity in expression pattern may be functionally related and under the same genetic control mechanism. Therefore, a common strategy undertaken already in early microarray studies was to analyse data by clustering genes into groups based on their expression profiles as scored in multiple experiments (Brown and Botstein, 1999). In most cases, gene clusters comprise both known and unknown genes allowing researchers to associate putative functions to the unknown genes by employing the concept of "guilt by association".

At present, both cDNA microarrays and oligonucleotide microarrays are used for gene expression monitoring. The first demonstration of the use of cDNA microarrays for quantitative monitoring of gene expression described the differential expression between *Arabidopsis* leaf and

root tissues using a small, 45-element array (Schena *et al*., 1995). Essentially, microarrays may be used to analyse any kind of variability in gene expression between given samples. These differences can be either naturally occurring or induced. Natural variation may occur between different plant cultivars, tissues, developmental stages, environmental conditions or during circadian rhythm. In a recent study, oligonucleotide-based arrays representing 8200 *Arabidopsis* genes were used to examine temporal patterns of gene expression in plant tissues harvested every 4 hours over a period of 2 days. Of the genes monitored, 6% were found to be under circadian control and "cycling" clusters were found in plant responses to light and key metabolic pathways (Harmer *et al*., 2000). More than 25% of the circadian regulated genes had not previously been characterised.

Induced changes in gene expression may arise from experimental exposure to different environmental conditions or result from mutagenesis. Schenk *et al*., (2000) used a 2,375- feature *Arabidopsis* cDNA array in order to perform comparative analyses of gene expression between plants inoculated with the fungal pathogen *Alternaria brassicicola* or treated with the defence-related signalling molecules, salicylic acid, methyl jasmonate and ethylene. A considerable network of regulatory interactions and co-ordination occurring during the plant defence response was identified, which had not been observed before by analysing only a few genes at a time. Microarray experiments, in which the response to drought and cold stresses (Seki *et al*., 2001), mechanical wounding and insect feeding (Reymond *et al*., 2000), herbivory (Arimura *et al*., 2000) and nitrate treatments (Wang *et al*., 2000) were analysed further, have already demonstrated the capability of microarray-assisted expression studies to identify novel response genes including those encoding regulatory factors.

In plants, microarrays are most suited for the analysis of mutant populations generated by a range of methods such as chemical or physical mutagenesis, knockouts (T-DNA tags, transposon tagging) and activation tagging (Pereira, 2000; Weigel *et al*., 2000). Other variants containing more specific mutations, induced by means of site-specific deletions, gene silencing, anti-sense, over-expression or the introduction of modified genes (induction of dominant mutations) may also be used. Maleck *et al*., (2000) used *Arabidopsis* cDNA microarrays to profile gene expression under 14 different conditions related to the capability of plants to develop a long lasting enhanced resistance to pathogens termed systemic acquired resistance (SAR). The 14 different conditions used were either generated by chemical treatment or involved mutants having a constitutive or a

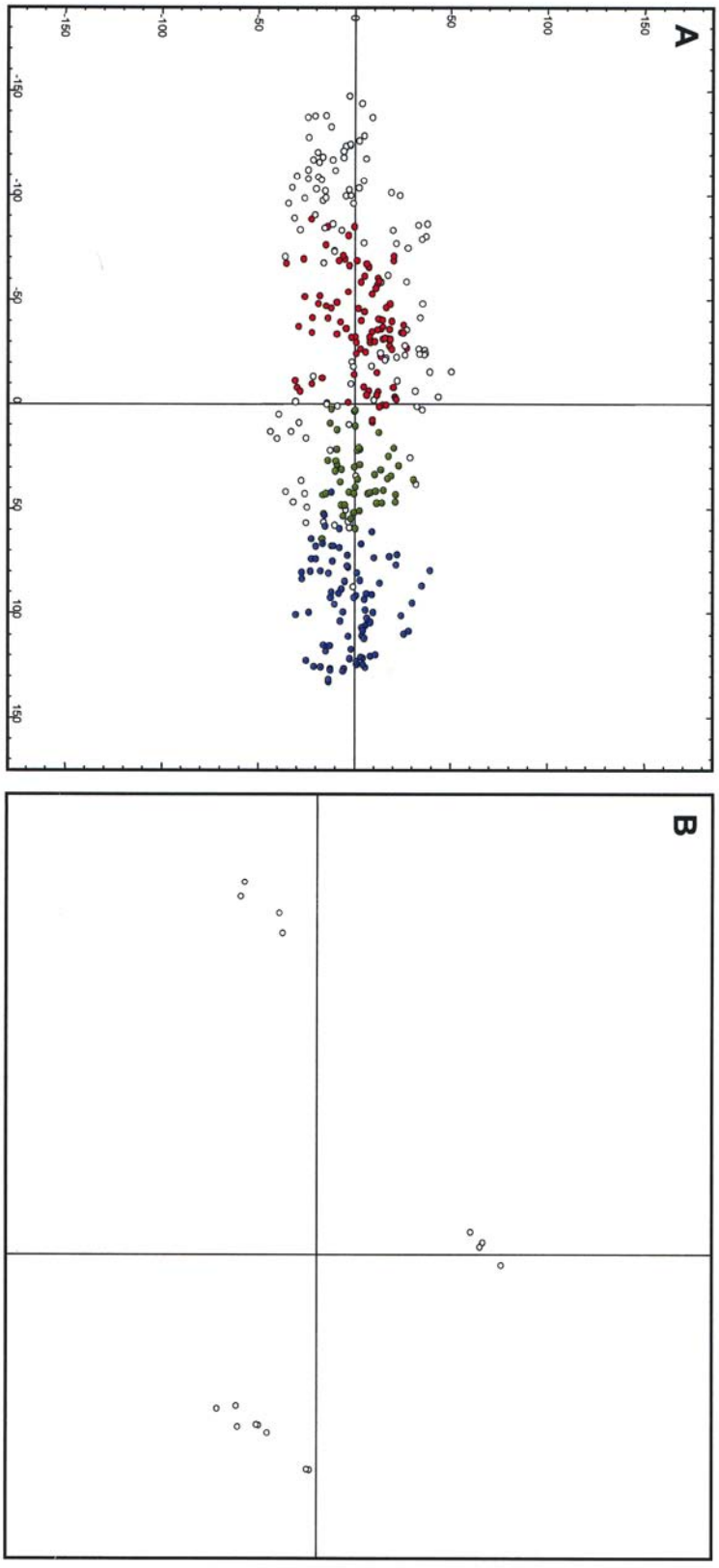

A
-150
-100
-50
0
50
100
150
B

Figure 7 (previous page) Principal component analysis (PCA) of expression data (identical to the one used for Figure 5). (A) Representation of the genes under study in principal component space. The first component, on the x-axis, and the second component, on the y-axis, explain 74.8% of the original variation. The colours indicate three gene clusters (B, G and R) based on the hierarchical clustering described in Figure 5; members of each of the gene clusters clearly group together. (B) Representation of the experiments in the same space, showing the contribution of each of them to the displayed components. This also gives an indication of their contribution to the separation of the three gene clusters in (A). Furthermore this clearly reveals that the different experiments are duplicates, repeats and alternative normalisations of only three biological samples, as they form three distinct groups. The PCA-analysis was done using GenMaths (Applied Maths) software.

repressed SAR phenotype. Apart from the identification of novel genes possibly functioning during SAR, by "scoring" for frequencies of potential binding motifs in promoters of co-regulated genes, the authors also identified the plant WRKY transcription factors as the most probable regulators of the SAR response. Similarly, monitoring expression in a plant containing a loss of function mutation or over-expression of a transcription factor or genes associated with signalling cascades may result in the identification of downstream genes (Petersen *et al*., 2000; Seki *et al*., 2001).

Genome-wide expression profiling

In expression monitoring, a single hybridisation experiment can provide quantitative results for 18,000 (Alizadeh *et al*., 2000) genes simultaneously. The high density and miniaturisation make genome-wide expression studies feasible, by using either cDNA or oligonucleotide arrays. Genome-wide expression profiling at the transcript level is one of the most exciting tools to study the cell and its integrative processes. Firstly, it is possible to measure transcript levels of every gene. This is something that is not yet feasible for proteins or metabolites. Secondly, expression patterns of genes can provide strong clues to elucidate their function. This assumption is based on numerous examples in which gene function was tightly connected to precise expression patterns under certain conditions. Consequently, global observation of gene expression patterns allows evaluation of the association between conditions of gene expression and function as well as the generality and strength of this link. Clearly, genome-wide expression data are linked to the study of promoters and regulatory elements which determine the levels of specific gene transcription. Understanding the information conveyed by the promoters will influence our ability to comprehend similarities in expression profiles. Last but not least, a broad picture of genes co-ordinately expressed in a cell might provide a dynamic molecular view and help to understand the operative biochemical and regulatory networks.

The ability to monitor simultaneously the expression of a large set of genes is one of the main spin-offs of genome sequencing efforts. Current reports on genome-wide expression analysis in plants also describe the use of microarrays (either oligonucleotide or cDNA) and already cover approximately one-third of the *Arabidopsis* genome (Wisman and Ohlrogge, 2000; Zhu and Wang,

2000). Nationally funded facilities for genomic surveys such as the Arabidopsis Functional Genomics Consortium (AFGC, http://afgc.stanford.edu/) in the USA and GARNET in the UK (http://www.york.ac.uk/res/garnet/garnet) are aimed at producing microarrays containing half to a complete *Arabidopsis* genome by the year 2003. *Arabidopsis* and possibly rice microarrays representing entire genomes will soon also be commercially available or provided as a service to the scientific community (Wisman and Ohlrogge, 2000; Zhu and Wang, 2000). During the coming years results from hundreds of microarray experiments will be collected in special gene expression databases [e.g. Stanford Microarray Database (SMD), http://genome-www4.stanford.edu/MicroArray/SMD/, NCGI's Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo/, NCGR's GeneX, http://www.ncgr.org/research/genex/]. This wealth of information might be most valuable in combination with additional non-related experiments, as shown by Schaffer *et al*., (2001). Conservation in sequence may also make existing microarray data applicable to related plant species (e.g. *Arabidopsis* and *Brassica napus*) (Girke *et al*., 2000).

The flexible nature of the fabrication and hybridisation methods of cDNA microarrays allows the application of the technology to non-model organisms. An early example of the application of cDNA microarrays to a non-model plant described the use of strawberry microarrays containing 1701 cDNAs for analysing gene expression during fruit development (Aharoni *et al*., 2000). A significant product of these experiments was the identification and characterisation of a novel gene involved in fruit flavour production. This clearly demonstrated the capability of microarray expression profiling to link gene to function, particularly when an exceptionally complicated and poorly described biological process is of interest. Results from microarrays representing genes derived from a range of sources, from lower plants to trees will no doubt be reported in the near future. As a consequence, microarray technology will effectively narrow the gap in molecular biology between model species and less exploited plant species.

Novel applications

Data from genome-wide gene expression studies provide a novel approach to the identification of new cis-regulatory elements in promoter regions and to the classification of genes in similar regulatory circuits according to the elements identified (Bucher, 1999). For example, the discovery of an "evening element motif" conferring rhythmic gene expression in *Arabidopsis* was identified solely by computational means (Harmer *et al*., 2000).

Another interesting application of gene expression arrays for the identification of secreted and membrane-associated gene products was described in yeast and humans by Diehn *et al*., (2000). The authors took advantage of the fact that membrane and secreted proteins are translated by membrane-bound polysomes which can be physically separated from free polysomes (synthesising cytoplasmic proteins). By comparing mRNA samples derived from membrane bound polysomes

and free polysomes they could identify hundreds of transcripts putatively encoding secreted or membrane associated proteins. Some of these were already identified, while others had not previously been recognised as such.

Table 1. Reports on the use of DNA microarrays in plants.

Biological Context	Plant species	Microarray Type and Scale[a]	Reference
Expression in roots and leaves		cDNA; 48 clones	Schena *et al.,* 1995
Expression in major plant organs	*Arabidopsis*	cDNA; 1443 clones	Ruan *et al.,* 1998
Strawberry ripening and flavour, flower development	Strawberry, Petunia	cDNA; 1701 strawberry and 480 petunia clones	Lemieux *et al.,* 1998; Aharoni *et al.,* 2000
Expression in rosette leaves of two accessions	*Arabidopsis*	cDNA; 673 clones	Kehoe *et al.,* 1999
Mapping the trait for defense response to fungal pathogen	*Arabidopsis*	Oligo; 412 polymorphisms	Cho *et al.,* 1999
Response to mechanical wounding and insect feeding	*Arabidopsis*	cDNA; 150 clones	Reymond *et al.,* 2000
Response to nitrate treatments	*Arabidopsis*	cDNA; 5524 clones	Wang *et al.,* 2000
Response to treatments with defense-related signaling molecules & fungal pathogen	*Arabidopsis*	cDNA; 2375 clones	Schenk *et al.,* 2000
Expression regulated by the circadian clock	*Arabidopsis*	Oligo; 8200 genes represented	Harmer *et al.,* 2000
Expression associated with systematic acquired resistance (SAR)	*Arabidopsis*	cDNA; 10,000 clones	Maleck *et al.,* 2000
Phytochrome A mediated response	*Arabidopsis*	Oligo; 412 polymorphisms	Spiegelman *et al.,* 2000
Expression in developing seeds	*Arabidopsis*	cDNA; 2715 clones	Girke *et al.,* 2000
Expression analysis of the glutathione-*S*-transferase gene family	Maize	cDNA; 42 clones	McGonigle *et al.,* 2000
Identification of downstream genes in MAP kinase 4 signaling pathway	Arabidopsis	cDNA; 9861 clones	Petersen *et al.,* 2000
Response to herbivory and herbivore-induced volatiles	Lima bean	cDNA; 2032 clones	Arimura *et al.,* 2000
Expression in different tissues, organs, genetic conditions & growth environments	*Arabidopsis*	Oligo; 8200 genes represented	Zhu and Wang, 2000
Large-set of biological questions, e.g. amino acid-metabolism, apoptosis, development, environmental-conditions, hormone treatment, metals, mitochondria, mutants, pathogen, stress, RNA stability, comparative-genomics, virus.	*Arabidopsis*	cDNA; 11,521 clones	Schaffer *et al.,* 2000; Wisman and Ohlrogge, 2000 (AFGC)[b]
Expression in different tissues	Rice	cDNA; 1265 clones	Yazaki *et al.,* 2000
Diurnal and circadian-regulated genes	*Arabidopsis*	cDNA; 11,521 clones	Schaffer *et al.,* 2001
Expression under drought and cold stresses	*Arabidopsis*	cDNA; 1300 clones	Seki *et al.,* 2001
Identification of repetitive genomic elements in 17 *Vicia* species; phylogenetic reconstruction	Vicia *spp*.	Repetitive genomic fragments; 1152 clones	Nouzová *et al*., 2001
Response to high light	Synechocystis *sp PCC6803*	1.0 kb PCR fragments; 3079 clones	Hihara *et al*., 2001
Expression of cold-regulated genes in hik33 mutant	Synechocystis *sp PCC6803*	1.0 kb PCR fragments; 3079 clones	Suzuki *et al*., 2001
Salt stress induced gene expression	Rice	CDNA; 1728 clones	Kawasaki *et al*., 2001

[a]cDNA, cDNA microarray; oligo, oligonucleotide array.

[b]AFGC, Arabidopsis Functional Genomics Consortium (http://afgc.stanford.edu); results obtained both from experiments performed as a service to the public and from surveys of the AFGC team itself.

Two recent publications (Ren *et al*., 2000; Iyer *et al*., 2001) demonstrated the use of DNA microarrays in combination with chromatin immunoprecipitation methods, for the identification of *in vivo* binding sites for yeast transcription factors. Ren *et al.,* could identify novel target genes of *GAL4*, which is one of the most investigated yeast transcription factors. In both cases, microarrays

composed specifically of all yeast intergenic regions were used thus limiting the use of the method to organisms with sequenced genomes. In addition such an approach would be difficult to apply in situations such as in higher plants where the genome is larger, the intergenic regions are more extensive and the promoter regions are difficult to define.

DNA sequence variation

Oligonucleotide arrays are well suited for the detailed analysis of DNA variation as they allow the detection of single nucleotide mismatches during hybridisation. These analyses can include both the discovery of novel DNA variants and the determination of known variants, e.g. in large-scale genotyping. Sequence variations, such as single nucleotide polymorphisms (SNPs), can serve as genetic markers. Several different oligonucleotide array designs, which are composed of probes complementary to sub-sequences of a target, can be used to determine the identity and abundance of the target sequence.

With a 'tiling' oligonucleotide array it is possible to scan a target sequence for mutations (Figure 8A). In such an array each overlapping 25-mer in the sequence is covered by four complementary oligonucleotide probes that differ only by having A, T, C or G substituted at the central position. In this way, each nucleotide to be determined is represented by four oligonucleotides. This renders it a most accurate and sensitive way of base-by-base DNA re-sequencing. The entire, 16.6 kb, human mitochondrial DNA was resequenced by performing in a single hybridisation to a 64,000 probe array (Chee *et al*., 1996). Resequencing of much larger DNA portions such as > 2 Mb of human genomic DNA composed of 16,725 short genomic sequences (or sequence –tagged sites; STSs) has also been demonstrated (Wang *et al*., 1998). In this study a total of 3241 candidate SNPs were identified.

When oligonucleotide arrays are used to detect known polymorphisms, such as SNPs, instead of a tiling array, another design has been applied, combining two tilling arrays, termed variant detector arrays (VDAs). In such a genotyping array, each SNP is tested by two VDAs corresponding to the two alternative alleles (Figure 8B). Each VDA will interrogate a few (e.g. seven) nucleotides centred around the polymorphic site by a set of four oligonucleotides for each of the bases investigated. Fragments corresponding to each SNP-containing region are amplified in a multiplex PCR reaction. The PCR products are then labelled and hybridised to the array. In this way a few hundred polymorphisms could be investigated in parallel. Typically, the genotyping array will include VDAs for the complementary strand as well. The technique also allows the reliable detection of heterozygous (Figure 8B).

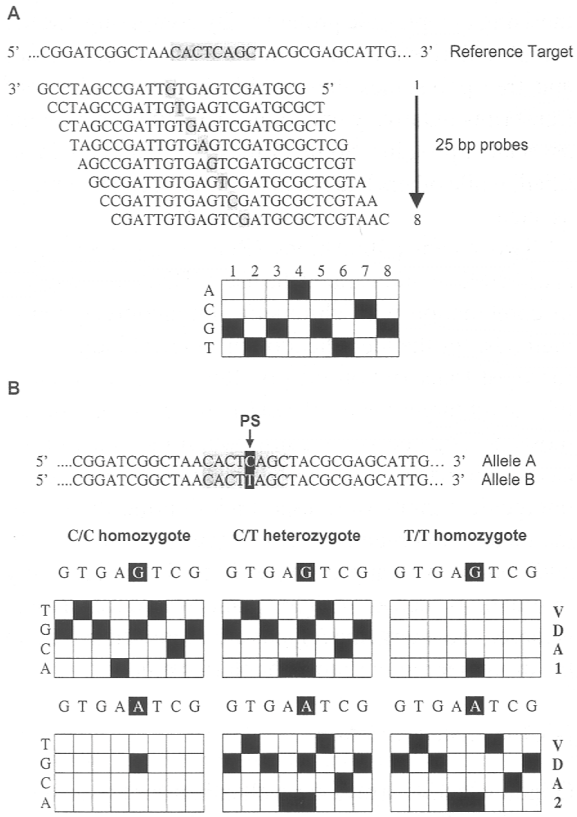


Figure 8 (A) Tiling strategy for the construction of an oligonucleotide array to be used for DNA resequencing. Each position in a reference target (nucleotides with gray background) is queried by a group of four complementary probes on the array. Each of the four probes is identical except for a single nucleotide at the central position, which is either a A, T, C or G. Subsequent groups (indicated 1 through 8) query adjacent positions in the reference target, indicated with gray background. Hybridization of a fluorescent sample to the array results in the hybridization pattern shown (black boxes). This approach allows the base by base resequencing of target DNA. (B) Oligonucleotide microarray for genotyping. The polymorphic site (PS) and several nucleotides centered around it (depicted with gray background) are interrogated by two variant detector arrays (VDAs) corresponding to the two alternative alleles (A and B). Each VDA consists of a set of four oligonucleotides, for each bases investigated, containing either an A, T, C, or G. Fragments corresponding to each SNP-containing region are amplified in a multiplex PCR reaction. The PCR products are then labeled and hybridized to the array. The resulting hybridization pattern for samples with the respective genotypes CC, CT, and TT is shown (hybridization is indicated in black). The presence of an allele should be reflected in strong hybridization to the corresponding VDA. In this way a few hundred polymorphisms could be investigated in parallel. Typically, the genotyping array will include VDAs for the complementary strand as well.

Alternatively, SNPs can be detected on spotted oligonucleotide arrays using a method called mini-sequencing, or single base extension (SBE) (Syvanen, 1999). In the mini-sequencing procedure an immobilised 'probe' oligonucleotide is extended after hybridisation to the target that functions as template. DNA polymerase is applied to incorporate a single appropriately labelled dideoxyribonucleoside triphosphate that matches the nucleotide at the variable site of the target. This single base extension reaction can be performed directly on the array and the differentially labelled nucleotides incorporated may be scored and used for the detection of point mutations and

SNPs. Recent improvements to the SBE method using spotted arrays such as SBE-TAGS (Hirschhorn *et al*., 2000) and allele specific primer extension (Pastinen *et al*., 2000) or those synthesised by photolithography (Fan *et al*., 2000) allow high-throughput genotyping of SNPs in a large set of samples.

For plants, reports on the use of microarray technology to detect polymorphisms are still scarce. Two reports on this topic were published recently for *Arabidopsis*. Cho *et al.*, (1999) describe the mapping of the trait for defence against the fungal pathogen *Erysiphe orontii* by high-throughput generation of meiotic maps of F_2 individuals using oligonucleotide array-based genotyping. In the second report Spiegelman *et al*., (2000) used hybridisation to the same oligonucleotide array (used by the previous authors) for rough initial mapping in combination with a denaturing HPLC technique for fine mapping of the RSF1 locus, defined by a mutant with reduced sensitivity to far-red light. The gene cloned using this strategy encodes a basic helix-loop-helix transcription factor which mediates phytochrome A signalling.

The Future of Plant Microarrays

Efficient use of microarray technology will eventually rely on the interconnection of diverse and accurate transcriptional profiles to data produced by other functional genomic tools. Generating and accumulating thousands of transcription profiles from a vast array of tissues, developmental stages, treatments etc. will make available a fingerprint of a large set of possible transcriptional scenarios in the cell. Transcription profiles gathered together into a reference database or "compendium" will allow the matching of expression patterns of uncharacterised mutants with known profiles in the database. Thus, one microarray assay may be sufficient to associate a mutation in a gene with a change in phenotype.

In plants, mutation machines, such as insertional mutagenesis using either transposons or T-DNA tags, are required for constructing a powerful microarray-based expression "compendium". In *Arabidopsis*, it is already feasible to acquire a mutant of every second open reading frame in the genome by using publicly available populations of insertional mutagenesis lines (Parinov and Sundaresan, 2000). However, even total genome saturation will not be sufficient to detect changes in gene expression for all transcripts, as genes that are expressed constitutively at low levels, or need specific factors for activation, may be overlooked. It will therefore be necessary to develop a panel of conditions, including crossings between mutant lines, that will cause a significant "transcriptional phenotype" for each mutant analysed. The mutated lines will be all screened for alterations in morphologic or metabolic characteristics. The results of such a "compendium" approach would be a standardised set of mutants and phenotypes, which have been profiled using microarrays.

The utility of the "compendium" approach has recently been validated in *S. cerevisiae* by examining 300 genome-wide profiles caused by deletions in both characterised and

uncharacterised genes, as well as treatments with compounds with known molecular targets (Hughes *et al*., 2000). Such experiments resulted in function identification for eight proteins encoded by previously uncharacterised open reading frames and also the discovery of a novel target for a commonly-used drug. Such a "compendium" could also result in a comprehensive identification of co-regulated transcript groups, which may lead to function identification of genes based on their regulatory characteristics.

Although applications of gene expression arrays are extensive, as described above, one should realise that it is not simply mRNA levels, but also the amount and modification of expressed proteins within the particular cellular context which determines true gene activity. It is therefore most important to couple transcriptome data to other functional "maps" such as those derived from DNA, protein and metabolite analyses. Protein expression data obtained either by 2D gel analysis coupled to mass-spectrometry or other more sensitive methods may provide clues to the mode of regulation when coupled to gene expression data (Dutt and Lee, 2000). High throughput protein interaction assays such as those performed for all yeast open reading frames (Uetz *et al*., 2000) will link protein partners to microarray gene expression clusters.

At the metabolite level, two main approaches, namely, metabolic profiling and biochemical genomics, may prove to integrate well with microarray data. The first approach allows us to obtain "snap-shots" of low molecular weight metabolites produced in different plant genotypes, genetically modified plants and plants exposed to different environmental conditions (Roessner *et al*., 2000). Currently, levels of only a few hundred metabolites out of the few hundred thousand predicted to accumulate in the plant cell can be determined (Fiehn *et al*., 2001). Although metabolites are, generally speaking, the ultimate products of genes, complex connections between metabolite levels and RNA expression can be expected. In the case of a positive correlation, detecting metabolic changes might lead to the identification of unknown genes through being part of a cluster induced or repressed under the same biological context. Negative correlation will teach us about the point of regulation of different metabolic pathways. Combined investigations of gene expression and metabolite levels will also aid in deciding on strategies for metabolic engineering.

A recently-described approach termed "biochemical genomics" allows high throughput identification of genes encoding proteins capable of performing a specific biochemical activity (Martzen *et al*., 1999). The method uses ORFs fused to a tag in order to systematically express, purify and assay enzymatic activity of individual proteins. Apart from identifying new enzymatic activities the purified proteins themselves can be analysed as putative substrates, e.g. for phosphorylation by regulatory protein kinases. Data arising from this approach will provide insight into the possible links between metabolic pathways, enzymes and gene expression clusters.

Conclusions

A large proportion of discoveries made in all scientific disciplines can be attributed to advances in tools and methodologies. DNA microarray technology will be one of the technologies which will take us rapidly forward in our understanding of plant biology in the years to come. At present the technology is widely used for monitoring gene expression on a large scale. The power of the microarray approach is in the possibility to monitor RNA levels for the complete set of transcripts of an organism. In the coming years the focus for expression analysis should be on constructing an accurate, detailed and large reference data base of known expression profiles. This will allow the association of a function to a gene by "searching" the data base for a matching expression phenotype. Such a procedure, in combination with information derived from other functional genomic fields, will enable a rapid method for the identification of gene function to become established.

In the field of DNA variation analysis, microarray-based analysis methods enables the determination of alleles at hundreds of thousands of loci, from numerous samples. This will facilitate a closer understanding of the genetic contribution to complex plant traits. As has been shown in the medical field, the technology is expected to expand outside the research laboratories to industrial applications and services. Although little has been discussed on this aspect of microarrays in this review, the method has much to offer in a commercial context e.g. in the diagnostics of food quality and safety, and accelerating breeding programs. The number of reports on the use of microarray methodology has increased steadily in the last year and it would not be surprising to observe a similar future trend in exciting biological discoveries in plant science.

Acknowledgements

The authors would like to thank Robert Hall and Ruud de Maagd for critical reading of the manuscript.

References

Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J., Sun, Z.K., Alvarez Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A., De Vos, R.C.H., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J. and O`Connell A.P. 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12: 647-661.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.G., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L.M., Marti, G.E., Moore, T., Hudson, J., Lu, L.S., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O. and Staudt, L.M. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403: 503-511.

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815.

Arimura, G., Tashiro, K., Kuhara, S., Nishioka, T., Ozawa, R. and Takabayashi, J. 2000. Gene responses in bean leaves induced by herbivory and by herbivore-induced volatiles. Biochem. Biophys. Res. Commun. 277: 305-310.

Bachem, C.W.B., Van Der Hoeven, R.S., De Bruijn, S.M., Vreugdenhil, D., Zabeau, M. and Visser, R.G.F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. Plant J. 9: 745-753.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S.J., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J. and Corcoran, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature Biotech. 18: 630-634.

Breyne, P. and Zabeau, M. 2001. Genome-wide expression analysis of plant cell cycle modulated genes. Curr. Opin. Plant Biol. 4: 136-142.

Brown, P.O. and Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. Nature Genet. 21 Suppl.: 33-37.

Bucher, P. 1999. Regulatory elements and expression profiles. Curr. Opin. Struct. Biol. 9: 400-407.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P. 1996. Accessing genetic information with high-density DNA arrays. Science 274: 610-614.

Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., Theologis, A., Yang Wei, H., Hubbell, E., Au, M., Chung, E.Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R.J., Ausubel, F.M., Davis, R.W. and Oefner, P.J. 1999. Genome-wide mapping with biallelic markers in Arabidopsis thaliana. Nature Genet. 23: 203-207.

DeRisi, J.L., Iyer, V.R. and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278: 680-686.

Desprez, T., Amselem, J., Caboche, M. and Hofte, H. 1998. Differential gene expression in Arabidopsis monitored using cDNA arrays. Plant J. 14: 643-652.

Diatchenko, L., Lau, Y.F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. 1996. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl Acad. Sci. USA 93: 6025-6030.

Diehn, M., Eisen, M.B., Botstein, D. and Brown, P.O. 2000. Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. Nature Genet. 25: 58-62.

Duggan, D.J., Bittner, M., Chen, Y.D., Meltzer, P. and Trent, J.M. 1999. Expression profiling using cDNA microarrays. Nature Genet. 21 Suppl.: 10-14.

Dutt, M.J. and Lee, K.H. 2000. Proteomic analysis. Curr. Opin. Biotechnol. 11: 176-179.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl Acad. Sci. USA 96: 10943.

Fan, J.B., Chen, X., Halushka, M.K., Berno, A., Huang, X., Ryder, T., Lipshutz, R.J., Lockhart, D.J. and Chakravarti, A. 2000. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. Genome Res. 10: 853-860.

Fiehn, O., Kloska, S. and Altmann, T. 2001. Integrated studies on plant biology using multiparallel techniques. Curr. Opin. Biotechnol. 12: 82-86.

Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T. and Solas, D. 1991. Light-directed, spatially addressable parallel chemical synthesis. Science 251: 767-773.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. 2000. Microarray analysis of developing Arabidopsis seeds. Plant Physiol. 124: 1570-1581.

Harmer, S.L., Hogenesch, L.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A. and Kay, S.A. 2000. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science 290: 2110-2113.

Hertzberg, M., Sievertzon, M., Aspeborg, H., Nilsson, P., Sandberg, G. and Lundeberg, J. 2001. cDNA microarray analysis of small plant tissue samples using a cDNA tag target amlification protocol. Plant J. *In press*.

Hihara, Y., Kamei, A., Kanehisa, M., Kaplan, A. and Ikeuchi, M. 2001. Dna microarray analysis of cyanobacterial gene expression during acclimation to high light. Plant Cell 13: 793-806.

Hirschhorn, J.N., Sklar, P., Lindblad-Toh, K., Lim, Y.M., Ruiz-Gutierrez, M., Bolk, S., Langhorst, B., Schaffner, S., Winchester, E. and Lander, E.S. 2000. SBE-TAGS: An array-based method for efficient single-nucleotide polymorphism genotyping. Proc. Natl Acad. Sci. USA 97: 12164-12169.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M. and Friend, S.H. 2000. Functional discovery via a compendium of expression profiles. Cell 102: 109-126.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Synder, M. and Brown, O.P. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409: 533-538.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. and Kim, S.K. 2001. Genome-wide analysis of developmental and sex-regulated gene expression profiles in Caenorhabditis elegans. Proc. Natl Acad. Sci. USA 98: 218-223.

Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D. and Madore, S.J. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucl. Acids Res. 28: 4552-4557.

Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D. and Bohnert, H.J. 2001. Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13: 889-906.

Kehoe, D.M., Villand, P. and Somerville, S.C. 1999. DNA microarrays for studies of higher plants and other photosynthetic organisms. Trends Plant Sci. 4: 38-41.

Lander, E.S. 1999. Array of hope. Nature Genet. 21 Suppl.: 3-4.

Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. and Davis, R.W. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl Acad. Sci. USA 94: 13057-13062.

Lemieux, B., Aharoni, A. and Schena, M. 1998. Overview of DNA chip technology. Mol. Breeding 4: 277-289.

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. Nature Genet. 21 Suppl.: 20-24.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E.L. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotech. 14: 1675-1680.

Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., Lawton, K.A., Dangl, J.L. and Dietrich, R.A. 2000. The transcriptome of Arabidopsis thaliana during systemic acquired resistance. Nature Genet. 26: 403-410.

Martzen, M.R., McCraith, S.M., Spinelli, S.L., Torres, FM., Fields, S., Grayhack, E.J. and Phizicky, E.M. 1999. A biochemical genomics approach for identifying genes by the activity of their products. Science 286: 1153-1155.

McGonigle, B., Keeler, S.J., Lan, S.M. C., Koeppe, M.K. and O`Keefe, D.P. 2000. A genomics approach to the comprehensive analysis of the glutathione S-transferase gene family in soybean and maize. Plant Physiol. 124: 1105-1120.

Mir, K.U. and Southern, E.M. 1999. Determining the influence of structure on hybridization using oligonucleotide arrays. Nature Biotech. 17: 788-792.

Nouzová, M., Neumann, P., Navrátilová, A., Galbraith, D.W. & Macas, J. 2001. Microarray-based survey of repetitive genomic sequences in Vicia spp. Plant Mol Biol. 45: 229-244.

Okamoto, T., Suzuki, T. and Yamamoto, N. 2000. Microarray fabrication with covalent attachment of DNA using Bubble Jet technology. Nature Biotech. 18: 438-441.

Parinov, S. and Sundaresan, V. 2000. Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. Cur. Opin. Biotechnol. 11: 157-161.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L. and Syvanen A.C. 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res. 10: 1031-1042.

Pereira, A. 2000. A transgenic perspective on plant functional genomics. Transgenic Res. 9: 245-260.

Petersen, M., Brodersen, P., Naested, H., Andreasson, E., Lindhart, U., Johansen, B., Nielsen, H.B., Lacy, M., Austin, M.J., Parker, J.E., Sharma, S.B., Klessig, D.F., Martienssen, R., Mattsson, O., Jensen, A.B. and Mundy, J. 2000. Arabidopsis MAP kinase 4 negatively regulates systemic acquired resistance. Cell 103: 1111-1120.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannet, N., Kanin, E., Volkart, T.L., Wolson, C.J., Bell, S.P. and Young, R.A. 2000. Genome-wide location and function of DNA binding proteins. Science 290: 2306-2309.

Reymond, P., Weber, H., Damond, M. and Farmer E.E. 2000. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell 12: 707-719.

Richmond, T. and Somerville, S.C. 2000. Chasing the dream: plant EST microarrays. Curr. Opin. Plant Biol. 3: 108-116.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N. and Willmitzer, L. 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J. 23: 131-142.

Ruan, Y., Gilmore, J. and Conner, T. 1998. Towards Arabidopsis genome analysis: Monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. 15: 821-833.

Schaffer, R., Landgraf, J., Perez Amador, M. and Wisman, E. 2000. Monitoring genome-wide expression in plants. Curr. Opin. Biotechnol. 11: 162-167.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M. and Wisman, E. 2001. Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. Plant Cell 13: 113-123.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470.

Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., Somerville, S.C., and Manners, J.M. 2000. Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc. Natl Acad. Sci. USA 97: 11655-11660.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y. and Shinozaki, K. 2001. Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13: 61-72.

Shimkets, R.A., Lowe, D.G., Tai, J.T., Sehl, P., Jin, H., Yang, R., Predki, P.F., Rothberg, B.E., Murtha, M.T., Roth, M.E., Shenoy, S.G., Windemuth, A., Simpson, J.W., Simons, J.F., Daley, M.P., Gold, S.A., McKenna, M.P., Hillan, K., Went, G.T., and Rothberg, J.M. 1999. Gene expression analysis by transcript profiling coupled to a gene database query. Nature Biotech. 17: 798-803.

Spiegelman, J.I., Mindrinos, M.N., Fankhauser, C., Richards, D., Lutes, J., Chory, J. and Oefner, P.J. 2000. Cloning of the arabidopsis RSF1 gene by using a mapping strategy based on high-density DNA arrays and denaturing high-performance liquid chromatography. Plant cell 12: 2485-2498.

Suzuki, I., Kanesaki, Y., Mikami, K., Kanehisa, M. and Murata, N. 2001. Cold-regulated genes under control of the cold sensor Hik33 in Synechocystis. Mol. Microbiol. 40: 235-244.

Syvanen, A.C. 1999. From gels to chips: "Minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Human Mut. 13: 1-10.

Uetz, P., Giot, L., Cagney, G., Mansfield T.A., Judson R.S., Knight J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi E.A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. and Rothberg J.M. 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403: 623-627.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. 1995. Serial analysis of gene expression. Science 270: 484-487.

Vorst, O., Van Hal, N.L.W., Keijer, J. and Hall, R.D. 2001. cDNA microarrays as a tool for gene expression studies - theoretical and practical considerations. In: Focus on Biotechnology 5: Plant Biotechnology. Kluwer Academic Publishers, Dordrecht. *In press.*

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N.P., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M. and Lander, E.S. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082.

Wang, R., Guegler, K., LaBrie Samuel, T. and Crawford N.M. 2000. Genomic analysis of a nutrient response in Arabidopsis reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. Plant Cell 12: 1491-1509.

Weigel, D., Ahn, J., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F., Chory, J., Ahn, J.H., Wang, Z.Y. and Xia, Y.J. 2000. Activation tagging in Arabidopsis. Plant Physiol. 122: 1003-1013.

Welford, S.M., Gregg, J., Chen, E., Garrison, D., Sorensen, P.H., Denny, C.T. and Nelson, S.F. 1998. Detection of differentially expressed genes in primary tumor tissues using representation differences analysis coupled to microarray hybridization. Nucl. Acids Res. 26: 3059-3065.

Wisman, E. and Ohlrogge, J. 2000. Arabidopsis microarray service facilities. Plant Physiol. 124: 1468-1471.

Yang, G.P., Ross, D.T., Kuang, W.W., Brown, P.O. and Weigel, R.J. 1999. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. Nucl. Acids Res. 27: 1517-1523.

Yazaki, J., Kishimoto, N., Nakamura, K., Fujii, F., Shimbo, K., Otsuka, Y., Wu, J.Z., Yamamoto, K., Sakata, K., Sakaki, T. and Kikuchi, S. 2000. Embarking on rice functional genomics via cDNA microarray: Use of 3 ' UTR probes for specific gene expression analysis. DNA Res. 7: 367-370.

Zhu, T. and Wang, X. 2000. Large-scale profiling of the Arabidopsis transcriptome. Plant Physiol. 124: 1472-1476.

Zou, S., Meadows, S., Sharp, L., Jan, L.Y. and Jan, Y.N. 2000. Genome-wide study of aging and oxidative stress response in Drosophila melanogaster. Proc. Natl Acad. Sci. USA 97: 13726-13731.

CHAPTER 3

Asaph Aharoni, Leopold C. P. Keizer, Hetty C. van den Broeck, Gregory Bois, Patrick Smit, Ric C. H. De Vos and Ann P. O'Connell

Novel Insight into Vascular, Stress and Auxin Dependent and Independent Gene Expression Programs in Strawberry, a Non-Climacteric Fruit

Plant Physiology 129: 1019-1031 (2002)

NOVEL INSIGHT INTO VASCULAR, STRESS AND AUXIN DEPENDENT AND INDEPENDENT GENE EXPRESSION PROGRAMS IN STRAWBERRY, A NON-CLIMACTERIC FRUIT

Using cDNA microarrays a comprehensive investigation of gene expression was carried out in strawberry fruit in order to understand the flow of events associated with its maturation and non-climacteric ripening. We detected key processes and novel genes not previously associated with fruit development and ripening, related to vascular development, oxidative stress and auxin response. Microarray analysis during fruit development and in receptacle and seed (achene) tissues established an interesting parallelism in gene expression between the transdifferentiation of tracheary elements in *Zinnia elegans* and strawberry. One of the genes, *CAD*, common to both systems and encoding the lignin related protein cinnamyl alcohol dehydrogenase was immunolocalised to immature xylem cells of the vascular bundles in the strawberry receptacle. To examine the importance of oxidative stress in ripening, gene expression was compared between fruit treated on-vine with a free radical generator and non-treated fruit. Out of 46 genes induced, 20 were also ripening regulated. This might suggest that active gene expression is induced to cope with oxidative stress conditions during ripening or that the strawberry ripening transcriptional program is an oxidative stress induced process. In order to gain insight into the hormonal control of non-climacteric fruit ripening, an additional microarray experiment was conducted comparing gene expression in fruit treated exogenously with auxin and control fruit. Novel auxin dependent genes and processes were identified in addition to transcriptional programs acting independent of auxin mainly related to cell wall metabolism and stress response.

INTRODUCTION

The attractive characteristics of strawberry fruit are not only aroma, taste, color and texture but also their essential nutrient, mineral, vitamin content and anti-oxidant properties. Their anti-oxidant properties coupled with high dietary fibre content have been medically recognised as having positive influences on protecting against the risk of many diseases (Brownleader et al., 1999). To date, we still lack valuable information on the molecular events that control strawberry fruit development, ripening, and adaptation to environmental cues which are all complex biological processes involving the co-ordinated regulation of genes and biochemical pathways.

Unlike fruit botanically defined as arising from the expansion of the ovary, strawberry is actually the swollen base of the flower (receptacle) with one seeded fruit (termed achenes) located on the outer surface (Perkins-Veazie, 1995). Vascular bundles supply nutrients which move acropetally to the developing embryos in the achenes and surrounding cells of the receptacle (Hancock, 1999). In the ripening stage of strawberry fruit development the vascular tissue comprises long fibres composed of cellulose, protein, pectin and lignin (Suutarinen et al., 1998). Since strawberry fruit is composed of approximately 90% water and 10% total soluble solids, it is not inconceivable that the vascular system beginning from the achenes and connecting to the pith plays an important role in the texture and structural integrity of the ripe fruit (Jewell et al., 1973; Suutarinen et al., 1998). To date studies have neglected to explore the role of the vascular system in the development and ripening of strawberry fruit and instead have focused on the remainder of the receptacle tissue.

In plants tracheary element (TE) differentiation/xylogenesis has been extensively studied using *Zinnia elegans* mesophyll cell system. It commences with rearrangements of the microtubules in a cortical banding pattern that reflects the position of future secondary thickenings (Fukuda, 1997). Subsequently, cellulose is deposited in the initial thickenings, followed by lignification and cell death (Domingo et al., 1998). Programmed cell death (PCD) is an active process that occurs in plants during development and in response to environmental cues. Cell death occurring during differentiation of procambium into TE is one such example (Greenberg, 1996). During the PCD process, TEs degrade their cellular contents and become hollow corpses serving as a water conducting system.

Organ senescence is an example of a PCD process occurring in plants. Senescence is a dynamic and tightly regulated developmental process that involves an array of changes at both physiological and biochemical levels including gene expression. Fruit ripening is considered by some to be a specialised form of senescence (Seymour et al., 1993). A large number of biotic and abiotic factors accelerate the process. In fruit, external environmental factors such as heat (Cheng et al., 1988; Kagan-Zur et al., 1995), cold (Masia, 1998), salt (Avsian-Kretchmer et al., 1999) and ozone (Kirtikara and Talbot, 1996) have been proven to induce oxidative stress. Ripening itself however may impose stress conditions on the fruit. In grape, the accumulation of ten cDNAs

encoding putative stress response proteins upon ripening was recently reported (Davies and Robinson, 2000). To date, no studies testing the hypothesis that a transcriptional program related to stress, and in particular oxidative stress, exists in ripening fruit have been reported.

Concurrent with the supply of nutrients to the achenes (described above), the hormone auxin is translocated basipetaly through the phloem of the vascular bundles from the achenes to the peduncle (Perkins-Veazie, 1995). It has been unequivocally demonstrated that growth and early fruit development of strawberry is stimulated by auxin originating in the achenes (Nitsch, 1950). Later in fruit development (middle green stage) prior to ripening, auxin levels decline in the receptacle possibly due to the cessation of auxin transport from the achenes, and this invokes the ripening process (Given et al., 1988). Ripening triggered by reduced auxin levels is accompanied by *de novo* synthesis of specific mRNAs, which encode proteins responsible for the dramatic changes in fruit such as pigmentation and texture (Manning, 1994; Manning, 1998). In climacteric fruit such as tomato, banana, apple and melon, ethylene is the hormonal signal that triggers ripening; however not all ripening processes are ethylene dependent (Lelièvre et al., 1997). While it has been well documented that exogenous ethylene has no effect on the ripening process in non-climacteric fruit, it appears that both in strawberry and other fruit such as citrus and pineapple it may play a role (Alonso et al., 1995; Goldschmidt et al., 1993; Cazzonelli et al., 1998). Strawberry exhibits a low and slightly elevated level of ethylene production during the late stage of ripening (Perkins-Veazie et al., 1996). It would thus seem that in both climacteric and non-climacteric fruit, not all ripening processes are affected by the same hormone.

In this study, our goal was to better understand the processes underlying strawberry fruit maturation and non-climacteric ripening. By using DNA microarray technology we were able to perform large-scale and simultaneous investigation of gene expression during fruit development, in different tissues and after exposure to stress (oxidative stress) and hormonal treatments. The results highlighted two key processes active during fruit development and ripening relating to vascular development and oxidative stress. They also showed that not all the processes associated with strawberry ripening are under the same genetic control, and are probably a collection of processes regulated in a discrete manner.

RESULTS AND DISCUSSION

First and Second Generation Microarrays

In this paper we refer to strawberry 'fruit' as the receptacle including the seeds (achenes). The main stages of strawberry fruit development are depicted in Fig. 1A. The time course from anthesis (full petal opening) to medium green, large white, turning and red (ripe) stages of fruit development is approximately 10, 21, 24 and 30 days, respectively. We used microarrays comprising 1701 strawberry cDNAs (probes) derived from a red fruit cDNA library to perform 4

first-generation microarray experiments (Fig. 1B). The focal point of these 4 experiments was to identify ripening related genes and processes not previously disclosed. The first three experiments (hybridisations) compared green versus red (I), white versus red (II), and turning versus red (III) stages of fruit development. A fourth microarray experiment was performed comparing achene versus receptacle (IV), in order to differentiate between genes expressed in either of this two fruit tissues.

Combining the results from all four experiments, a total of 537 unique cDNAs were identified as differentially expressed at least once (Fig. 2A). 259 cDNA clones (48%) showed higher expression in the achenes (AchA, Fig. 2A) and 182 (34%) showed higher expression in the receptacle (RecA, Fig. 2A) (441 in total expressed in either achene or receptacle). Eighty eight percent of the achene cDNAs (228) and 56% of the receptacle cDNAs (102) were not developmentally regulated. Ninety-six cDNAs (18%) were equally expressed in either tissue type but were differentially expressed during development. A large number (42%) of the 537 cDNA clones identified were “unknown” or “novel”.

Based on these data a second generation microarray was prepared comprising 384 probes. This new microarray allowed us to focus our analysis primarily on ripening regulated receptacle associated cDNAs. Array elements included: 1) those showing elevated expression in the receptacle tissue, 2) those that were differentially expressed during development including ripening regulated cDNAs (112 individual cDNAs were identified as ripening regulated and 80 of them **(**RipR, Fig. 2B**)** were arrayed on the second generation microarray), 3) cDNAs identified in our original EST collection that did not show differential expression in the first 4 experiments and 4) appropriate controls (for a detailed description of array elements see Material and Methods).

Two additional microarray experiments were performed, an oxidative stress experiment (V) and an auxin experiment (VI). The first experiment was designed to identify ripening regulated cDNAs which might be also oxidative stress-induced (in the receptacle). The second experiment was performed in order to detect auxin dependent and independent ripening related cDNAs and processes. A schematic diagram showing the experimental outline of the first and second-generation microarray experiments is shown in Fig. 1B.

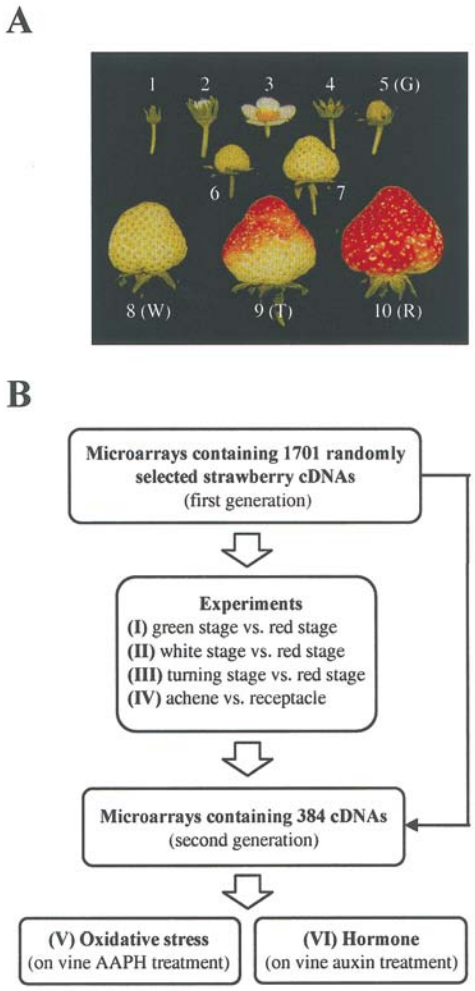


Figure 1 Using cDNA microarrays to follow gene expression patterns in strawberry fruit during development, in different tissues, under oxidative stress conditions and hormonal treatment. A, Strawberry fruit developmental stages. 1, small flower bud; 2, large flower bud; 3, anthesis; 4, small green; 5, medium green; 6, large green; 7, small white; 8, large white; 9, turning; 10, red. Stages 5, 8, 9 and 10 were used for microarray experiments (referred as G, W, T and R, respectively). B, Experimental schematic diagram of first (I to IV) and second-generation (V and VI) microarray experiments.

The capability and reproducibility of our microarray experiments in scoring differential gene expression was described in detail in a previous paper (Aharoni et al., 2000). Each of the microarray experiments described in this study was performed twice with the dyes reversed between the two replicates. A statistical analysis of variance model was used to evaluate the data and to determine a threshold value, which indicates a significant up or down regulation of gene expression (see Materials and Methods). We further demonstrated the quality and reliability of our microarray experiments by comparing gene expression results from RNA gel blots analyses with expression ratios originating from microarray data (Table I). It is beyond the scope of this report to describe in detail all the genes that were differentially expressed and novel processes identified. Several of our main discoveries are provided below.

A
Dev
207
AchA
259
RecA
182
80
96
31
96
B
AuxRep
45
RecA
182
AuxRep
45
RipR
80
54
StrInd
46
AuxRep
45
39
29
30
39
6
18
10
20
30
6
6
6
AuxRep + RipR + RecA + StrInd

Figure 2 (previous page) Differential gene expression in strawberry fruit. The diagrams show the numbers of overlapping and non-overlapping genes differentially expressed during fruit development (Dev), receptacle associated (RecA) and achene associated (AchA) as detected in experiments I to IV (A). Those genes auxin repressed (AuxRep), ripening regulated (RipR), receptacle associated (RecA) and stress induced (StrInd) detected in experiments V and VI and based on experiments I to IV (B).

The Importance of the Vascular System and Lignification in the Developing Receptacle

Gene expression analysis during strawberry fruit development (experiments I-III) revealed a group of 112 unique (i.e. distinct) ripening regulated genes. Ripening regulated genes were only those which showed higher levels of expression in the red stage compared to either the green, white or turning stages (in one or more of the three cases). The majority of the cDNA clones were either receptacle associated (64 out of 112) or did not show any difference in expression between the achene and receptacle tissues (41) and are possibly expressed in both tissues (as deduced from experiment IV). Several of them were previously associated with a specific ripening process (e.g. pigmentation, cell wall) while others had no previous recognised role in any or a certain aspect of fruit ripening. One such group comprised cDNAs encoding putative cinnamoyl-CoA reductases (CCR: JB116; JB196) and cinnamyl alcohol dehydrogenases (CAD: F193; F138; F122), enzymes performing the last committed steps in the biosynthesis of lignin (Chapple and Carpita, 1998).

Detailed analysis of *CCR* and *CAD* expression in different strawberry tissues and during fruit development and maturation using RNA gel-blots confirmed the microarray data and showed elevated levels of both transcripts in the red stage (Fig. 3A). Whilst the expression of *CCR* gradually increased during ripening, *CAD* expression decreased after the green stage (in the white and turning stages) before increasing again at the red stage (Fig. 3A). Expression of both genes could be detected in achene and receptacle (fruit with no achenes), petioles, leaves and flowers. Since these genes were strongly expressed in the ripening receptacle tissue we suspected that some of them might be actively expressed in the vascular bundles and associated with their lignification (Fig. 3B). To localise where active lignification is occurring in the fruit, we performed histochemical staining on sections from the four different stages of fruit development (green, white, turning and red) using the Weisner reagent (phloroglucinol-HCl). This reacts with aldehyde groups (cinnamaldehydes, benzaldehydes) in the lignin, giving characteristic deep reddish\purple coloration in the xylem of the vascular bundles (Clifford, 1974). Strong staining indicating the presence of lignin was detected in all stages of development in immature xylem cells of the fibrovascular strands of the receptacle (Fig. 3C and 3D).

Expression of a *CAD* cDNA homolog (F193: gene bank accession number U63534) in yeast cells and enzymatic activity assays demonstrated a cinnamyl alcohol dehydrogenase activity of the recombinant enzyme (i.e. only a clear activity was found using sinapylaldehyde and coniferaldehyde as substrates) (Blanco-Portales et al., *In press*). Immunological detection of the strawberry CAD (U63534) protein in the receptacle using its corresponding primary anti-

strawberry CAD polyclonal antiserum showed that this particular CAD protein was present during all stages of fruit growth and development and localised specifically to immature xylem cells undergoing active lignification (Fig. 3E to 3H). At this stage it cannot be ruled out that CAD enzyme activity in the receptacle might also be associated with wound response or with the biosynthesis of flavour compounds as suggested in an earlier study by Mitchell and Jelenkovic (1995). The authors (Mitchell and Jelenkovic) reported on a ripening regulated and receptacle specific CAD enzyme activity and correlated it with the interconversion of aldehydes and alcohols implicated in flavour of ripe strawberry fruit. However, substrate specificity of the recombinant CAD including the immunolocalisation data presented here clearly suggest a role for this particular CAD in the lignification of vascular elements in the receptacle.

Table I. Confirmation of differential expression detected by microarrays with RNA gel-blot analysis

Clone	Homolog definition (GenBank accession number) [a]	R/G ratio[b]	Gel blot[c]
			L G R
F193	Cinnamyl alcohol dehydrogenase (U63534)	8.8	
H159	Ripening induced protein (AJ001445)	8.8	
C121	UDP- glucose:flavonoid 3-O-glucosyl transferase (AF117267)	8.3	
A135	Chalcone synthase 2 (D26594)	5.5	
JB173	Chalcone reductase like (AC007259)	4.6	
G13	Dioxygenase (AC007504)	4.6	
JB172	Stress related protein (AF178990)	4.4	
E89	Limonoid UDP- glucosyl transferase (AB033758)	3.7	
B72	Ribosomal protein L7Ae-like (AL033545)	1.0	

[a]Definition and accession of nucleotide sequence of the first BLAST X homolog.
[b]Expression ratio detected by microarray experiment (I) comparing red (R) and green (G) fruit stages of development.
[c]Gel blots loaded with total RNA derived from leaf (L), green fruit stage (G) and red fruit stage (R).

Analogy in Gene Expression between Strawberry Fruit Development and Tracheary Element Differentiation in *Zinnia Elegans*

Apart from genes associated with lignification the expression pattern and putative identity of other clones suggested that a large set of genes detected in this study might be related to processes occurring in the vascular tissue. From our first three microarray experiments (I-III), we could deduce that 31 out of 112 (28%) distinct genes identified as ripening regulated, show similarity *to Zinnia elegans* genes expressed during the process of tracheary element (TE) differentiation. In

Table II we show the previously reported *Zinnia* genes associated with TE differentiation together with their strawberry ripening counterparts. It is feasible that the strawberry genes (or other members of a particular gene family) suggested here as vascular associated might function in other strawberry fruit ripening processes and tissues aswell.

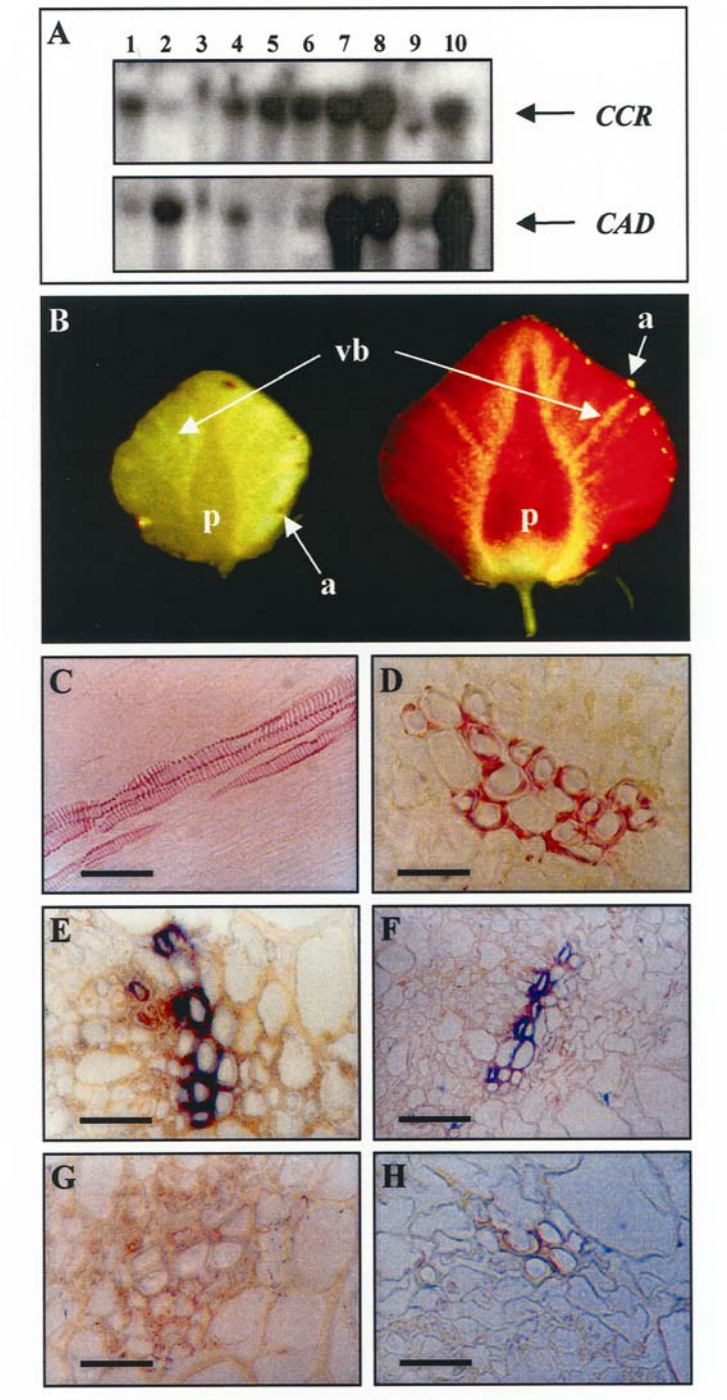


Figure 3 The vascular system and lignin associated gene expression and protein localisation in strawberry fruit. A. RNA gel-blot analysis of strawberry *CCR* and *CAD* expression in various strawberry tissues and during fruit development. 1- petiole; 2- leaf; 3- flower; 4- green fruit; 5- white fruit; 6- turning fruit; 7- red fruit; 8- red fruit without achenes; 9- achenes; 10- overripe fruit. The entire strawberry *CCR* and *CAD* cDNAs were used as probes for hybridisations. B, Section of a green and red ripe strawberry fruit showing fibrovascular strands (vb, vascular bundles)

connecting the achenes (a) to the interior of the receptacle (p, pith). C and D, show the presence of lignin in the vascular system (xylem vessels) in the receptacle, visualised after staining with phloroglucinol, E and F, show cross-sections of the receptacle stained by immunolocalization of CAD in the lignified vascular tissue (immature xylem) with the strawberry anti-CAD (F193) antiserum, G and H, show the receptacle stained by immunolocalization with pre-immune antiserum (negative controls). Sections C, E, G and D, F, H are green and red stage strawberry receptacle, respectively. In: C and F, bar = 12 μm; D and E, bar = 6 μm; H and G, bar = 7 μm.

Genes involved in the first phase of *Zinnia* TE differentiation are dominated by components of RNA and protein turnover machinery, such as ribosomal proteins and elongation factors type 1 (Table II; Fukuda, 1997). During strawberry ripening, a dramatic induction of genes related to DNA/RNA/protein turnover, such as those encoding elongation factors (type 1 and 2) and ribosomal proteins were also observed in the receptacle. A putative strawberry lipid transfer protein, similar to the *Zinnia TED4* gene, was also identified as ripening regulated. The TED4 protein is secreted into the apoplastic space and associated with morphological changes of TEs (Endo et al., 2001). TED4 is suggested to act as an inhibitor of the proteasome which induces TE differentiation and the progression of TE program in committed cells. By inhibiting the proteasome, TED4 protects healthy cells from injury due to proteolytic activities exudated from dying TEs.

During the second phase of *Zinnia* TE development, prior to secondary wall thickening, the cytoskeleton undergoes dynamic changes reflected by the accumulation of transcripts encoding tubulins (Fukuda, 1997). Tubulin synthesis increases the amount of microtubules, facilitating the regulation of secondary cell-wall formation in subsequent stages of TE development. The identification in strawberry of *Zinnia* tubulin homologs (*ZeTub1*, *ZeTub2*) never previously associated with ripening of soft fruit, showing a dramatic increase in expression during fruit development, provided strong supporting evidence to the analogy between the two systems.

Along with the increase in tubulin synthesis, changes in actin organisation occur, in which actin filaments form thick cables functioning in cytoplasmatic streaming (Fukuda, 1997). An important added value for gene expression analysis using microarrays is the association of genes with an unclear role in a certain process to those already identified showing similar expression profile (i.e. “guilt by association”). Using a similar approach we suggest that the dramatic accumulation of profilin in ripening (14 fold increase in expression between the green to red stages) combined with its specificity to the receptacle tissue is possibly related to its role in vascular bundle development. Profilin is an actin binding protein and therefore affects the structure of the cytoskeleton by regulating the organisation of actin filaments (at high concentrations, profilin prevents the polymerisation of actin, whereas it enhances it at low concentrations) (Kovar et al., 2000).

Table II. Parallels in gene expression between tracheary element differentiation in *Zinnia elegans* and strawberry development and ripening

***Zinnia* Gene**	**Reference[a]**	**Strawberry (PRI Clone)**	**Homolog Definition (GenBank accession) (E-value)[b]**	**Ratio[e]** R/G	R/W	R/T
			Activation of Protein Synthesis Apparatus and Dedifferentiation			
ZeRPS	Fukuda, 1997	F187	40S Ribosomal protein S13 (AB043974) (3e-54)	0.9	1.5	**2.6**
ZeRPS3a		C23	60S Ribosomal protein L13E (AJ132537) (6e-60)	**10.6**	2.1	**2.9**
ZeRPPO		H88	QM-like protein (AF295636) (1e-52)	**2.7**	1.8	**2.2**
ZeEF1β	Fukuda, 1997	C197	Elongation factor 1α (1) (AJ223969) (3e-69)	**4.0**	1.5	1.7
TED4	Fukuda, 2000	B90	Lipid transfer protein (2) (X79604) (4e-18)	1.3	**3.8**	1.7
	Koonce, 1995[c]	B105	Nuclear protein (AL357612) (2e-11)	0.7	1.5	**2.4**
			Changes in Cytoskeleton Including Actin Reorganisation			
ZeTub1,2	Yoshimura et al.,	G84	Beta-tubulin (1) (D63138) (3e-80)	**12.5**	**6.5**	1.5
	1996	H54	Alpha-tubulin (2) (U12589) (2e-66)	**3.4**	2.3	1.8
Ngr		C122	Profilin (AF129427) (4e-49)	**13.9**	**8.2**	1.8
Ngr		H78	Annexin (AF188832) (3e-59)	2.5	**3.5**	1.2
			Primary & Secondary Cell Wall Metabolism			
ZePel	Domingo et al., 1998	D86	Pectate lyase (1) (U63550) (3e-78)	**4.8**	0.8	0.6
ZeExp1,2,3	Im et al., 2000	F39	Expansin (1) (AF159563) (3e-28)	**13.3**	1.6	1.2
Ngr	Fukuda, 1997[d]	D56	Extensin (2) (AF026382) (3e-43)	**5.5**	0.5	0.6
	Ohdaira et al., 2001[c]	D15	Polygalacturonase (2) (U20431) (1e-19)	**4.0**	1.2	0.6
			Phenylpropanoid Pathway and Lignification			
ZCAD1	Fukuda, 2000	F122	Cinnamyl ADH (1) (U63534) (8e-57)	**4.3**	**5.2**	**3.5**
		F138	Cinnamyl ADH (2) (D13991) (3e-65)	**4.2**	**3.5**	1.5
		F193	Cinnamyl ADH (3) (U63534) (2e-53)	**16.0**	**8.2**	1.8
Ngr		JB116	Cinnamoyl-CoA reductase (1) (AJ295838) (3e-49)	**3.5**	1.5	0.8
		JB196	Cinnamoyl-CoA reductase (2) (X79566) (1e-06)	**4.0**	1.6	0.9
CAOMT	Ye and Varner, 1995	F102	O-methyl transferase (2) (AF220491) (1e-18)	**2.6**	1.0	1.3
			Programmed Cell Death			
ZCP-4	Fukuda, 1997	C56	Cysteine proteinase (3) (Z14028) (8e-42)	**4.1**	1.3	1.0
P48h-17	Ye and Varner, 1996	JB202	Cysteine proteinase (7) (Z99954) (3e-68)	**4.7**	1.6	0.6
TED2	Fukuda, 2000	E149	Quinone reductase-like protein (AL16372) (3e-36)	**3.3**	**3.7**	1.8
Ngr	Woffenden et al.,	JB75	Ubiquitin (1) (U29159) (2e-81)	**4.9**	0.9	1.5
	1988[d]	D53	Ubiquitin extension protein (AJ223329) (7e-48)	**2.9**	1.3	1.4

[a]Report on expression of the corresponding gene in *Zinnia elegans* (depicted under *Zinnia* Gene) during TE differentiation.

[b]Definition, accession (nucleotide sequence) and E-value (see Altschul et al., 1990) of the first BLAST X homolog. Not always the described strawberry cDNA shows first homology to the corresponding *Zinnia* gene in the BLAST search result. The number in parentheses following the putative definitions represent the number of the sequence contig in the case when more than one sequence showed similar BLAST result but did not align in the sequence alignment.

[c]Unpublished data (only sequence is currently available).

[d]Cases in which protein activity has been reported or suggested but not expression of the corresponding gene (ngr, no gene reported). If no report on protein exists the association with TE differentiation was suggested in this study.

[e]Expression ratios obtained in first three microarray experiments comparing red versus green (R/G), red versus white (R/W) and red versus turning (R/T) stages of fruit development. Values depicted in bold and shaded represent significant changes in expression (see Materials and Methods).

The processes depicted above each gene set (in bold) are known to occur during TE differentiation in *Zinnia elegans*.

Zinnia genes known to be involved in primary and secondary cell wall metabolism (prior to cell wall thickening) include pectate lyase (*ZePel*), expansins (*ZeEXP 1,2,3*), polygalacturonase (*ZePG1*), caffeic acid 3-O-methyltransferase (*CAOMT*), and cinnamyl alcohol dehydrogenases (*ZCAD1*). Apart from extensins, which are specifically associated with secondary walls of TEs (Fukuda, 1997), the expression of expansin genes was recently correlated with primary cell wall expansion and secondary cell wall thickening during *Zinnia* TE development *in-vitro* (Im et al., 2000). It is possible that the pectate lyase and expansin enzymes previously identified in strawberry as ripening regulated and associated with cell wall metabolism in the receptacle cells (Medina-Escobar et al., 1997; Civello et al., 1999) might be involved in remodelling the cell wall during the development of the vascular system.

In the third phase of *Zinnia* TE development, the deposition of secondary cell-wall components (secondary cell wall thickening and lignification) in conductive tissues consisting of dead TEs is tightly coupled to PCD (Fukuda, 2000). Hydrolytic activities of enzymes, such as cysteine proteases, and of the ubiquitin and proteasome systems have been implicated in the PCD process during organ senescence and tracheary element differentiation, acting both as mediators of signal transduction and as effectors of programmed cell death (Groover and Jones, 1999). A strawberry homologue of the *TED2* gene (E149) was identified as ripening regulated. *TED2* gene previously shown to be expressed in developing vasculature has homology to crystallin, a quinone oxidoreductase (Demura and Fukuda, 1994). It was also previously shown that plant γ cystallins play a distinct role in plant oxidative stress tolerance (Babiychuk et al., 1995).

Ripening Regulated Genes in the Receptacle are Induced by Oxidative Stress

Results from the first four experiments (I-IV) revealed more than a dozen putative stress related ripening regulated cDNAs which were preferentially expressed in the receptacle. This prompted us to initiate a stress experiment, in order to identify whether ripening regulated genes identified through our microarray study could form part of a transcriptional program responsive to oxidative stress conditions, which could arise in the receptacle during the fruit maturation process. Oxidative stress conditions develop from reactive oxygen species (ROS), that can be generated as a result of uncontrolled respiration and damaged electron flow in mitochondria, leading to the induction of stress- and detoxification-related gene expression (Leprince et al., 2000). ROS are natural by products of metabolism, and often result from exposure to free radical generating compounds such as natural quinones, xenobiotics and pollutants (Babiychuk et al., 1995).

In order to impose oxidative stress conditions, white stage fruit were treated on-vine with the free radical generating compound 2,2'-Azobis (2-amidinopropane) dihydrochloride (AAPH), a water soluble substance that decomposes thermally, yielding two carbon centered radicals, which subsequently react with oxygen to form peroxy radicals (Henkow et al., 1996). A concentration of 100 mM was selected for the AAPH treatment after performing two independent RNA gel blot

analyses using the strawberry ferritin cDNA as a probe (Fig. 4A), which is known to be induced under stress conditions (Deak et al., 1999). Gene expression in AAPH treated fruit and those treated with buffer only was compared using the second-generation microarrays (experiment V). The strong induction of ferritin transcript by oxidative stress was confirmed by this experiment, which revealed an additional 45 significantly induced cDNAs.

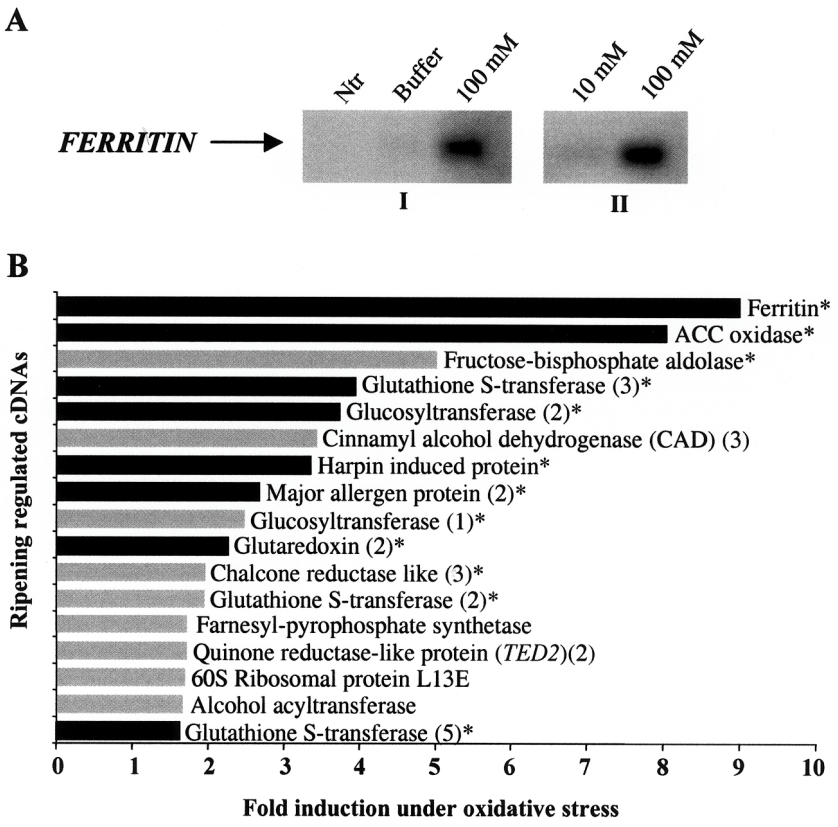


Figure 4 Identification of ripening regulated genes induced by oxidative stress. A, Ferritin gene expression in strawberry fruit following treatment with the free radical generator AAPH. Two RNA gel blot experiments were performed: (I) comparing expression in non-treated fruit (Ntr), fruit treated with buffer only (Buffer) and treated with buffer containing 100 mM AAPH (100 mM), (II) comparing expression in fruit treated with 10 mM and 100 mM AAPH (depicted as 10 mM and 100 mM respectively). Gene expression in fruit treated with 100 mM AAPH and control fruit treated with buffer only was compared using the second-generation microarray (experiment V). B, Graph representing cDNAs shown to be ripening regulated from microarray experiments (I-III) and oxidative stress induced from microarray experiment V. Grey bars represent genes showing differential and elevated expression in receptacle tissue compared to achenes, while black bars represent genes showing no differential expression between the two tissues as deduced from experiment IV. The number in parentheses following the putative definitions represent the number of the sequence contig in the case when more than one sequence showed a similar BLAST result but did not align in the sequence alignment. Genes marked with an asterisk did not show any change in expression upon application of exogenous auxin (as detected by experiment VI).

Table III. Identification of auxin dependent genes and processes in strawberry

PRI Clone	Auxin Response (fold)[a]	R/A[b]	Homolog Definition (GenBank accession number) (E-value)[c]	Putative Function
Auxin Repressed and Ripening Regulated Genes				
A104	5.9	8.4	Glutathione *S*-transferase (Y07721) (6e-67)	Anthocyanin sequestration
A135	6.7[d]	22.2	Chalcone synthase 2 (D26594) (6e-69)	Flavonoid pathway
C122	6.5	4.3	Profilin (AF129427) (4e-49)	Actin binding
C179	4.9[d]	6.5	Flavanone 3-hydroxylase (X69664) (9e-75)	Flavonoid pathway
C23	2.3	9.7 ↑	60S Ribosomal protein L13E (Z22620) (6e-60)	Unknown
D117	1.9[d]	3.7	Acyl carrier protein (AJ001446) (1e-29)	Fatty acid synthesis
D23	1.9[d]	10.2	NADPH-dependent oxidoreductase (AF108438) (2e-43)	Phytoalexin synthesis
E149	1.9	5.0 ↑	Quinone reductase-like protein (AL163972) (3e-36)	Reduction of quinones
E27	6.5[d]	16.3 ↑	Alcohol acyltransferase (AF193789) (3e-50)	Ester formation
E30	1.8[d]	8.5	Pectate lyase (U63550) (3e-54)	Cell wall degradation
E80	2.2[d]	8.7	Endo-1,4-beta-glucanase (AF074923) (3e-55)	Cell wall degradation
F1	2.0	n.d.	LTP (Q43681) (2e-22)	Lipid transfer
F102	3.6[d]	8.4	O-methyltransferase (AF220491) (1e-18)	Metabolites methylation
F157	2.6	20.3	Chalcone-flavonone isomerase (AB010692) (6e-38)	Flavonoid pathway
F193	2.0	3.3 ↑	Cinnamyl alcohol dehydrogenase (U63534) (2e-53)	Lignin biosynthesis
G13	2.7	8.8	Dioxygenase (AC007504) (4e-19)	Unknown
G175	2.1	3.5	Isoflavone reductase related (AF071477) (9e-46)	Phytoalexin synthesis
G84	7.2	3.7	Beta-tubulin (D63138) (3e-80)	Cytoskeleton organisation
H142	8.7	21.1	Dioxygenase (U97530) (1e-08)	Unknown
H159	4.8[d]	3.6	Ripening-induced protein (AJ001445) (9e-28)	Unknown
H51	2.0[d]	6.2	Pyruvate decarboxylase (AF193791) (2e-73)	Aldehyde formation
H61	3.3[d]	16.0	Flavanone 3-hydroxylase (X69664) (7e-73)	Flavonoid pathway
H81	2.4	5.5	Unknown (AL138646) (2e-30)	Unknown
JB136	2.9	4.6 ↑	Farnesyl pyrophosphate synthase (U15777) (2e-54)	Terpene metabolism
JB173	2.5[d]	12.6	Chalcone reductase-like (AC007259) (0.002)	Phytoalexin synthesis
JB19	2.0[d]	4.6	Malonyl-CoA decarboxylase (AF069441) (2e-47)	Fatty acid synthesis
JB202	3.0[d]	4.0	Cysteine proteinase (Z99954) (3e-68)	Protein breakdown
JB77	3.6	19.1	Anthocyanidin synthase (X7160) (4e-31)	Flavonoid pathway
Auxin Induced and Early-Mid Development Regulated Genes				
C176	2.2	n.d.	Copper homeostasis factor (AL163763) (8e-23)	Protection against oxidative stress
E1	1.7	3.5	Metallothionein-like protein (AJ001444) (2e-15)	Binding various heavy metals
F29	1.9	n.d.	ABC transporter-like (AL138656) (3e-32)	Unknown
H117	1.8	n.d.	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (X83499) (2e-56)	Methionine synthesis
H163	1.8	n.d.	Pectin esterase-like (AJ237985) (1e-21)	Cell wall metabolism
JB67	2.1	n.d.	SAM synthetase (AF271220) (7e-52)	*S*-adenosyl methionine synthesis

Pigmentation · Flavour · Stress Response
Fatty Acid Metabolism · Cell Structure · Unknown
Cell Wall Metabolism · Other

[a]Expression ratio detected in microarray experiment VI examining response to auxin. In the first set of genes it represents fold repression due to auxin treatment, while in the second set it represents fold induction by the auxin treatment.

[b]Expression ratio in the fourth microarray experiment (IV) comparing expression in receptacle tissue (R) vs. achene tissue (A). Genes identified in experiment V as stress induced are marked by ↑. n.d., No difference in expression between achene and receptacle.

[c]Definition, accession (nucleotide sequence) and E-value (Altschul et al., 1990) of the first BLAST X homolog.

[d]Described earlier as ripening regulated and auxin repressed (e.g. Manning, 1998).

According to their putative function, genes were associated to one of the processes depicted below the table.

Of the 46 induced cDNA clones, 20 were detected in earlier microarray experiments (I-III) as ripening regulated (StrInd; Fig. 2B) and 17 of them showing significant homology to other genes in the public databases are represented in Figure 4B. Results derived from experiment (IV) showed that 9 cDNAs (out of the 20) displayed elevated expression in receptacle tissue compared with achene tissue, while the rest (11 cDNAs) did not show differential expression between the two tissue types. As depicted in Figure 4B, the developing fruit appears to respond to the oxidative stress treatment with an increase in the production of detoxifying enzymes [glutathione *S*-transferases, glutaredoxin, quinone reductase-like protein (*TED2*)], protective enzymes (ferritin, 60S ribosomal protein L13E) and pathogenesis related proteins (harpin induced protein, CAD). Overexpression of ferritin was previously shown to confer resistance to free radical toxicity in tobacco plants (Deak et al., 1999). Interestingly, ferritin, which was the most responsive cDNA clone in the oxidative stress experiment (9 fold induction), contains an electrophile response element (EpRE) with sequence similarity to EpRE motifs found in antioxidant response genes, such as glutathione *S*-transferases and quinone reductase (Tsuji et al., 2000), also identified in this study as ripening regulated and oxidative stress induced. The strawberry glutaredoxin (2.3 fold increase) is a homolog of a glutaredoxin from *Ricinus communis L.*, an abundant sieve-tube exudute protein that was previously shown to prevent oxidative damage and to regulate the redox status of other sieve tube proteins (Szederkenyi et al., 1997).

As in other non-climacteric fruit, ethylene levels in strawberry fruit are very low compared to climacteric fruit such as tomato and banana. However, it has been observed that the levels of ethylene in strawberry decrease from green to white stage and then rise again in the red ripening stage (Perkins-Veazie et al., 1996). A strawberry ACC oxidase putatively encoding the enzyme catalysing the terminal step in the biosynthesis of ethylene was identified in this study. It showed 3 fold higher expression in red stage compared to the turning stage and a strong induction due to the oxidative stress treatment (8 fold). This could suggest an alternative function for ethylene in strawberry as a potential regulator of a stress response induced by ripening rather than triggering ripening itself. As it was previously reported that ethylene might play a role in the induction or progression of *Zinnia* TE differentiation (Fukuda, 1997), we tentatively suggest its possible involvement in a stress response associated with TE differentiation in strawberry (such as for example PCD in the vascular bundles). However, the correlation between ACC oxidase gene expression and ethylene formation in strawberry remains to be clarified. It has been previously demonstrated that the activity of the preceding enzyme (ACC synthase) in the ethylene biosynthetic pathway is the key step in ethylene biosynthesis (Cazzonelli et al., 1998). In this study we did not identify a strawberry cDNA encoding ACC synthase.

Correlating microarray expression data from the five experiments (I-V) has provided us with preliminary data to support our hypothesis that strawberry fruit may contain a transcriptional program responsive to stress (more particularly oxidative stress), induced during ripening.

Table IV. Identification of auxin independent and ripening regulated genes and processes in strawberry

PRI Clone	R/A[a]	Homolog Definition (GenBank accession number) (E-value)[b]	Putative Process
A145	n.d.	Unknown (AL138641) (7e-49)	Unknown
B1	6.0	Aldo keto reductase (AC002292) (5e-14)	Unknown
B73	n.d.	Protein kinase (AC009325) (1e-33)	Signal transduction
B85	n.d.	Myo-inositol-1-phosphate synthase (U38920) (4e-60)	Stress
B109	13.1	Adenosine kinase (AJ012281) (3e-79)	Primary metabolism
C54	6.9	Stearoyl-acyl carrier protein desaturase (X56508) (9e-49)	Fatty acid metabolism
C94	n.d.	HMGR (L10390) (3e-59)	Primary metabolism
C112	6.0	Ascorbate peroxidase (AF022213) (2e-51)	Stress
D5	2.4	Dehydrin (U69633) (2e-13)	Stress
D15	4.9	Polygalacturonase precursor (U20431) (1e-19)	Cell wall
D120	14.1	Globulin like protein (AF206627) (9e-24)	Unknown
D135	2.6	Unknown (AP002031) (3e-35)	Unknown
F22	12.5	Expansin (AF159563) (4e-54)	Cell wall
F93	3.7	Proline rich protein (AF026382) (2e-34)	Cell wall
G128	4.7	Inorganic pyrophosphatase (AF093629) (3e-25)	Primary metabolism
G144	n.d.	Short chain alcohol dehydrogenase (AC009273) (9e-29)	Stress
H78	11.0	Annexin (AF188832) (3e-59)	Stress
H168	2.4	Unknown (AF325033) (9e-20)	Unknown
H182	6.2	Ubiquitin (X98063) (3e-51)	Stress

[a]Expression ratio in microarray experiment (IV) comparing expression in receptacle (R) vs. achenes (A). n.d., No difference in expression between achene and receptacle.

[b]Definition, accession (nucleotide sequence) and E-value (see Altschul et al., 1990) of the first BLAST X homolog.

Auxin and Gene Expression in Strawberry Development and Ripening

Early in strawberry research, it was demonstrated that the decline in auxin levels supplied from the achenes to the receptacle tissue during fruit development was associated with the onset of strawberry fruit ripening (Given et al., 1988) and triggered ripening by inducing expression of ripening related genes (Manning 1994; Medina-Escobar et al., 1997; Manning, 1998; Harpster et al., 1998; Moyano et al., 1998). It is therefore expected that by artificially treating green strawberry fruit on the vine with exogenous auxin, one would suppress the transcription of ripening related genes. In this manner auxin repressed ripening related cDNAs as deduced from the microarray experiment (VI) are those normally up-regulated in the receptacle during ripening.

We used the second-generation microarray to perform a comprehensive examination of auxin action on gene expression and ripening processes in strawberry (experiment VI). By doing so, we could discriminate between auxin dependent (repressed or induced) and independent ripening genes and processes. Strawberry fruit on the vine were covered with paste with or without the auxin [1-naphthaleneacetic acid (NAA)]. Auxin treated fruit were morphologically similar to the non-treated fruit however, they did not accumulate anthocyanins, indicated by lack of red coloration typical of ripe strawberries. Samples generated from treated and non-treated fruit were used for comparative hybridization on the second-generation microarray.

Thirty ripening regulated cDNAs were repressed by the auxin treatment (out of 45 repressed in total) and 28 of them are depicted in Table III (cDNAs classified as "no hit" are not presented). Fruit ripening processes which appear to be auxin dependent include pathways related to pigmentation, stress/defence, cell wall metabolism, cell structure, fatty acid metabolism and flavour/aroma synthesis (aldehyde, ester and possibly terpene biosynthesis). Fourteen of the genes reported in this study as ripening regulated and auxin repressed were previously reported by Manning (1998) and others (e.g. Medina-Escobar et al., 1997; Harpster et al., 1998) and this provides an additional evidence to the quality and reliability of the microarray hybridisation data obtained. A cDNA encoding a dioxygenase like protein (H142) with unknown function showed the strongest repression (8.7 fold) by the auxin treatment. As expected, many cDNAs related to flavonoid metabolism and pigmentation were relatively strongly repressed by the auxin treatment (A104, A135, C179, F157, H61 and JB77). Interestingly, expression of the two receptacle and cell structure associated genes profilin (C122) and tubulin (G84) was strongly reduced by the auxin treatment (6.5 and 7.2 fold respectively).

Twenty-five cDNAs were induced by the auxin treatment. None of them was ripening regulated and 19 did not show any change in expression during development, as deduced from microarray experiments (I-III). However, the remaining 6 were both induced by auxin and showed elevated gene expression in early to mid strawberry fruit development compared to the red stage (green to white stage, Table III). Among them we identified a pectin esterase like protein (H163), which may be involved in early cell wall metabolism and fruit softening related to expansion, and other two cDNAs related to methionine biosynthesis [5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (H117) and S-adenosylmethionine (SAM) synthetase (JB67)].

Of the 80 individual ripening regulated cDNA clones identified in the microarray experiments (I-III) and arrayed on the second-generation microarray (RipR; Fig. 2B), 48 cDNAs (61.5%) did not show repression or induction by auxin and thus represent auxin independent ripening processes. Nineteen selected cDNAs out of the 48 are listed in Table IV and another 12 are depicted in Figure 4, as induced by the oxidative stress treatment (indicated by an asterisk). Although we have identified ripening regulated and auxin repressed cDNAs associated with certain metabolic processes such as fatty acid metabolism, cell wall and stress, other cDNA clones related to the same processes appeared to be auxin independent. For example, ripening regulated and cell wall related cDNAs such as expansin (F22), extensin-like/proline rich protein (F93) and polygalacturonase (D15) did not show any change in expression as a result of the auxin treatment (Table IV), while pectate lyase (E30) and endo 1,4 beta glucanase (E80) were repressed (Table III). Our observation that not all ripening regulated cell wall related genes in strawberry are auxin dependent is supported by a previous study on the strawberry expansin gene *FaExp2*, which was reported to be auxin insensitive (Civello et al., 1999). Interestingly, *FaExp2* expression was also not affected by ethylene treatment.

CONCLUSION

In this paper we have employed microarray technology to provide a comprehensive view of gene expression patterns during strawberry fruit development, in different tissues, under oxidative stress and auxin treatment conditions. The broad picture of gene expression obtained by our microarray analysis enabled new biological insights, which could not have been identified using conventional single observation methodologies. Combining the expression data from six different microarray experiments resulted in three major findings in relation to: i) a novel yet uncharacterised ripening process in strawberry namely the development of the vascular system; ii) the association between ripening related gene expression and oxidative stress response iii) hormone (auxin) dependent and independent processes.

One of the intriguing outcomes emerging from our microarray data analysis was the parallelism in gene expression patterns between TE differentiation in *Zinnia elegans* and strawberry. Based on the experimental data, although at this stage mainly correlative we would like to put forward the hypothesis that the development of the vascular system is a significant event coupled to fruit maturation. Whether vascular development in strawberry proceeds in the same way as in *Zinnia* still however remains to be established. This finding on the importance of gene expression in the vascular tissue of strawberry receptacle has some important implications concerning the possible function of genes identified earlier as ripening regulated in strawberry. Part of this genes might play a specific or additional role in the developing vasculature rather than only in the receptacle tissue itself.

The identification of 20 ripening regulated cDNAs induced by an oxidative stress treatment implies that oxidative stress could be part of certain strawberry ripening processes. The genes identified in this study might be triggered in order to actively respond to the stress conditions and/or play a role in different ripening processes induced by stress. Stress may arise in the fruit during ripening from changes in osmotic potential due to the accumulation and storage of osmotically active substances (e.g. hexoses), or from abiotic or biotic factors. A potential source of stress could also be (possibly in addition to other sources) the lignifying vasculature. Recently a basic peroxidase isozyme was located in the concentric array of the vascular bundles and in the vascular connections with the achenes in strawberry (Lopez-Serrano and Barcelo, 2001). Peroxidases are involved in the oxidation of phenolic compounds in cell walls, polymerisation of lignin and suberin and several other oxidation processes. Whether the activity of this peroxidase could contribute to oxidative stress conditions in the receptacle remains to be established.

Finally, we have identified novel ripening induced genes that were either repressed by auxin or not affected by the auxin treatment suggesting that another signal molecule(s) besides auxin may regulate the developmental ripening process in strawberry. A set of 25 genes were induced by the hormone, none of them were ripening regulated while 6 showed high expression levels in early to mid fruit development. High auxin levels are known to promote early fruit growth in strawberry

(Nitsch, 1950) and our data provides supporting evidence at the level of gene expression to this early observation. Auxin influence on gene expression in early fruit development was also reported in grapes, another non-climacteric fruit (Davies et al., 1997).

The results presented demonstrate the complexity of the hormonal control of non-climacteric fruit ripening, and indicates that the ripening process is likely to be a collection of sub-processes differentially regulated yet co-ordinated into a general ripening program. Further experiments to examine gene expression and protein localisation in the vascular bundles compared to cortical tissue, and in fruit treated with other phytohormones (e.g. ethylene, abscisic acid, gibberellins and cytokinins) will provide additional valuable data on the genetic controls governing ripening in strawberry.

MATERIALS AND METHODS

Plant Material and Preparation of mRNA

For developmental microarray experiments, medium-size green fruit, white fruit with no sign of pigmentation, turning (fruit are partially pigmented) and red ripe stage fruit obtained from the domesticated strawberry (*Fragaria* X *ananassa*) cultivar *Elsanta* were used. For the comparison of receptacle and achenes, 5 kg of red ripe fruits were blended with water and the achenes sinking to the bottom of the beaker were collected and used for RNA isolation. Achenes were removed manually from frozen red ripe fruits, and the remaining receptacle tissue was used for the comparison with achene tissue. Total RNA was prepared as described by Schultz et al. (1994). For mRNA preparation, the mRNA purification kit (Pharmacia Biotechnology) was used.

Production of First Generation Microarrays

The first generation microarrays were produced as described previously (Aharoni et al., 2000). Briefly, the source of the probes arrayed was a red ripe strawberry fruit tissue cDNA library including the achenes. The library was constructed in the UNI–XR vector (Stratagene La Jolla, CA). Following mass excision, plasmid DNA from 1701 strawberry picked randomly was extracted using the BioROBOT 9600 (Qiagen, Chatsworth, CA). The cDNAs were amplified by PCR using the T3 and T7 universal primers using the GeneAmp PCR system 9600 (Perkin Elmer, Foster City, CA). The primers contained a six-carbon amino modification (Isogen Bioscience BV, Maarssen). PCR products were purified using the QIAquick PCR purification kit (Qiagen) and eluted in 100 µL of 0.1 x Tris-EDTA, pH 8.0. Samples were dried to completion, resuspended in 7.5 µL of 5 X SSC (approximately 1 mg mL^{-1}) and transferred to a 384-format plate to be subsequently used for spotting. Amplified cDNAs were spotted in duplicate onto silylated

microscope slides (CEL Associates, Houston, TX) using a 16-pin print-head and a custom built arraying robot. After arraying, the slides were air dried and stored in the dark.

Production of Second Generation Microarrays

The second generation array contained 384 probes: 356 strawberry cDNAs, 16 peach fruit cDNAs, 1 petunia cDNA, 1 *NPTII* gene and 10 controls (5 fragments of the firefly luciferase gene, 2 mouse cDNAs and 3 probes not containing DNA). Amplification, purification and arraying of the probes were performed as described for first generation microarray with a few modifications. Samples were resuspended in 10 µL 5 X SSC prior to arraying using the PixSys 5500 (Chartesian Technologies, Irvine, CA) including ArrayIt, ChipMaker 3 micro spotting device and pins (TeleChem, Sunnyvale, CA). Arraying was performed on amino-silane coated slides (Corning, Corning NY). Each array was printed a second time at the opposite side of the same slide. After printing, the microarrays were re-hydrated above a beaker containing hot water for 5 s and then snap-dried for 2 s on a hot plate (100°C). The DNA was then ultraviolet (UV) crosslinked to the surface by subjecting the slides to 20 mJ energy (Stratagene, Stratalinker).

Hybridization and Data Analysis

For first and second generation microarrays, hybridization, scanning and data acquisition and statistical analysis were performed as described previously (Aharoni et al., 2000). Each of the microarray experiments was performed in a duplicate with the dyes reversed. For the first generation arrays we used two separate slides and for the second generation arrays, the duplicate experiment was performed on the same slide using the duplicated arrays and hybridising under separated coverslips. For the first three experiments green/red, white/red, and turning/red, the threshold ratio for detection (minimum ratio for differential expression) was 2.60, 3.32, and 2.24, respectively. For the microarray experiments comparing achene and receptacle tissues, oxidative stress and auxin, the threshold ratio of detection was 1.97, 1.63 and 1.73 respectively (in all experiments significant at single test $p<0.05$).

RNA Gel Blot Analysis

For RNA gel blot analyses, ten micrograms of glyoxal (1.5 M)-denaturated total RNA was electrophorized and blotted onto Hybond N^+ membrane (Amersham). After fixation (2 h 80°C) blots were hybridized as described by Angenent et al. (1992). The hybridization probes were made by random labelling oligonucleotide priming (Feinberg and Vogelstein, 1984) of the entire cDNAs. Blots were washed two times for half an hour each in 0.1 X SSC and 0.1% SDS at 65°C.

Immunolocalization of the Strawberry CAD Protein and Structural Studies

For the cyto-localization of the CAD polypeptide, tissues were fixed in ethanol-acetic acid (3:1, v/v), dehydrated through an ethanol-tertiary butanol series, and embedded in Paraplast Plus (Sherwood Med. Co., St Louis, MO). Five µm microtome sections were mounted on slides covered with gelatin, deparaffinized in xylene and re-hydrated through an ethanol series, and blocked with 2% non fat dried milk in TBS. Immunological detection was performed using a primary anti-strawberry CAD (clone F193) polyclonal antiserum diluted 1/25, and a secondary anti-rabbit alkaline phosphatase-conjugated antibody (Sigma) diluted 1/250. The reaction of alkaline phosphatase was developed with nitroblue tetrazolium and 5-bromo-4-chloro-3 indolyl-phosphate for 15-30 min.. The sections were dehydrated through graded ethanols, cleared in xylene and mounted in Entellan New (Merck). An Olympus AH-2 (Japan) photomicroscope was utilized for sample visualisation and photography. Lignified structures were visualised by performing the Weisner reaction using phloroglucinol-HCl (Clifford, 1974).

Oxidative Stress Treatment

For the oxidative stress experiment, white stage fruits (attached to the plants, approximately 20 fruit of 8 plants per treatment) were submerged in a solution containing 100 mM 2,2'-Azobis (2-amidinopropane) dihydrochloride (AAPH, Polysciences, Warrington, PA) dissolved in 10 mM 2-[N-Morpholino] ethanesulfonic acid (MES) buffer, pH 6.0, and 0.05% (w/v) Tween 20. Control fruits were submerged in the same buffer lacking AAPH. Fruit were submerged two times for 30 min in the solutions with a 17 h gap in between treatments. Six hours after the second treatment, fruits were picked, immediately frozen in liquid nitrogen and stored at -80°C until mRNA isolation. The AAPH concentration used in the microarray experiment was chosen from preliminary RNA gel blot experiments, using RNA from fruit treated with buffer only, 10 mM, and 100 mM AAPH. Four known stress related genes served as probes in these experiments.

Auxin Treatment

For hormone treatment we used 1-naphthaleneacetic acid (NAA) at a concentration of 0.5 mM in a lanolin paste containing 1% (v/v) Dimethylsulfoxide (DMSO). The paste was applied gently using a spatula to the entire fruit, on the vine, at the middle green stage of development (fruit grown on 20 plants were used). Seven days after the treatment, berries were picked, wiped clean of lanolin and used for total RNA isolation (25 berries for each sample). Control fruit were treated in a similar manner except for the absence of NAA in the paste.

Sequence Analysis

1100 cDNA out of a total of 1700 cDNAs were partially sequenced from the 5' end before performing the microarray experiments. Other non-sequenced cDNAs, which showed differential expression in the microarray experiments, were sequenced using the Applied Biosystems (Foster City, CA) dye terminator cycle sequencing Ready Reaction kit and the 310 DNA sequencer. Comparison analysis of the sequences was conducted with the advanced basic local alignment search tool, BLAST X (2.2.1) server (Altschul et al., 1990) and the National Center for Biotechnological Information (www.ncbi.nlm.nih.gov) non-redundant protein database. DNA and protein analyses was performed using Geneworks (IntelliGenetics, Oxford, UK) and DNASTAR (DNASTAR Inc. Madison, WI). cDNA clones showing differential expression that could not be classified to any functional category were annotated as unknown or novel. The "unknown" category included sequences showing significant homology to genes with unknown function. The "novel" category included sequences showing no homology at all (no hit) or low homology to database sequences.

ACKNOWLEDGEMENTS

The authors wish to thank Raffaella Greco and Dirk Bosch for critically reading the manuscript and Jan Schaart for providing the image in Figure 1.

REFERENCES

Aharoni A, Keizer LCP, Bouwmeester HJ, Sun ZK, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen A, De Vos RCH, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O'Connell AP (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12: 647-661

Alonso JM, Chamaro J, Granell A (1995) Evidence for the involvement of ethylene in the expression of specific RNAs during maturation of the orange, a non-climacteric fruit. Plant Mol Biol 29: 385-390

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Angenent GC, Busscher M, Franken J, Mol JN, van Tunen AJ (1992) Differential expression of two MADS box genes in wild-type and mutant petunia flowers. Plant Cell 4: 983-993

Avsian-Kretchmer O, Eshdat Y, Gueta-Dahan Y, Ben-Hayyim G (1999) Regulation of stress-induced phospholipid hydroperoxide glutathione peroxidase expression in citrus. Planta 209: 469-477

B abiychuk E, Kushnir S, Belles Boix E, Montagu Mv, Inze D, Van Montagu M (1995) Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance of yeasts toward the thiol-oxidizing drug diamide. J Biol Chem 270: 26224-26231

Brownleader MD, Jackson P, Mobasheri A, Pantelides AT, Sumar S, Trevan M, Dey PM (1999) Molecular aspects of cell wall modifications during fruit ripening. Crit Rev Food Sci 39: 149-164

Cazzonelli CI, Cavallaro AS, Botella JR (1998) Cloning and characterization of ripening-induced ethylene biosynthetic genes from non-climacteric pineapple (Ananas comosus) fruits. Aust J Plant Physiol 25: 513-518

Chapple C, Carpita N (1998) Plant cell walls as targets for biotechnology. Curr Opin Plant Biol 1: 179-185

Cheng TS, Floros JD, Shewfelt RL, Chang CJ (1988) The effect of high-temperature stress on ripening of tomatoes. J Plant Physiol 132: 459-464

Civello PM, Powell ALT, Sabehat A, Bennett Alan B (1999) An expansin gene expressed in ripening strawberry fruit. Plant Physiol 121: 1273-1279.

Clifford MN (1974) Specificity of acidic phloroglucinol reagents. J Chromatogr 94: 321-324

Davies C, Boss PK, Robinson SP (1997) Treatment of grape berries, a nonclimacteric fruit with a synthetic auxin, retards ripening and alters the expression of developmentally regulated genes. Plant Physiol 115: 1155-1161

Davies C, Robinson SP (2000) Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins. Plant Physiol 122: 803-812

Deak M, Horvath GV, Davletova S, Torok K, Sass L, Vass I, Barna B, Kiraly Z, Dudits D (1999) Plants ectopically expressing the ironbinding protein, ferritin, are tolerant to oxidative damage and pathogens. Nat Biotechnol 17: 192-196

Demura T, Fukuda H (1994) Novel vascular cell-specific genes whose expression is regulated temporally and spatially during vascular system development. Plant Cell 6: 967-981

Domingo C, Roberts K, Stacey NJ, Connerton I, Ruiz TF, McCann MC (1998) A pectate lyase from Zinnia elegans is auxin inducible. Plant J 13: 17-28

Endo S, Demura T, Fukuda H (2001) Inhibition of proteasome activity by the TED4 protein in extracellular space: A novel mechanism for protection of living cells from injury caused by dying cells. Plant Cell Physiol 42: 9-19

Feinberg AP, Vogelstein B (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum Anal Biochem 137: 266-267

Fukuda H (1997) Tracheary element differentiation. Plant Cell 9: 1147-1156

Fukuda H (2000) Programmed cell death of tracheary elements as a paradigm in plants. Plant Mol Biol 44: 245-253

Given NK, Venis MA, Grierson D (1988) Hormonal regulation of ripening in the strawberry, a non-climacteric fruit. Planta 174: 402-406

Goldschmidt EE, Huberman M, Goren R (1993) Probing the role of endogenous ethylene in the degreening of citrus fruit with ethylene antagonists. Plant Growth Regul 12: 325-329

Greenberg JT (1996) Programmed cell death: A way of life for plants. Proc Natl Acad Sci USA 93: 12094-12097

Groover A, Jones AM (1999) Tracheary element differentiation uses a novel mechanism coordinating programmed cell death and secondary cell wall synthesis. Plant Physiol 119: 375-384

Hancock JF (1999) Strawberries. CABI Publishing, Wallingford, UK, pp 131-148

Harpster MH, Brummell DA, Dunsmuir P (1998) Expression analysis of a ripening-specific, auxin-repressed endo-1, 4-beta-glucanase gene in strawberry. Plant Physiol 118: 1307-1316

Henkow L, Strid A, Berglund T, Rydstrom J, Ohlsson AB (1996) Alteration of gene expression in Pisum sativum tissue cultures caused by the free radical-generating agent 2,2'-azobis (2-amidinopropane) dihydrochloride. Physiol Plant 96: 6-12

Im KH, Cosgrove DT, Jones AM (2000) Subcellular localization of expansin mRNA in xylem cells. Plant Physiol 123: 463-470

Jewel GG, Rantsios A and Scholey J (1973) Factors influencing the breakdown of fruit in strawberry jam. J Texture Stud 4: 363-370

Kagan-Zur V, Tieman DM, Marlow SJ, Handa AK (1995) Differential regulation of polygalacturonase and pectin methylesterase gene expression during and after heat stress in ripening tomato (Lycopersicon esculentum Mill.) fruits. Plant Mol Biol 29: 1101-1110

Kirtikara K, Talbot D (1996) Alteration in protein accumulation, gene expression and ascorbate-glutathione pathway in tomato (Lycopersicon esculentum) under paraquat and ozone stress. J Plant Physiol 148: 752-760

Kovar DR, Drobak BK, Staiger CJ (2000) Maize profilin isoforms are functionally distinct. Plant Cell 12: 583-598

Lelièvre JM, Latche A, Jones B, Bouzayen M, Pech Jean C (1997) Ethylene and fruit ripening. Physiol Plant 101: 727-739

Leprince O, Hoekstra FA, Harren FJM (2000) Unravelling the responses of metabolism to dehydration points to a role for cytoplasmic viscosity in desiccation tolerance. *In* M Black, KJ Bradford, J Vazquez Ramos eds, Seed biology: advances and applications. CABI Publishing; Wallingford; UK, pp 57-66

Lopez-Serrano M, Barcelo AR (2001) Histochemical localization and developmental expression of peroxidase and polyphenol oxidase in strawberries. J Am Soc Hortic Sci 126: 27-32

Manning K (1994) Changes in gene expression during strawberry fruit ripening and their regulation by auxin. Planta 194: 62-68

Manning K (1998) Isolation of a set of ripening-related genes from strawberry: their identification and possible relationship to fruit quality traits. Planta 205: 622-631

Masia A (1998) Superoxide dismutase and catalase activities in apple fruit during ripening and post-harvest and with special reference to ethylene. Physiol Plant 104: 668-672

Medina-Escobar N, Cardenas J, Moyano E, Caballero JL, Munoz-Blanco J (1997) Cloning, molecular characterization and expression pattern of a strawberry ripening-specific cDNA with sequence homology to pectate lyase from higher plants. Plant Mol Biol 34: 867-877

Mitchell WC, Jelenkovic G (1995) Characterizing NAD- and NADP-dependent alcohol dehydrogenase enzymes of strawberries. J Am Soc Hort Sci 120: 798-801

Moyano E, Portero-Robles I, Medina-Escobar N, Valpuesta V, Munoz-Blanco J, Caballero JL (1998) A fruit-specific putative dihydroflavonol 4-reductase gene is differentially expressed in strawberry during the ripening process. Plant Physiol 117: 711-716

Nitsch JP (1950) Growth and morphogenesis of the strawberry as related to auxin. Am J Bot 37: 211-215

Perkins-Veazie P (1995) Growth and ripening of strawberry fruit. Hort Rev 17: 267-297

Perkins-Veazie PM, Huber DJ, Brecht JK (1996) In vitro growth and ripening of strawberry fruit in the presence of ACC, STS or propylene. Ann Appl Biol 128: 105-116

Schultz DJ, Craig R, Cox-Foster DL, Mumma RO, Medford JI (1994) RNA isolation from recalcitrant plant tissue. Plant Mol Biol Rep 12: 310-316

Seymour GB, Taylor JE, Tucker GA (1993) Biochemistry of fruit ripening. Chapman & Hall, London, pp 454

Suutarinen J, Anakainen L, Autio K (1998) Comparison of light microscopy and spatially resolved Fourier transform infrared (FT-IR) microscopy in the examination of cell wall components of strawberries. Lebensm Wiss u-Technol 31: 595-601

Szederkenyi J, Komor E, Schobert C (1997) Cloning of the cDNA for glutaredoxin, an abundant sieve-tube exudate protein from Ricinus communis L. and characterisation of the glutathione-dependent thiol-reduction system in sieve tubes. Planta 202: 349-356

Tsuji Y, Ayaki H, Whitman SP, Morrow CS, Torti SV, Torti FM (2000) Coordinate transcriptional and translational regulation of ferritin in response to oxidative stress. Mol Cell Biol 20: 5818-5827

Woffenden BJ, Freeman TB, Beers EP (1998) Proteasome inhibitors prevent tracheary element differentiation in zinnia mesophyll cell cultures. Plant Physiol 118: 419-429

Ye ZH, Varner JE (1995) Differential expression of two O-methyltransferases in lignin biosynthesis in Zinnia elegans. Plant Physiol 108: 459-467

Ye ZH, Varner JE (1996) Induction of cysteine and serine proteases during xylogenesis in Zinnia elegans. Plant Mol Biol 30: 1233-1246

Yoshimura T, Demura T, Igarashi M, Fukuda H (1996) Differential expression of three genes for different beta-tubulin isotypes during the initial culture of Zinnia mesophyll cells that divide and differentiate into tracheary elements. Plant Cell Physiol 37: 1167-1176

CHAPTER 4

Asaph Aharoni and Ann P. O'Connell

Microarray Gene Expression Analysis During Strawberry Achenes and Receptacle Maturation

Journal of Experimental Botany, *In Press* (2002)

GENE EXPRESSION ANALYSIS OF STRAWBERRY ACHENE AND RECEPTACLE MATURATION USING DNA MICROARRAYS

Large scale single pass sequencing and parallel gene expression analysis using DNA microarrays were employed for the comprehensive investigation of ripening in strawberry fruit. A total of 1701 cDNA clones (comprising 1100 strawberry ESTs and 601 un-sequenced cDNAs) obtained from a strawberry (*Fragaria* x *ananassa*) ripe fruit cDNA library were displayed on microarrays, and utilised for monitoring concurrent gene expression in receptacle and achene tissues. Analysis of expression ratios identified 66 out of the 259 (25%) achene related clones and 80 out of 182 (44%) receptacle related clones with more than 4 fold difference in expression between the two tissue types. Half of the achene associated genes putatively encode proteins with unknown function, and a large number of the remainder were proteins predicted to form part of signal and regulation cascades related to achene maturation and acquisition of stress and desiccation tolerance. These included phosphatases, protein kinases, 14-3-3 proteins, transcription factors and others. In the receptacle we identified key processes and novel genes that could be associated with ripening. Genes putatively encoding proteins related to stress, cell-wall, DNA/RNA/protein and primary metabolism where highly represented. Apart from providing a global observation on gene expression programs and metabolic pathways in developing strawberry our study has made available a large database and unique information for gene discovery, promoter selection and markers for molecular breeding approaches.

INTRODUCTION

Strawberry is deemed a false fruit in that what is commonly called the fruit originates from the expansion of the flower base (the receptacle) as a pseudocarp, whereas the real one-seeded fruits (achenes) are actually on the epidermal layer. The mature achene is composed of a hard and relatively thick pericarp, a thin testa, an endosperm consisting of one cell layer, and a small embryo (Perkins Veazie, 1995). Strawberry fruit is defined as non-climacteric since it does not exhibit a peak in respiration and ethylene production during ripening. In climacteric fruit, ethylene provides the signal for ripening by activating the transcription of many genes related to fruit ripening. Ethylene is present in strawberry, but it's influence on non-climacteric fruit ripening is yet not fully understood. Strawberry fruit development, on the other hand, is strongly influenced by auxin which positively effects the initial growth phase of the receptacle. Later in fruit development (middle green stage prior to ripening), auxin levels decline in the receptacle possibly due to the cessation of auxin transport from the achenes, and this invokes the ripening process (Given *et al.,* 1988).

In many plants fruit development can be divided into four distinct phases (Gillaspy *et al.,* 1993). The first phase commences after flower opening (anthesis) and involves fertilisation and development of the ovary (in so-called true fruit) and is generally referred to as fruit set. In the second phase fruit growth by cell division is the most prominent process and is accompanied by seed and early embryo formation. In the third phase, following cell divisions, fruit growth is mainly due to an increase in cell volume. During this stage of fruit expansion, the embryo passes through a maturation phase. This phase often leads to the induction of seed dormancy and is characterised by (a) the accumulation of storage products (b) the suppression of precocious germination (c) the acquisition of desiccation tolerance and (d) water loss (Bewley and Black, 1994). The fourth phase is the ripening phase. Ripening is an aspect of fruit development that is initiated after seed maturation has almost been completed, supporting the hypothesis that seeds influence fruit development and ripening. As in other fruit, strawberry ripening is characterised by a rise in the content of soluble solids in the receptacle, the production of natural aroma and flavour compounds and alterations to fruit shape, size, texture and pigmentation.

In recent years a number of groups reported on the cloning and characterisation of genes associated with various aspects of strawberry fruit ripening. Of main interest were genes related to cell-wall metabolism since fruit softening, especially in strawberry, is an important post-harvest quality trait. In most cases the genes identified were expressed preferentially in receptacle tissue, whilst just a few could be associated with the developing achenes. Various strategies were utilised for gene discovery, such as the use of degenerate oligonucleotides for either direct amplification of cDNA by Reverse Transcriptase-PCR (Kim and Chung, 1998; Llop-Tous *et al.,* 1999; Harrison *et al.,* 2001; Trainotti *et al.,* 2001) or for screening a cDNA library (Harpster *et al.,* 1998). Other approaches involved either screening a cDNA library using known cDNAs (Civello *et al.,* 1999) or

the use of several different methods for analysis of differential gene expression (i.e. differential display, Wilkinson *et al.,* 1995; differential screening of cDNA libraries, Manning, 1998 and Nam *et al.,* 1999; differential screening of a subtracted cDNA library combined with Southern blot analysis, Medina-Escobar *et al.,* 1997).

This paper describes a different strategy for the isolation of genes associated with ripening of strawberry. The approach combines large-scale cDNA sequencing followed by public database searching for putative gene homologs, with comprehensive gene expression analysis using DNA microarrays. The generation of a large collection of single pass cDNA sequences known as ESTs (Expressed Sequence Tags) has emerged in the last years as an alternative gene discovery tool. An EST collection from non-model, exotic plants, with hardly any sequence information, and with a typical metabolic process often contains the relevant genes of a specific pathway under investigation (Lange *et al.,* 2000). Although highly abundant mRNAs are represented at high frequency, the use of a random sequencing approach has proven most effective for the discovery of genes involved in specific biochemically characterised plant metabolic pathways (Ohlrogge and Benning, 2000).

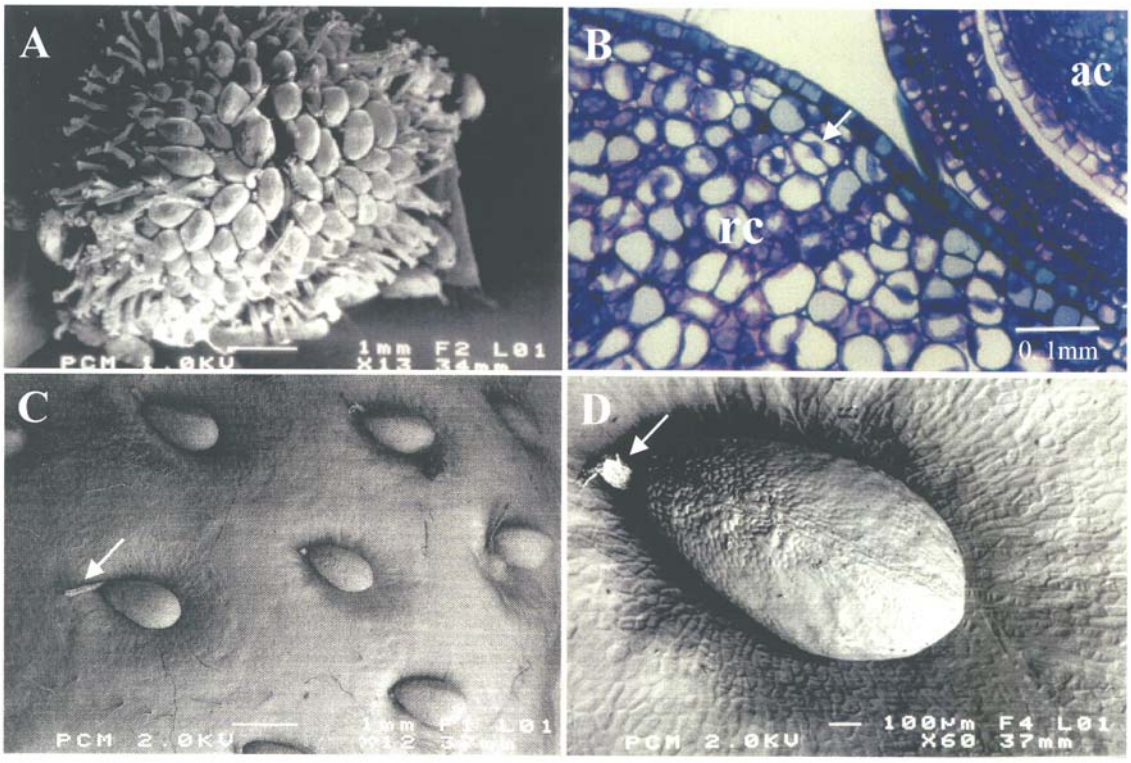


Figure 1 Cytological examination of strawberry receptacle and achenes **(A)** Scanning electron micrograph of an early green stage fruit, showing the close arrangement of the achenes. **(B)** Aniline blue stained section of an early green stage fruit, showing the high frequency of cell divisions in the receptacle (rc) parenchymous tissue (a cell divided is marked with an arrow). In the upper right corner a transverse section of an achene (ac) is visible. **(C)** Scanning electron microscopical picture of a red stage strawberry receptacle with achenes embedded in the epidermal layer. **(D)** A closer look at an achene. In (C) and (D) the arrows point to the remainder of the style connected to the achenes.

Complementing the major improvements in sequencing technologies, several methods have been developed in recent years which allow large-scale measurements of gene expression (Aharoni and Vorst, 2002). One such method employs a highly dense array of mechanically deposited DNA samples on a glass surface for the hybridisation of flourecently labelled reference and test RNA populations, allowing parallel and quantitative comparison of transcript levels in the samples investigated. When amplified cDNAs are used to construct such a microarray the method is often referred to as "cDNA microarrays". Simultaneous gene expression analyses of each individual EST under various conditions can provide evidence for the function of an unknown gene mainly when its expression profile is similar to the one of genes with known function. Correlating expression profiling data of genes with a known or putative function provides a global view of active expression programs.

Such "mining of gene expression data" has recently proven successful for identifying several candidate genes in strawberry for deeper functional investigation (Aharoni *et al.,* 2000, Schwab *et al.,* 2001). In this study we describe the identification of key processes and genes active simultaneously in two fruit tissues (achene and receptacle) during strawberry development and ripening. The results demonstrate the substantial difference in gene expression programs active in parallel in these two tissues and provide the basis for investigating how responses are co-ordinated between the two tissues to ensure seed dispersal.

RESULTS AND DISCUSSION

Production of a Strawberry EST Collection

In this paper we refer to strawberry 'fruit' as the receptacle (pseudocarp) including the seeds (achenes). The achenes are formed from the carpal and a single seed (combination of seed and ovary tissue) (Perkins Veazie, 1995). They are embedded in the receptacle epidermis and fibrovascular strands connect them to the interior of the receptacle (Figure 1). We used red ripe strawberry fruit tissue of the cultivated octaploid strawberry (*Fragaria* x *ananassa*) variety *Elsanta* for construction of a cDNA library. After performing mass excision of the library, 1100 cDNA clones where randomly picked and sequenced from their 5' end. The average length of the sequences obtained was 500 bp. Sequence information was analysed for homology to other gene sequences publicly available (using the BLASTX and BLASTN programs) and the data was transferred into a home made database. Forty percent of the sequences did not show significant homology to sequences present in the public databases (BLASTX score below 80). Amongst the clones that showed significant homology, 70% showed similarity to known sequence from the plant kingdom.

Production of Strawberry cDNA Microarrays

The entire set 1701 strawberry cDNAs (comprising 1100 sequenced strawberry ESTs and 601 un-sequenced cDNAs) were amplified and spotted in high density on glass microscope slides. The first three hybridisations compared a) green versus red b) white versus red and c) turning versus red stages of fruit development. After correction for redundancy (performed by sequence alignment), 239 unique differentially expressed cDNA clones were identified. A detailed technical description of these three microarray experiments including statistical analysis is provided in a recent publication (Aharoni *et al.,* 2000). Microarray data will not always reflect gene expression levels (nor protein abundance or enzyme activity) however, in some cases we will refer to the changes in mRNA abundance measured by the method as changes in gene expression as well.

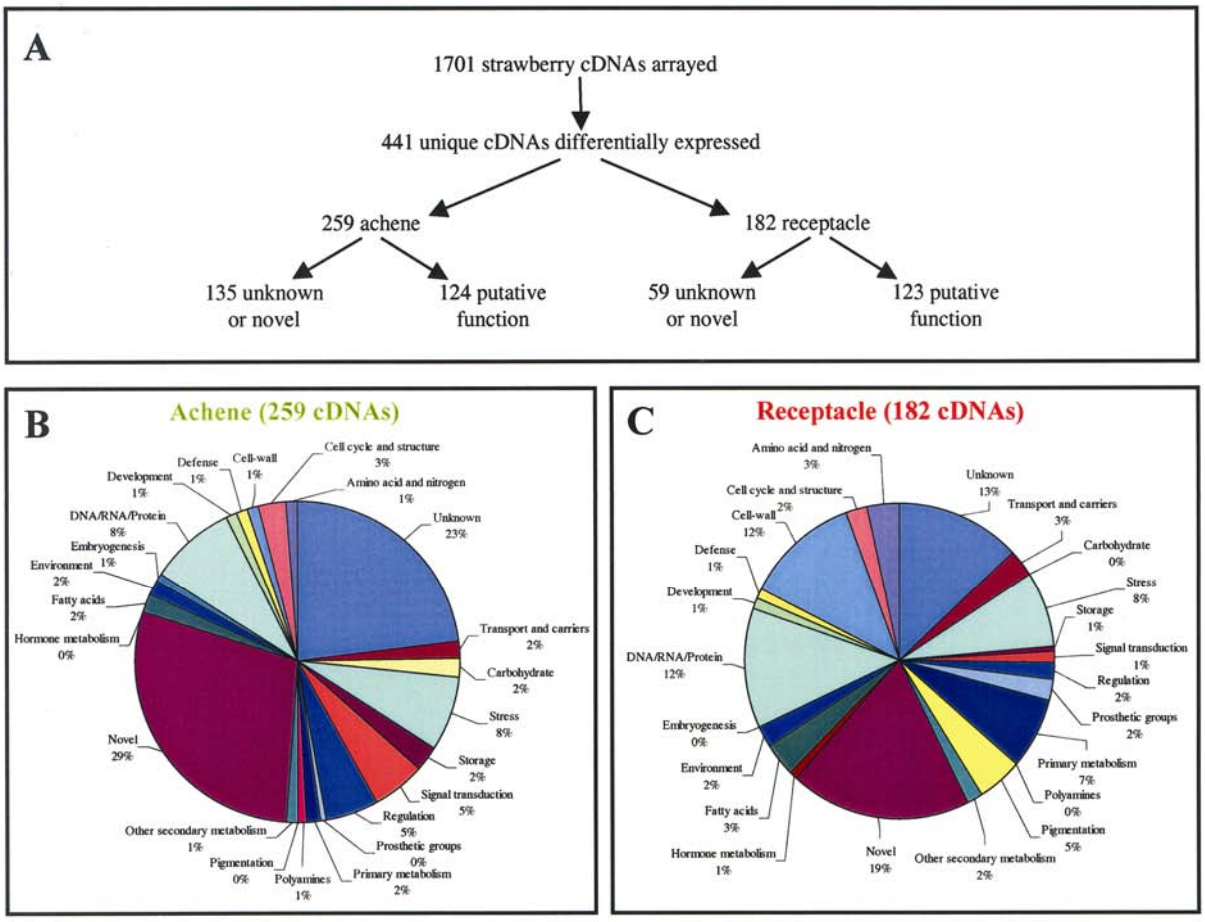


Figure 2 Classification of achene or receptacle associated cDNAs to functional categories. **(A)** The microarray experiment comparing RNA derived from achene and receptacle tissues revealed 441 unique cDNAs (based on sequence alignments) showing significant levels of differential expression between the 2 tissues. **(B)** and **(C)** represent the classification of cDNAs associated with each tissue to functional categories based on their homology to genes in the public databases. Percentages are calculated based on the total cDNAs showing elevated expression in each tissue (259 achenes and 182 receptacle). Definitions of cDNAs classified as “Unknown” or “Novel” are explained in the main text.

As the cDNA library from which the probes for microarray was prepared from whole fruit containing a mixture of achene and receptacle tissue, we were capable of performing a fourth experiment to compare gene expression between the ripening stage achene and receptacle tissues. From this experiment after correction for redundancy 441 unique differentially expressed cDNA clones were identified (Figure 2). Two hundred and fifty nine cDNA clones (48%) showed higher expression in the achenes and 182 (34%) higher expression in the receptacle. A large number of genes can be considered as achene or receptacle associated. Sixty six out of the 259 (25%) achene related clones and eighty out of 182 (44%) receptacle related clones showed more than 4 fold difference in expression between the two tissue types. One hundred and ninety four unique clones (44%) showed differential expression between the two tissues and could not be assigned to any functional category. These were classified as "unknown" or "novel". The "unknown" category included sequences showing significant sequence identity to genes with unknown function while the "novel" category included sequences showing no hit in the search.

Combining the results from all four experiments, after correction for redundancy, a total of 537 unique cDNAs differentially expressed at least once were identified. Each clone was assigned to one of 23 categories (including "unknown" and "novel" genes) on the basis of its BLAST search output. It is beyond the scope of this paper to describe in detail all the genes that were differentially expressed. Instead, we ordered and presented 313 cDNA clones (58%, excluding genes belonging to the categories "unknown" and "novel") on the basis of their expression profiles and putative functions of their closest homologs following the procedure of (Eisen *et al.,* 1998) (Figure 3A-U). It must be recognised that some of the genes may belong to more than one category.

Key Processes Taking Place in Achene and Receptacle

Although physically associated, the achene and receptacle are shown to possess distinct transcriptional programs during fruit development and maturation. A list containing 15 of the most highly differentially expressed cDNAs (as deduced from their expression ratios) pertaining to either the achene or receptacle tissue is depicted in Table 1. From the microarray study (Figure 2, Figure 3, Table 1) two categories, well represented in both tissues during the fruit maturation process, relate to the turnover of DNA, RNA, and proteins (category A) and to stress responses (category C). In the achenes other well represented categories relate to signal transduction (category U), regulation (category T), storage (category Q) and carbohydrates (M). In the receptacle other well represented categories relate to cell wall modification (category D), pigmentation (category E) and primary metabolism (category B). Achene related genes showed quite different expression profiles and less dramatic fluctuations in expression ratios (i.e. relative mRNA abundance levels) during development compared with receptacle specific cDNAs. Receptacle expressed genes were more prominent in the latter stages of fruit development.

Gene Expression Profiling of Developing Achenes

Gene expression patterns in achenes reflect physiological events occurring during seed development and maturation. Events occurring normally in seeds during the early stages of development include morphogenesis, embryogenesis and cell elongation, followed by storage material deposition and acquisition of stress tolerance. During maturation, seeds prepare for survival in a quiescent state and accumulate the necessary storage and protective components to enable survival for prolonged periods prior to germination. Protection against desiccation-induced injury, including damage by reactive oxygen species (ROS) resulting from imbalance of the pro-oxidant/antioxidant homeostasis (oxidative stress), is a necessary component of genetic programs active during late seed development (Stacy *et al.,* 1999). Plants have adopted the potential of interactions with oxygen (i.e. active oxygen species) for metabolic regulation. There are many signal transducing molecules (i.e. ethylene, abscisic acid and salicylic acid) implicated in the modification of gene expression (i.e. detoxification and protection/defence related cDNAs) mediated by redox regulated transcription factors (i.e. secondary messengers).

ABA and Strawberry Achene Maturation

Microarray profiling of achene related cDNAs revealed several distinct gene clusters. One prominent cluster contained simultaneously up regulated cDNAs related to signal transduction and transcriptional regulation, some of which were previously identified as being abscisic acid (ABA) regulated (Figure 3T and 3U, Table 1). Previous studies showed that ABA plays a crucial role in the seed maturation process. It induces many genes essential during this period (i.e. seed storage proteins, lipid and embryogenesis genes etc.), and is required for active repression of germination. In addition ABA plays a role in the adaptation to abiotic environmental stresses such as drought, salt and cold stress (Finkelstein *et al.,* 1998). In *Arabidopsis*, ABA insensitive seed mutants (abi) display reduced sensitivity to ABA or stress causing defects in seed storage reserve accumulation, maturation, dormancy, and expression of a variety of stressed induced genes (Finkelstein *et al.,* 1998). Members of the protein phosphatase 2C (PP2C) family (i.e. ABI1 and ABI2 of *Arabidopsis*; Leung *et al.,* 1994; Leung *et al.,* 1997; Meyer *et al.,* 1994) were identified as being required for wild-type ABA response. *Arabidopsis* ABI1 and ABI2 mutants showed altered vegetative and seed ABA regulated functions (e.g. reduced dormancy). In the strawberry signal transduction cluster (Figure 3U) we identified 3 different strawberry achene related cDNAs (Figure 3U, Table 1A: clone G81) showing homology to the PP2C family of protein phosphatases, closely related to *ABI1* and *ABI2*. In *Arabidopsis* the ABI1 is a negative regulator of ABA signalling (Gosti *et al.,* 1999). Other genes forming part of this cluster and which could be components of an ABA signal transduction pathway included a RAS related small GTP binding

protein, a putative serine/threonine protein kinase (*ARSK1*), two calcium binding EF hand proteins [one calmodulin and a second EF hand/embryo specific protein (EFA27 homolog) in category N] and two 14-3-3 proteins. Previous studies have shown that 14-3-3 proteins are involved in signalling pathways, generally functioning as adaptors, chaperones, activators and repressors through protein-protein interactions *via* their 14-3-3 domain (Chung *et al.,* 1999). Rice homologs of the strawberry 14-3-3 protein were shown to be capable of interacting with both site-specific DNA binding proteins and tissue specific regulatory factors as part of a transcriptional complex in the ABA response pathway (Schultz *et al.,* 1998). The RAS family of small GTP-binding proteins has been implicated in the transduction of signals from growth factor receptors to signalling cascades. The *Fagus sylvatica* homolog (GenBank accession number X98540) of the strawberry RAS related protein was previously shown to be induced by ABA and accumulated in the embryonic axis of dormant seeds (Nicolas *et al.,* 1998).

Recent evidence implicates cytosolic free Ca^{2+} as a second messenger in the ABA signal transduction cascade in seeds (Leung and Giraudat, 1998). In plant cells, the calcium binding protein calmodulin is considered the primary sensor for changes in cellular free Ca^{2+} levels (Roberts and Harmon, 1992). Ca^{2+} molecules bind to the EF hand binding motifs. Such motifs were identified in the rice *EFA27* gene induced by ABA treatment (Frandsen *et al.,* 1996). A strawberry cDNA showing homology to the *EFA27* gene homolog was found to be 7 fold higher expressed in achenes than in receptacle tissue. Other EF hand containing proteins are ABI1 and ABI2 (Leung *et al.,* 1994). A function in ABA signal transduction pathway was also ascribed to the *Arabidopsis* root-specific *ARSK1* gene, encoding a serine/threonine protein kinase activated upon dehydration, ABA and salt treatments (Hwang and Goodman, 1995).

Ethylene and Strawberry Achene Maturation

The *Arabidopsis ARSK1* gene may be regulated by ethylene through the GCC box sequence in its promoter region, which was shown to be the core sequence of the ethylene-responsive element (ERE) in tobacco (Ohme Takagi and Shinshi, 1995). ERE binding factor (ERF) proteins bind to the GCC box and were proposed to act as transcription factors for stress-responsive genes (Fujimoto *et al.,* 2000). A strawberry cDNA showing homology to the tobacco and *Arabidopsis* ERFs was identified, suggesting a role for ethylene in late achene development. Although the ABA response involves ethylene in many stress related processes, ethylene has also been implicated in the promotion of seed germination, ripening of climacteric fruit, pathogenesis, leaf abscission and flower senescence (Fluhr and Mattoo, 1996). Four additional strawberry cDNAs that were achene associated showed homology to recently identified ethylene responsive genes (*ER24* and *ER6*) from young green tomato fruit (Zegzouti *et al.,* 1999) (Figure 3T). The strawberry *ER24* homolog from tomato was suggested to encode a multi-protein-bridging factor required for transcription initiation. Interestingly the expression of an unknown AP2 domain transcriptional

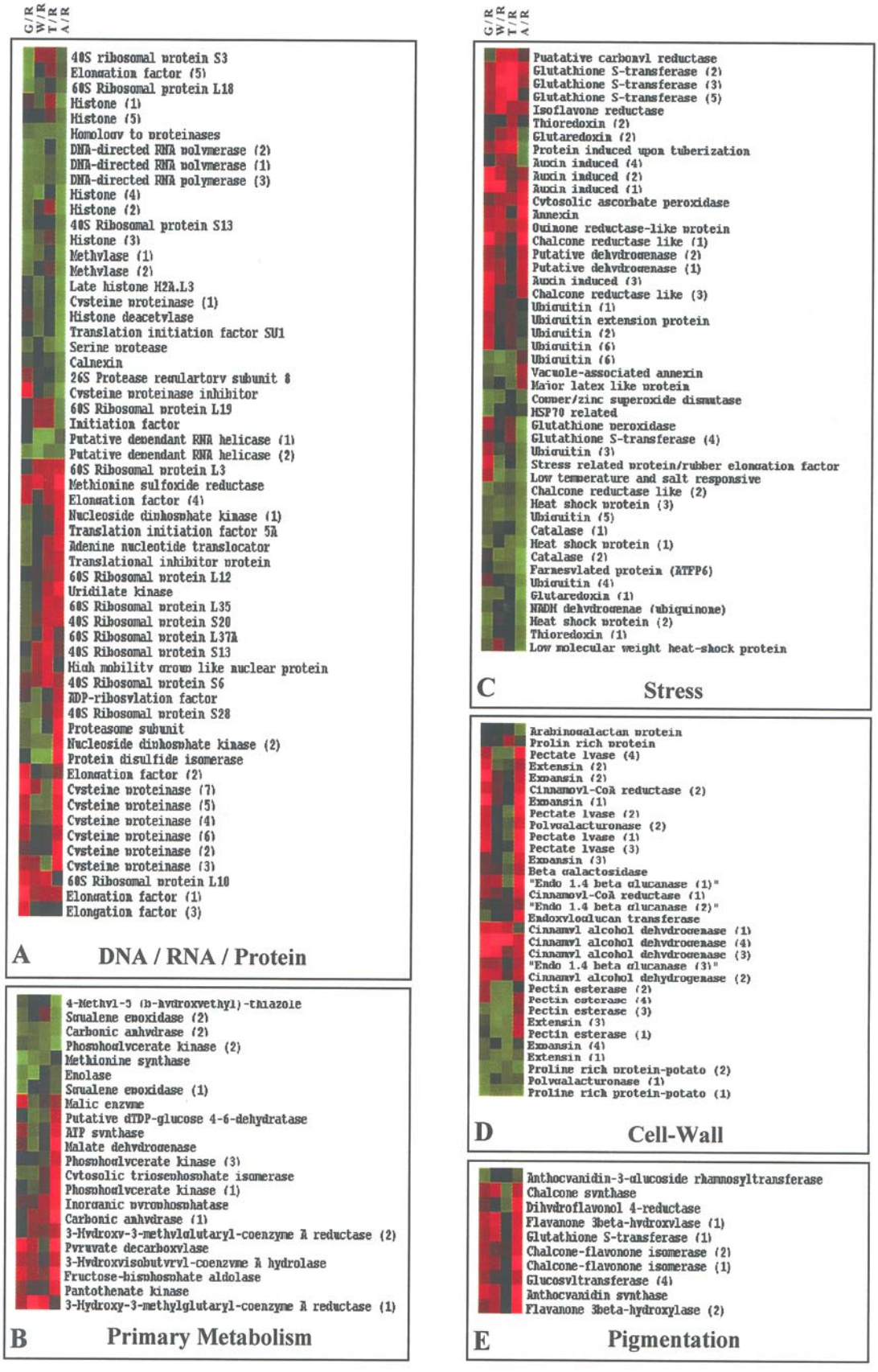
regulator gene (Figure 3U, Table 1A clone D61) was 10 fold higher in achenes than in the receptacle. The AP2 domain shares homology with the DNA binding domain of ERE-binding proteins.

Our microarray studies suggest an important regulatory role for ethylene in the achene maturation phase. The only reported role of ethylene in seed development concerns endosperm cell death during maize kernel or wheat seed development (Young *et al.,* 1997). Recently, the same authors suggested that the balance between ABA and ethylene might regulate the onset and progression of programmed cell death (PCD) in the developing maize endosperm (Young and Gallie, 2000).

The Acquisition of Stress Tolerance in Achenes During Development

The downstream activation of detoxification and protection/defense gene expression is one mechanism used by seeds to gain stress tolerance. A cluster of stress related genes expressed in achenes (Figure 3C) could form part of this protection gene expression program during seed maturation. The identified stress related genes included four different heat shock proteins (HSP), an NADH dehydrogenase (ubiquinone), thioredoxin, glutaredoxins, catalases, a glutathione S-transferase, a copper/zinc superoxide dismutase, a low temperature and salt responsive gene homolog, an aldo/keto reductase (chalcone reductase like), ubiquitins, a stress related protein homolog, a glutathione peroxidase like protein, a metallothionein (see category O) and a farnesylated protein homolog.

Many of these stress-related genes expressed in achenes putatively encode enzymes known to be involved in the metabolism of electrophilic compounds like xenobiotics. These enzymes may play an important role in protecting the seed from peroxidative damage (oxidative stress damage) arising from dehiscence. Phase 1 (transformation) enzymes such as cytochrome p450s induce functional groups onto substrates. Phase 2 (conjugation) enzymes such as glutathione S-transferase (GSTs) utilise the functional groups as a site of further conjugation, usually resulting in less toxic and more water soluble conjugates. Phase 3 (compartmentation) enzymes like ATP dependant pumps recognise and transfer conjugates across membranes for excretion or sequestration. Plant GSTs attach reduced glutathione (GSH) to electrophilic compounds, which tags them for vacuolar sequestration by ATP binding cassette (ABC) transporters. The relative high expression of these xenobiotic-related genes in achenes suggests that achene development coincides with an increase in the level of electrophilic compounds to be neutralised by GSH.


G/R
W/R
T/R
A/R
40S ribosomal protein S3
Elongation factor (5)
60S Ribosomal protein L18
Histone (1)
Histone (5)
Homology to proteinases
DNA-directed RNA polymerase (2)
DNA-directed RNA polymerase (1)
DNA-directed RNA polymerase (3)
Histone (4)
Histone (2)
40S Ribosomal protein S13
Histone (3)
Methylase (1)
Methylase (2)
Late histone H2A.L3
Cysteine proteinase (1)
Histone deacetylase
Translation initiation factor SUI
Serine protease
Calnexin
26S Protease regulartory subunit 8
Cysteine proteinase inhibitor
60S Ribosomal protein L19
Initiation factor
Putative dependant RNA helicase (1)
Putative dependant RNA helicase (2)
60S Ribosomal protein L3
Methionine sulfoxide reductase
Elongation factor (4)
Nucleoside diphosphate kinase (1)
Translation initiation factor 5A
Adenine nucleotide translocator
Translational inhibitor protein
60S Ribosomal protein L12
Uridilate kinase
60S Ribosomal protein L35
40S Ribosomal protein S20
60S Ribosomal protein L37A
40S Ribosomal protein S13
High mobility group like nuclear protein
40S Ribosomal protein S6
ADP-ribosylation factor
40S Ribosomal protein S28
Proteasome subunit
Nucleoside diphosphate kinase (2)
Protein disulfide isomerase
Elongation factor (2)
Cysteine proteinase (7)
Cysteine proteinase (5)
Cysteine proteinase (4)
Cysteine proteinase (6)
Cysteine proteinase (2)
Cysteine proteinase (3)
60S Ribosomal protein L10
Elongation factor (1)
Elongation factor (3)
A
DNA / RNA / Protein
4-Methyl-5 (b-hydroxyethyl)-thiazole
Squalene epoxidase (2)
Carbonic anhydrase (2)
Phosphoglycerate kinase (2)
Methionine synthase
Enolase
Squalene epoxidase (1)
Malic enzyme
Putative dTDP-glucose 4-6-dehydratase
ATP synthase
Malate dehydrogenase
Phosphoglycerate kinase (3)
Cytosolic triosephosphate isomerase
Phosphoglycerate kinase (1)
Inorganic pyrophosphatase
Carbonic anhydrase (1)
3-Hydroxy-3-methylglutaryl-coenzyme A reductase (2)
Pyruvate decarboxylase
3-Hydroxyisobutyryl-coenzyme A hydrolase
Fructose-bisphosphate aldolase
Pantothenate kinase
3-Hydroxy-3-methylglutaryl-coenzyme A reductase (1)
B
Primary Metabolism
G/R
W/R
T/R
A/R
Putative carbonyl reductase
Glutathione S-transferase (2)
Glutathione S-transferase (3)
Glutathione S-transferase (5)
Isoflavone reductase
Thioredoxin (2)
Glutaredoxin (2)
Protein induced upon tuberization
Auxin induced (4)
Auxin induced (2)
Auxin induced (1)
Cytosolic ascorbate peroxidase
Annexin
Quinone reductase-like protein
Chalcone reductase like (1)
Putative dehydrogenase (2)
Putative dehydrogenase (1)
Auxin induced (3)
Chalcone reductase like (3)
Ubiquitin (1)
Ubiquitin extension protein
Ubiquitin (2)
Ubiquitin (6)
Ubiquitin (6)
Vacuole-associated annexin
Major latex like protein
Copper/zinc superoxide dismutase
HSP70 related
Glutathione peroxidase
Glutathione S-transferase (4)
Ubiquitin (3)
Stress related protein/rubber elongation factor
Low temperature and salt responsive
Chalcone reductase like (2)
Heat shock protein (3)
Ubiquitin (5)
Catalase (1)
Heat shock protein (1)
Catalase (2)
Farnesylated protein (ATFP6)
Ubiquitin (4)
Glutaredoxin (1)
NADH dehydrogenae (ubiquinone)
Heat shock protein (2)
Thioredoxin (1)
Low molecular weight heat-shock protein
C
Stress
Arabinogalactan protein
Prolin rich protein
Pectate lyase (4)
Extensin (2)
Expansin (2)
Cinnamoyl-CoA reductase (2)
Expansin (1)
Pectate lyase (2)
Polygalacturonase (2)
Pectate lyase (1)
Pectate lyase (3)
Expansin (3)
Beta galactosidase
"Endo 1.4 beta glucanase (1)"
Cinnamoyl-CoA reductase (1)
"Endo 1.4 beta glucanase (2)"
Endoxyloglucan transferase
Cinnamyl alcohol dehydrogenase (1)
Cinnamyl alcohol dehydrogenase (4)
Cinnamyl alcohol dehydrogenase (3)
"Endo 1.4 beta glucanase (3)"
Cinnamyl alcohol dehydrogenase (2)
Pectin esterase (2)
Pectin esterase (4)
Pectin esterase (3)
Extensin (3)
Pectin esterase (1)
Expansin (4)
Extensin (1)
Proline rich protein-potato (2)
Polygalacturonase (1)
Proline rich protein-potato (1)
D
Cell-Wall
Anthocyanidin-3-glucoside rhamnosyltransferase
Chalcone synthase
Dihydroflavonol 4-reductase
Flavanone 3beta-hydroxylase (1)
Glutathione S-transferase (1)
Chalcone-flavonone isomerase (2)
Chalcone-flavonone isomerase (1)
Glucosyltransferase (4)
Anthocyanidin synthase
Flavanone 3beta-hydroxylase (2)
E
Pigmentation

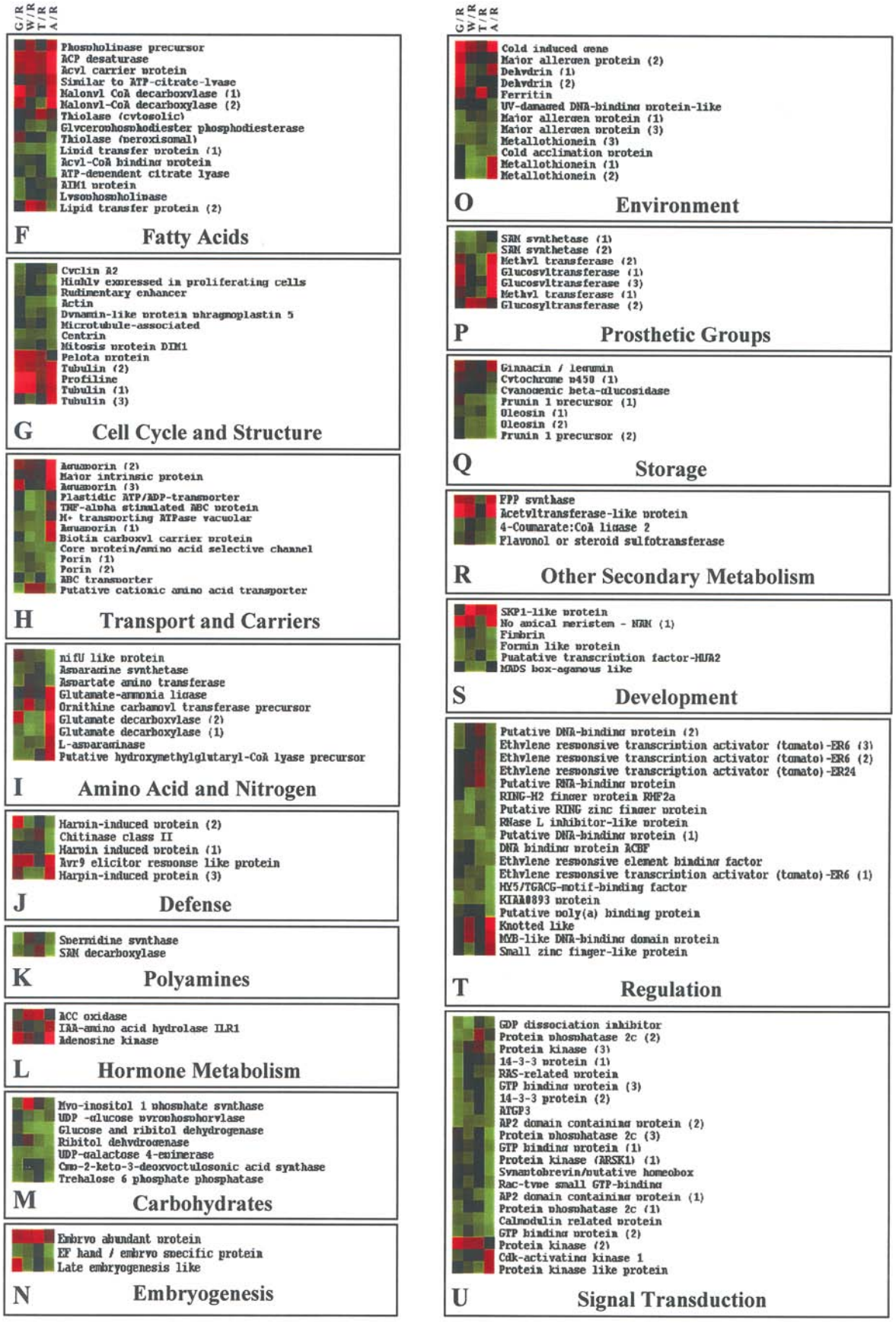

G/R
W/R
T/R
A/R
Phospholipase precursor
ACP desaturase
Acyl carrier protein
Similar to ATP-citrate-lyase
Malonyl CoA decarboxylase (1)
Malonyl-CoA decarboxylase (2)
Thiolase (cytosolic)
Glycerophosphodiester phosphodiesterase
Thiolase (peroxisomal)
Lipid transfer protein (1)
Acyl-CoA binding protein
ATP-dependent citrate lyase
AIM1 protein
Lysophospholipase
Lipid transfer protein (2)
F
Fatty Acids
Cyclin A2
Highly expressed in proliferating cells
Rudimentary enhancer
Actin
Dynamin-like protein phragmoplastin 5
Microtubule-associated
Centrin
Mitosis protein DIM1
Pelota protein
Tubulin (2)
Profiline
Tubulin (1)
Tubulin (3)
G
Cell Cycle and Structure
Aquaporin (2)
Major intrinsic protein
Aquaporin (3)
Plastidic ATP/ADP-transporter
TNF-alpha stimulated ABC protein
H+ transporting ATPase vacuolar
Aquaporin (1)
Biotin carboxyl carrier protein
Core protein/amino acid selective channel
Porin (1)
Porin (2)
ABC transporter
Putative cationic amino acid transporter
H
Transport and Carriers
nifU like protein
Asparagine synthetase
Aspartate amino transferase
Glutamate-ammonia ligase
Ornithine carbamoyl transferase precursor
Glutamate decarboxylase (2)
Glutamate decarboxylase (1)
L-asparaginase
Putative hydroxymethylglutaryl-CoA lyase precursor
I
Amino Acid and Nitrogen
Harpin-induced protein (2)
Chitinase class II
Harpin induced protein (1)
Avr9 elicitor response like protein
Harpin-induced protein (3)
J
Defense
Spermidine synthase
SAM decarboxylase
K
Polyamines
ACC oxidase
IAA-amino acid hydrolase ILR1
Adenosine kinase
L
Hormone Metabolism
Myo-inositol 1 phosphate synthase
UDP -glucose pyrophosphorylase
Glucose and ribitol dehydrogenase
Ribitol dehydrogenase
UDP-galactose 4-epimerase
Cmp-2-keto-3-deoxyoctulosonic acid synthase
Trehalose 6 phosphate phosphatase
M
Carbohydrates
Embryo abundant protein
EF hand / embryo specific protein
Late embryogenesis like
N
Embryogenesis
G/R
W/R
T/R
A/R
Cold induced gene
Major allergen protein (2)
Dehydrin (1)
Dehydrin (2)
Ferritin
UV-damaged DNA-binding protein-like
Major allergen protein (1)
Major allergen protein (3)
Metallothionein (3)
Cold acclimation protein
Metallothionein (1)
Metallothionein (2)
O
Environment
SAM synthetase (1)
SAM synthetase (2)
Methyl transferase (2)
Glucosyltransferase (1)
Glucosyltransferase (3)
Methyl transferase (1)
Glucosyltransferase (2)
P
Prosthetic Groups
Ginnacin / legumin
Cytochrome p450 (1)
Cyanogenic beta-glucosidase
Prunin 1 precursor (1)
Oleosin (1)
Oleosin (2)
Prunin 1 precursor (2)
Q
Storage
FPP synthase
Acetyltransferase-like protein
4-Coumarate:CoA ligase 2
Flavonol or steroid sulfotransferase
R
Other Secondary Metabolism
SKP1-like protein
No apical meristem - NAM (1)
Fimbrin
Formin like protein
Putative transcription factor-HUA2
MADS box-agamous like
S
Development
Putative DNA-binding protein (2)
Ethylene responsive transcription activator (tomato)-ER6 (3)
Ethylene responsive transcription activator (tomato)-ER6 (2)
Ethylene responsive transcription activator (tomato)-ER24
Putative RNA-binding protein
RING-H2 finger protein RHF2a
Putative RING zinc finger protein
RNase L inhibitor-like protein
Putative DNA-binding protein (1)
DNA binding protein ACBF
Ethylene responsive element binding factor
Ethylene responsive transcription activator (tomato)-ER6 (1)
MY5/TGACG-motif-binding factor
KIAA0893 protein
Putative poly(a) binding protein
Knotted like
MYB-like DNA-binding domain protein
Small zinc finger-like protein
T
Regulation
GDP dissociation inhibitor
Protein phosphatase 2c (2)
Protein kinase (3)
14-3-3 protein (1)
RAS-related protein
GTP binding protein (3)
14-3-3 protein (2)
ATGP3
AP2 domain containing protein (2)
Protein phosphatase 2c (3)
GTP binding protein (1)
Protein kinase (ARSK1) (1)
Synaptobrevin/putative homeobox
Rac-type small GTP-binding
AP2 domain containing protein (1)
Protein phosphatase 2c (1)
Calmodulin related protein
GTP binding protein (2)
Protein kinase (2)
Cdk-activating kinase 1
Protein kinase like protein
U
Signal Transduction

Figure 3. Multigenic expression profiles with microarrays. Genes belonging to the same functional category (21 categories in total excluding the "unknown" and "novel" genes) were ordered by using a hierarchical clustering program (Eisen *et al.,* 1999). Each row represents a separate unique cDNA on the microarray (313 unique cDNAs in total) showing differential expression at least once in the 4 microarray experiments (columns). The results presented represent the ratio of hybridization of fluorescent cDNA targets in the four experiments [green/red (GR); white/red (WR); turning/red (TR) and achene/receptacle (A/R)]. These ratios are a measure of relative gene expression in each experimental sample and were depicted according to the colour scale shown. Red colours indicate higher expression in the red stage of development or in receptacle. Green colours indicate higher expression in either green or white or turning stages of development or in achenes. It must be recognised that some of the genes may belong to more than one category. Each cDNA was annotated according to the putative function of its closest NCBI database homologs. The number in parentheses beside part of the putative definitions represent the number of the sequence contig in the case when more then one sequence showed similar BLAST result but did not align in the sequence alignment. These might be for example in the case of gene family members.

Transport and Storage

The presence of an ABC transporter transcript showing elevated abundance in achenes (3.7 fold compared to the receptacle; Figure 3H) suggests that it might be active in transfer of metabolites such as the transport of GSH conjugates. However, this particular strawberry ABC transporter bears homology to a *GCN20* gene which is part of the translation initiation pathway in amino acid starved yeast cells, suggesting that it may encode an amino acid transporter (Vazquez de Aldana *et al.,* 1995). The accumulation of storage compounds (i.e. storage proteins) for seed dormancy and germination is preceded by import of amino acids (e.g. for the supply of building blocks, organic nitrogen and/or for the repair of damaged proteins) and metabolites such as sugars. Both processes appear to be finally co-ordinated at the mRNA level (Hirner *et al.,* 1998). We suspect similar co-regulation in strawberry achenes, as the expression of a putative cationic amino-acid transporter (Figure 3H) correlates with expression of the storage proteins prunin and oleosin (Figure 3Q). The expression of a putative amino acid selective channel protein and two porin homologs (porins allow diffusion of small hydrophilic molecules) might also be part of the intensive process of import into the achenes and storage of amino acids and other metabolites.

Turnover of DNA, RNA and Proteins

The microarray data clearly showed active synthesis of transcripts implicated in turnover of DNA, RNA and proteins in achenes during the green and white stages of fruit development, prior to achene maturation (Figure 3A). This was indicated by the co-ordinate accumulation of transcripts encoding a) methylases, different histone types, and histone deacetylase, b) DNA-directed RNA

polymerases, RNA helicase and c) ribosomal proteins, calnexin, protein translation factor SU1, and proteases. The accumulation of DNA during this phase may be related to endoreduplication or simply storage of deoxynucleotides for the post germination period (Bewley and Black, 1994). The accumulation of RNA and protein may be related to active storage protein accumulation.

Carbohydrate Metabolism

The accumulation of storage proteins, oligosaccharides and of late embryogenesis abundant (LEA) proteins in the seeds prior to or during drying suggests that they may be involved in the protection of seed tissues against the harshness of desiccation (Bewley and Black, 1994). Monosaccharides such as glucose, mannose, fructose and galactose that are predominant in the desiccation intolerant phase are replaced with the disaccharide sucrose and the oligosaccharides raffinose and stachyose when the seed acquires desiccation tolerance. We have identified several strawberry clones encoding putative carbohydrates metabolising enzymes showing elevated expression in early to mid achene development suggesting active metabolism of different carbohydrates such as galactose, glucose, ribitol and surprisingly the disaccharide trehalose. These clones putatively encode UDP-galactose 4-epimerase, trehalose-6-phosphate phosphatase (TPP), glucose and ribitol dehydrogenase, and ribitol dehydrogenase (Figure 3M). The primary plant homologs of the latter two enzymes have been identified in barley embryo (glucose and ribitol dehydrogenase, *pG31*, GenBank accession number S72926) and during development of oilseed rape pods (ribitol dehydrogenase, *SAC25*, GenBank accession number X74225). Interestingly the *pG31* homolog transcript identified from strawberry (Figure 3M, Table 1: clone JB120) was 300 fold more abundant in achene than in receptacle tissue.

In plants, cloning of genes encoding enzymes from the metabolic pathway leading to trehalose was recently reported (Blazquez *et al.,* 1998; Vogel *et al.,* 1998). However, to date there are no studies reporting the synthesis of trehalose in seeds. A wide range of functions have been attributed to trehalose such as tolerance to desiccation, osmotics, temperature and ethanol, control of sugar influx in glycolysis, sugar sensing in plants, and acting as a blood sugar in insects or storage carbohydrate in fungi (Goddijn and van Dun, 1999). The presence of the TPP homolog suggests a role for treholase in the process of maturation and the acquisition of stress and desiccation tolerance in strawberry achenes. Interestingly trehalose-6 phosphate synthase (TPS), which catalyses the formation of trehalose-6-phosphate which in turn is dephosphorylated into trehalose by TPP, contains phosphorylation sites able to interact with 14-3-3 proteins (Goddijn and van Dun, 1999). The co-ordinate up-regulation of genes encoding UDP-galactose 4'-epimerase and UDP-glucose pyrophosphorylase (which did not show a difference in transcript levels between achene and receptacle) involved in UDP-sugar formation and interconversion supports our observation on the importance of carbohydrate metabolism at this stage of achene development. UDP-sugars can serve directly or indirectly as substrates for sugar metabolism (formation of

galactolipids and cell wall polysaccharides), glycolysis and as prosthetic groups for the glycosylation of for instance proteins and phenolic compounds (e.g. flavonoids).

Microarray data can be most useful in providing evidence for the biosynthesis of specific metabolites in a given tissue. Similar to trehalose and polyamine related genes (see Figure 3K) the increase in transcript abundance of genes putatively involved in the metabolism of cyanogenic glycosides may point to their biosynthesis in the achene tissue. Cyanogenic glycosides are an important group of nitrogen containing compounds and are carbohydrate derivatives of cyanohydrins (2-hydroxynitriles). These compounds are widespread in plants and in some instances are a source for HCN which can render a plant toxic. Although known to accumulate in plants of the *Rosaceae* family, they have not been reported in strawberry yet. They are often catabolised to the corresponding aldehyde or ketone and HCN by beta-glucosidases, similar to the cDNA we identified in strawberry (Figure 3Q). In *Sorghum bicolor*, two cytochrome p450 enzymes (CYP79A1 and CYP71E1) mediate the biosynthesis of the cyanogenic glucoside dhurrin (Kahn *et al.,* 1999). The identification of a strawberry cytochrome p450 homolog of *CYP71E1* expressed in achenes may provide a clue to its function in strawberry (Figure 3Q).

Gene Expression Profiling in the Developing Receptacle

Unlike for achene related cDNAs, the majority of cDNAs associated with receptacle show their highest abundance during the red stage of ripening. The green to white transition stage of strawberry development is considered to be the starting point of fruit ripening. At the onset of ripening dramatic changes occur a) textural changes (i.e. cell wall disassembly) b) pigmentation c) production of natural aroma and flavour compounds d) alterations in carbohydrate composition, hormone levels and phenolic constituents and e) assimilation of organic acids.

Enhanced Turnover of DNA, RNA and Protein and Primary Metabolism

Although fruit ripening is a specialised form of plant senescence, it is clearly not a process in which cellular organisation and control are randomly disintegrating (Seymour *et al.,* 1993). Expression of genes related to protein synthesis and turnover indicate that strawberry fruit cells retain a dynamic protein synthesis (i.e. ribosomal proteins, elongation factors), maintenance (protein disulfide isomerase), and repair (methionine sulfoxide reductase) program (Figure 3A). As part of the protein turnover process, damaged and abnormal or non-essential proteins as well as important short-lived regulators are simultaneously removed. Degradation products may be recycled for the generation of nitrogen and other ripening metabolites and de-novo synthesis of proteins. Developmentally regulated degradation may be performed by proteolytic enzymes similar to the strawberry cysteine proteinases showing enhanced expression during ripening or possibly *via* ubiquitin and the 26 proteasome (either by a common pathway or independently)

(Belknap and Garbarino, 1996; Ito *et al.,* 1997). In strawberry, both types of ubiquitin structures (ubiquitin extension protein and polyubiquitin) show increased expression during development from green to red (Figure 3C). However, due to their high sequence conservation we could not distinguish achene and receptacle isoforms. As part of ubiquitin dependent proteolysis, multi-ubiquinated target proteins are selectively degraded to short peptides by the large ATP dependent 26S proteolytic complex. Although we identified a strawberry gene encoding a proteasome subunit with higher expression in receptacle than in achenes (Figure 3A), its expression profile was not similar to any of the identified ubiquitin cDNAs. This suggests that independent protein degradation pathways for ubiquitin as well as for the proteasome complex may exist in strawberry fruit.

Primary metabolism plays a major role in the production of energy and precursors for ripening related processes as for example flavour formation (i.e. amino acids, fatty acids). Out of thirteen different genes putatively encoding enzymes involved in primary metabolism (mainly in gluconeogenesis) and showing elevated expression in the receptacle tissue, only three showed an increased transcript levels during ripening (Figure 3B). These were malic enzyme, fructose-bisphosphate aldolase and pyruvate decarboxylase. Malic enzyme catalyses the reductive decarboxylation of malate to pyruvate. It allows carbon from malate to be fed into the TCA cycle without any requirement for production of pyruvate by glycolysis (Seymour *et al.,* 1993). The reduction in total acidity in the overripe strawberry fruit was previously attributed to the reduction in malic acid levels (Reyes *et al.,* 1982). These data suggest that the increase in transcript levels of the gene encoding malic enzyme may lead to the consumption of malate by respiration of the formed pyruvate and decrease in acidity during ripening. The NAD(P)H produced by the malic enzyme activity, TCA cycle and glycolysis is used for, amongst others, ATP synthesis via oxidative phosphorylation. Pyruvate metabolism might also generate a wide array of other metabolites important in strawberry ripening such as amino acids (i.e. methionine, alanine, valine) and acetaldehyde (reaction catalysed by pyruvate decarboxylase), ethanol and ethyl esters derived thereof.

Fatty Acid Metabolism and Pigmentation

The induction of several fatty acid (Figure 3F) and pigmentation related genes (Figure 3E) provides an example whereby the induction of mRNAs encoding enzymes performing the ultimate steps in the formation of ripening-related metabolites, is co-ordinated with the biosynthesis of their precursors. Fatty acids serve as the initial precursors for several groups of flavour and aroma compounds present in most fruits including strawberry (Schottler and Boland, 1996; Perez *et al.,* 1999). Aliphatic C-6 compounds (aldehydes, alcohols, acids and esters) contributing to strawberry fruit flavour, are formed from unsaturated aliphatic C-18 fatty acids, linoleic ($C_{18:2}$) and linolenic acid ($C_{18:3}$), through the lipoxygenase/hydroperoxide lyase pathway (Croteau and Karp, 1994).

Plant fatty acid unsaturation begins with the conversion of 16:0 acyl carrier protein (ACP) and 18:0-ACP into 16:1-ACP and 18:1-ACP, respectively, by a soluble plastid Δ9 stearoyl-ACP desaturase (Wang *et al.,* 1996). High level expression of the strawberry Δ9 stearoyl-ACP desaturase and an acyl carrier protein (ACP), and for a third gene putatively involved in fatty acid biosynthesis (malonyl-CoA decarboxylase) during the white stage of fruit development was observed (Figure 3F). These expression profiles matched the profile of the strawberry alcohol acyltransferase gene encoding the ester-forming enzyme (acetyltransferase-like protein, Figure 3R), which could implicate these genes in the process of volatile ester formation in strawberry (Aharoni *et al.,* 2000). Interestingly the activities of two other intermediate enzymes in the pathway, lipoxygenase and hydroperoxide lyase were previously reported to increase steadily from the white to the red stage of strawberry fruit development (Perez *et al.,* 1999).

At the onset of ripening there is degradation of early pigments (i.e. chlorophyll) and accumulation of newly synthesised pigments (i.e. anthocyanins). Confidence in the ability of the microarray system to observe co-regulation of an entire pathway is strengthened by the co-ordinated regulation (i.e. similar expression profiles) of six genes involved in the biosynthesis of the anthocyanin pelargonidin-3-glycoside (92% of total pigment in most strawberry varieties) (Perkins Veazie, 1995) (Figure 3E). Genes encoding enzymes devoted to anthocyanin biosynthesis, for example anthocyanidin synthase, showed a dramatic increase (7 fold) during strawberry receptacle development. However, dihydroflavonol 4-reductase which encodes a branch point enzyme involved in the formation of condensed tannins in the early green stage and later in the formation of anthocyanins did not show a significant change in expression patterns during receptacle development.

Cell Wall and the Vascular System

Cell wall loosening in the absence of cell growth is a central process taking place in the receptacle during the third and fourth phase of fruit development. It is now clear that cell wall disassembly is mediated by a concerted and synergistic action of several enzyme families and their unique isoforms (Rose and Bennett, 1999; Brummell and Harpster, 2001). In contrast to previous publications concerning cell wall related enzymes which focus solely on single enzyme families (e.g. endo 1-4 beta glucanase isoforms; see Harpster *et al.,* 1998; Trainotti *et al.,* 1999; Llop Tous *et al.,* 1999) our microarray data has allowed us to correlate the expression profiles of a large collection of transcripts putatively encoding cell wall related enzymes (Figure 3D). Two cell wall modifying enzymes, a pectate lyase (Table 1: clone F71) and an expansin (Table 1: F22), showed 13 fold higher expression in ripening stage receptacle tissue compared with achene tissue. Transcripts for the cell wall modifying enzymes, pectin esterases, show relatively early expression during development. Pectin methylesterase action in de-esterification of pectin may be a prerequisite for the action of the enzyme polygalacturonase (PG) (Brownleader *et al.,* 1999). PG is

a cell wall degrading enzyme that is capable of hydrolysing α(1-4) linkages between adjacent demethylated galacturonic acid residues. Studies with transgenic antisense PG tomato fruit suggested that even very low levels of PG activity may be sufficient to catalyze extensive pectin disassembly (Hadfield and Bennett, 1998). There are conflicting reports concerning PG enzyme presence and activity in strawberry fruit (Huber, 1984; Nogata *et al.,* 1993). Our results demonstrate that there are at least two PG isoforms. One showed decreasing transcript levels from the green to red stage. This correlate well with the observations of Nogata *et al.,* (1993) that PG activity decreases upon strawberry ripening. A second PG showed increasing transcript levels from the white to red stage of fruit development. It is possible that the early PG is responsible for cell wall modifications only during the first peak of fruit growth, which encompasses the green stage. However, the specificity of the early expressed PG mRNA to the receptacle is not clear. The second PG may be active only in the second peak of fruit growth, the white to red stage. The expression profiles of these PG mRNAs correspond with the sigmoid growth curve of strawberry fruit (Miura *et al.,* 1990). Recently, Redondo-Nevado *et al.,* (2001) cloned and characterised a strawberry endo PG gene predominantly expressed at the onset of fruit ripening.

Several of the cell-wall associated genes which show increased transcript levels during ripening are putatively related to a lignification process in the receptacle. In the ripening stage of strawberry fruit development the vascular tissue comprises long fibres composed of cellulose, protein, pectin and lignin (Suutarinen *et al.,* 1998). Thus the different cinnamyl alcohol dehydrogenases (CADs) and the cinnamoyl-CoA reductases clones isolated might be involved in the lignification process in the receptacle. For example, enzymatic activity assays with a recombinant protein encoded by a strawberry CAD gene homolog, identified as ripening regulated was shown to retain cinnamyl alcohol dehydrogenase activity and was immunolocalised to the vascular tissue in the receptacle (Blanco-Portales and Aharoni, unpublished data). Interestingly we observed that at least one-third of the strawberry ripening regulated and receptacle associated transcripts identified by ourselves and others could actually be attributed to the development of the vascular system in strawberry fruit. A detailed description of these genes is provided in a subsequent publication.

Gene Expression Related to Stress Response in the Receptacle

Sustaining ripening related processes in the receptacle requires increased respiration and energy consumption. Oxidative stress conditions arise from ROS that are likely generated as a result of uncontrolled respiration and damaged electron flow in mitochondria (Leprince *et al.,* 2000) leading to the induction of stress- and detoxification-related gene expression. Several oxidative stress-related cDNAs showing enhanced expression during different developmental stages were receptacle associated (Figure 3C and 3O). These transcripts encoded ROS-detoxifying and metabolising enzymes, including auxin induced proteins, cytosolic ascorbate peroxidase, annexin,

quinone reductase-like protein, chalcone reductase-like proteins, putative dehydrogenases, glutathione *S*-transferases, isoflavone reductase-like proteins, and metallothionein. Similar results were obtained by Davies and Robinson (2000) reporting on a group of 10 genes that putatively encoded proteins implicated in stress responses showing an increase in expression during grape berry ripening. The authors suggested that part of the adjustment to the rapid increase in vacuolar sugar levels might be the synthesis of stress proteins.

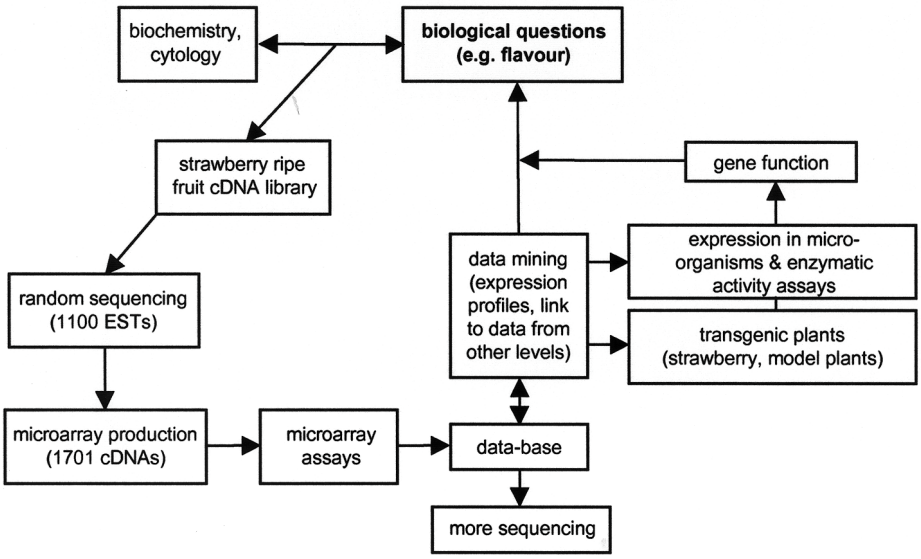


Figure 4 A schematic diagram of the strategy described in this study for the isolation of genes associated with strawberry fruit maturation. The approach combines large-scale cDNA sequencing and a comprehensive gene expression analysis using cDNA microarrays.

Table 1. Top 30 differentially Regulated Genes in Achene and Receptacle

Clone	Fold	Homolog Definition (GenBank accession number)[a]	Functional Category
A		Genes Showing Elevated Expression in the Achene Tissue	
JB120	301	Glucose and ribitol dehydrogenase (AC009177.5)	Carbohydrates
A171	293	Prunin (X78119)	Storage
B176	204	Oleosin (L00935)	Storage
JB59	167	No hit	Novel
JB115	131	No hit	Novel
D74	119	Member of the PF | 00903 glyoxalase family (AC007591)	Unknown
C6	102	KIAA0893 protein (AB020700)	Regulation
G81	74	Protein phosphatase 2c (D38109)	S. trans.[b]
H128	53	Abscisic acid-induced protein HVA (AL035523)	Unknown
C195	35	No hit	Novel
C150	20	Glutaredoxin (Z49699)	Stress
JB122	19	Farnesylated protein ATFP6 (AL035540)	Stress
C13	15	No hit	Novel
SX10	11	Arabidopsis expressed sequence tag (AC007591)	Unknown
D61	10	AP2 domain transcriptional regulator (AC012680)	S. trans.
B		Genes Showing Elevated Expression in the Receptacle Tissue	
E19	28	Chalcone-flavanone isomerase (AF061808)	Pigment
C114	25	Anthocyanidin synthase (X71360)	Pigment
A135	22	Chalcone synthase (D26594)	Pigment
H142	21	Ethylene-forming-enzyme-like dioxygenase (U97530)	Unknown
H109	17	Acyl carrier protein (AF041386)	Fatty acids
E27	16	Acetyltransferase-like protein (AB023041)	Other s. metabolism
H61	16	Flavanone 3-hydroxylase (X69664)	Pigment
B13	14	Legumin like protein (AC007171)	Storage
B109	13	Adenosine kinase (AJ012281)	Hormone metabolism
SX8	13	Putative Aldo/keto reductase (AC007258)	Stress
F71	13	Pectate lyase (U63550)	Cell wall
F22	13	Expansin (AF159563)	Cell wall
JB139	12	No hit	Novel
D75	12	Hypothetical protein (AC016529)	Unknown
F75	12	No hit	Novel

Table 1 Top 15 genes most strongly associated with either achene or receptacle tissues. The 30 most differentially regulated genes pertaining to either the achene (A) or receptacle (B) tissue in red fruit as deduced from the achene/receptacle (A/R) microarray experiment. It must be recognized that some of the genes may belong to more than one functional category.

[a] Definition and accession of nucleotide sequence of the first BLASTX homolog.

[b] S. trans, signal transduction.

CONCLUSION

Combining large-scale random sequencing with gene expression analysis has provided us with a unique and comprehensive overview of transcription relating to putative signal transduction cascades, regulatory pathways, and key metabolic pathways, coinciding in both achene and receptacle tissue, during the development and maturation of strawberry fruit. The information obtained, although correlative in nature, paves the way for a more focused functional analysis in which single genes or closely related groups of genes can be systematically investigated. For example genes identified as related to auxin and ethylene metabolism might be a good starting point for the investigation of hormonal control of non-climacteric fruit. Novel tissue specific genes identified through this study (e.g. glucose and ribitol dehydrogenase, 300 fold more highly expressed in achene than in receptacle tissue) provide new candidates for the identification of strong tissue specific promoters. These promoters can be useful for regulating the temporal and spatial expression of genes controlling important economic traits in transgenic plants. Although non-model plants such as strawberry may currently "suffer" from lack of "whole genome" information compared to model plants, they provide a better platform for elucidating complex biological processes such as the genetic controls governing biosynthesis of specific economically important secondary metabolites.

We believe that the strategy described here will serve as a paradigm for future research projects in strawberry and other fleshy fruit species (Figure 4). The flexible nature of the microarray approach makes it suited for comparative cross hybridisation studies with other fruit with similar genetic background as for example apple and peach. It can also be used for detailed studies of gene expression and patterns in fruit mutants either natural ones or those obtained by chemical/radiation or by insertional mutagenesis (e.g. transposon and T-DNA tagging) and transgenic approaches.

It is anticipated that results arising from this study will in the future be coupled to data provided by other functional genomic tools (proteomics and metabolomics) to assist the generation of multi-component databases linking together sequence, expression, metabolic and mutagenesis data.

METHODS

Scanning Microscopy and Cytology

Strawberry fruit were analysed by scanning electron microscopy, using the cryo stage of a JEOL scanning microscope. Fruits were first frozen in liquid nitrogen and coated with gold by sputtering. For the cytological analysis, fruits were cut into thin slices (less than 2 mm) and fixed with a 1% solution of glutaraldehyde in 0.1 M phosphate buffer, adjusted to pH 5.8. After overnight fixation,

the tissue blocks were dehydrated in an alcohol series of 30%, 50%, 70%, 90% and absolute ethanol, and imbedded in Technovit. After curing of the blocks, sections (5 to 10 μm) were cut on a Reichert microtome (Type 2040). For general observation, the sections were stained with aniline blue 1% (Sigma) in ethanol. Sections were observed and photographed on a Zeiss AXIOPHOT microscope, using Kodak 400 ASA Ektachrome film, using normal illumination with a blue correction filter for the aniline blue stained sections. Calibrations were performed by photographing a micrometer slide at the various magnifications specified.

Plant Material and Preparation of mRNA

For developmental microarray experiments, medium-size green fruits, white fruits with no sign of pigmentation, turning (fruits are partially pigmented) and red ripe stage fruits obtained from the domesticated strawberry (*Fragaria* x *ananassa*) cultivar *Elsanta* were used. Achene and receptacle tissue derived from a red ripe fruit were used for RNA isolation and comparison in the fourth microarray experiment. Total RNA was prepared as described by Schultz *et al.,* (1994). For mRNA preparation, an mRNA purification kit (Pharmacia Biotechnology) was used.

Microarray Experiments

Strawberry microarray production, hybridisation, scanning, data acquisition and statistical analysis were performed as described previously (Aharoni *et al.,* 2000). Briefly, the source of the clones arrayed was a red ripe strawberry fruit tissue cDNA library including the achenes. The library was constructed in the UNI–XR vector (Stratagene La Jolla, CA). Following mass excision, plasmid DNA from 1701 strawberry picked randomly was extracted using the BioROBOT 9600 (Qiagen, Chatsworth, CA). The cDNAs were amplified by polymerase chain reaction (PCR) using the T3 and T7 universal primers using the GeneAmp PCR system 9600 (Perkin Elmer, Foster City, CA). The primers contained a six-carbon amino modification (Isogen Bioscience BV, Maarssen). PCR products were purified using the QIAquick PCR purification kit (Qiagen) and eluted in 100 μL of 0.1 x TE, pH 8.0. Samples were dried to completion, resuspended in 7.5 μL of 5 X SSC (approximately 1 mg/mL) and transferred to a 384-format plate to be subsequently used for spotting. Amplified cDNAs were spotted in duplicate onto silylated microscope slides (CEL Associates, Houston, TX) using a 16-pin print-head and a custom built arraying robot. After arraying, the slides were air dried and stored in the dark. Each of the microarray experiments was performed in a duplicate with the dyes reversed. For the first three experiments green/red, white/red, and turning/red, the threshold ratio for detection (minimum ratio for differential expression) was 2.60, 3.32, and 2.24, respectively. For the microarray experiment comparing achene and receptacle tissues the threshold ratio of detection was 1.97 (in all experiments significant at single test $p<0.05$). The expression ratios for each cDNA determined by the

statistical analysis of each experiment was used for performing cluster analysis using the cluster algorithm of Eisen *et al.,* (1999).

Sequence Analysis

1100 cDNA out of a total of 1701 cDNAs were partially sequenced from the 5' end before performing the microarray experiments. Other non-sequenced cDNAs, which showed differential expression in the microarray experiments, were sequenced using the Applied Biosystems (Foster City, CA) dye terminator cycle sequencing Ready Reaction kit and the 310 DNA sequencer. Comparison analysis of the sequences was conducted with the advanced basic local alignment search tool, BLASTX server (Altschul *et al.,* 1990) and the National Center for Biotechnological Information (www.ncbi.nlm.nih.gov) non-redundant protein database. Software used for DNA and protein analysis was the Geneworks program (IntelliGenetics, Oxford, UK) and the DNASTAR program (DNASTAR Inc. Madison, WI).

ACKNOWLEDGEMENTS

We thank Harrie Verhoeven and Jan Blaas for the contribution of images, Lonneke v.d. Geest and Steven Groot for help with achenes collection and helpful discussions.

REFERENCES

Aharoni A, Keizer LCP, Bouwmeester HJ, Sun ZK, Alvarez Huerta M, Verhoeven HA, Blaas J, van Houwelingen A, De Vos RCH, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O'Connell AP 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *The Plant Cell* 12, 647-61.

Aharoni A, Vorst O. 2002. DNA microarrays for functional plant genomics. *Plant Molecular Biology* 48, 99-118.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-10.

Belknap WR, Garbarino JE. 1996. The role of ubiquitin in plant senescence and stress responses. *Trends in Plant Science* 1, 331-35.

Bewley JD, Black M. 1994. *Seeds: physiology of development and germination.* New York: Plenum Publishing Corporation.

Blazquez MA, Santos E, Flores CL, Martinez Zapater JM, Salinas J, Gancedo C. 1998. Isolation and molecular characterization of the *Arabidopsis* TPS1 gene, encoding trehalose-6-phosphate synthase. *The Plant Journal* 13, 685-89.

Brownleader MD, Jackson P, Mobasheri A, Pantelides AT, Sumar S, Trevan M, Dey PM. 1999. Molecular aspects of cell wall modifications during fruit ripening. *Critical Reviews in Food Science and Nutrition* 39, 149-64.

Brummell DA, Harpster MH. 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. *Plant Molecular Biology* 47, 311-340.

Chung H, Sehnke PC, Ferl RJ, Chung HJ. 1999. The 14-3-3 proteins: cellular regulators of plant metabolism. *Trends in Plant Science* 4, 367-71.

Civello PM, Powell ALT, Sabehat A, Bennett AB. 1999. An expansin gene expressed in ripening strawberry fruit. *Plant Physiology* 121, 1273-79.

Croteau R, Karp F. 1994. Origin of Natural Odorants. In: Muller PM, Lamparsky D eds. *Perfumes : art, science, and technology*. Glasgow: Blackie Academic & Professional, an imprint of Chapman & Hall, 101-26.

Davies C, Robinson SP. 2000. Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterisation of cDNAs encoding putative cell wall and stress response proteins. *Plant Physiology* 122, 803-12.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences, USA* 95, 14863-68.

Finkelstein RR, Wang M, Lynch TJ, Shashirekha R, Goodman HM, Wang ML, Rao S. 1998. The *Arabidopsis* abscisic acid response locus ABI4 encodes an APETALA2 domain protein. *The Plant Cell* 10, 1043-54.

Fluhr R, Mattoo AK. 1996. Ethylene - biosynthesis and perception. *Critical Reviews in Plant Science* 15, 5-6.

Frandsen G, Muller UF, Nielsen M, Mundy J, Skriver K. 1996. Novel plant Ca2+-binding protein expressed in response to abscisic acid and osmotic stress. *The Journal of Biological Chemistry* 271, 343-48.

Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme Takagi M. 2000. *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *The Plant Cell* 12, 393-404.

Gillaspy G, Ben David H, Gruissem W. 1993. Fruits: a developmental perspective. *The Plant Cell* 5, 1439-51.

Given NK, Venis MA, Grierson D. 1988. Hormonal regulation of ripening in the strawberry, a non-climacteric fruit. *Planta* 174, 402-06.

Goddijn OJM, van Dun K. 1999. Trehalose metabolism in plants. *Trends in Plant Science* 4, 315-19.

Gosti F, Beaudoin N, Serizet C, Webb AAR, Vartanian N, Giraudat J. 1999. ABI1 protein phosphatase 2C is a negative regulator of abscisic acid signaling. *The Plant Cell* 11, 1897-909.

Hadfield KA, Bennett AB. 1998. Polygalacturonases: many genes in search of a function. *Plant Physiology* 117, 337-43.

Harpster MH, Brummell DA, Dunsmuir P. 1998. Expression analysis of a ripening-specific, auxin-repressed endo-1, 4-beta-glucanase gene in strawberry. *Plant Physiology* 118, 1307-316.

Harrison EP, McQueen Mason SJ, Manning K. 2001. Expression of six expansin genes in relation to extension activity in developing strawberry fruit. *Journal of Experimental Botany* 52, 1437-46.

Hirner B, Fischer WN, Rentsch D, Kwart M, Frommer WB. 1998. Developmental control of H+/amino acid permease gene expression during seed development of *Arabidopsis*. *The Plant Journal* 14, 535-44.

Huber DJ. 1984. Strawberry fruit softening: The potential roles of polyuronides and hemicelluloses. *Journal of. Food Science* 49, 1310-15.

Hwang I, Goodman HM. 1995. An *Arabidopsis* thaliana root-specific kinase homolog is induced by dehydration, ABA, and NaCl. *The Plant Journal* 8, 37-43.

Ito N, Tomizawa KI, Tanaka K, Matsui M, Kendrick RE, Sato T, Nakagawa H. 1997. Characterization of 26S proteasome alpha- and beta-type and ATPase subunits from spinach and their expression during early stages of seedling development. *Plant Molecular Biology* 34, 307-16.

Kahn RA, Fahrendorf T, Halkier BA, Moller BL. 1999. Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor (L.) Moench*. *Archives of Biochemistry and Biophysics* 363, 9-18.

Kim IJ, Chung WI. 1998. Molecular characterization of a cytosolic ascorbate peroxidase in strawberry fruit. *Plant Science* 133, 69-77.

Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D, Croteau R. 2000. Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proceedings of the National Academy of Sciences*, *USA* 97, 2934-39.

Leprince O, Hoekstra FA, Harren FJM. 2000. Unravelling the responses of metabolism to dehydration points to a role for cytoplasmic viscosity in desiccation tolerance. In: Black M, Bradford KJ, Vazquez Ramos J eds. *Seed biology: advances and applications.* Wallingford; UK: CABI Publishing, 57-66.

Leung J, Bouvier Durand M, Morris PC, Guerrier D, Chefdor F, Giraudat J. 1994. *Arabidopsis* ABA response gene Abi1: features of a calcium-modulated protein phosphatase. *Science* 264, 1448-52.

Leung J, Giraudat J. 1998. Abscisic acid signal transduction. *Annual Review of* Plant *Physiology and Plant Molecular Biology* 49, 199-222.

Leung J, Merlot S, Giraudat J. 1997. The *Arabidopsis* ABSCISIC ACID-INSENSITIVE2 (ABI2) and ABI1 genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. *The Plant Cell* 9, 759-71.

Llop Tous I, Dominguez Puigjaner E, Palomer X, Vendrell M. 1999. Characterization of two divergent endo-beta-1,4-glucanase cDNA clones highly expressed in the nonclimacteric strawberry fruit. *Plant Physiology* 119, 1415-21.

Manning K. 1998. Isolation of a set of ripening-related genes from strawberry: their identification and possible relationship to fruit quality traits. *Planta* 205, 622-31.

Medina Escobar N, Cardenas J, Valpuesta V, Munoz Blanco J, Caballero JL. 1997. Cloning and characterization of cDNAs from genes differentially expressed during the strawberry fruit ripening process by a MAST-PCR-SBDS method. *Analytical Biochemistry* 248, 288-96.

Meyer K, Leube MP, Grill E. 1994. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* 264, 1452-55.

Miura H, Imada S, Yabuuchi S. 1990. Double sigmoid growth curve of strawberry fruit. *Journal of the Japanese Society of Horticultural Science* 59, 527-531.

Nam YW, Tichit L, Leperlier M, Cuerq B, Marty I, Lelievre JM. 1999. Isolation and characterization of mRNAs differentially expressed during ripening of wild strawberry (*Fragaria vesca L.*) fruits. *Plant Molecular Biology* 39, 629-36.

Nicolas C, Nicolas G, Rodriguez D. 1998. Transcripts of a gene, encoding a small GTP-binding protein from *Fagus sylvatica*, are induced by ABA and accumulated in the embryonic axis of dormant seeds. *Plant Molecular Biology* 36, 487-91.

Nogata Y, Ohta H, Voragen AGJ. 1993. Polygalacturonase in strawberry fruit. *Phytochemistry* 34, 617-20.

Ohlrogge, J. and Benning, C. 2000. Unraveling plant metabolism by EST analysis. *Current Opinion in Plant Biology* 3, 224-28.

Ohme Takagi M, Shinshi H. 1995. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *The Plant Cell* 7, 173-82.

Perez AG, Sanz C, Olias R, Olias JM. 1999. Lipoxygenase and hydroperoxide lyase activities in ripening strawberry fruits. *Journal of Agriculture and Food Chemistry* 47, 249-53.

Perkins Veazie P. 1995. Growth and ripening of strawberry fruit. *Horticultural Reviews* 17, 267-97.

Redondo Nevado J, Moyano E, Medina Escobar N, Caballero JL, Munoz Blanco J. 2001. A fruit-specific and developmentally regulated endopolygalacturonase gene from strawberry (*Fragaria x ananassa cv. Chandler*). *Journal of Experimental Botany* 52, 1941-45.

Reyes FGR, Wrolstad RE, Cornwell CJ. 1982. Comparison of enzymic, gas-liquid chromatographic, and high performance liquid chromatographic methods for determining sugars and organic acids in strawberries at three stages of maturity. *Journal of the Association of Analytical Chemistry* 65, 126-31.

Roberts, DM. and Harmon, AC. 1992. Calcium-modulated proteins: targets of intracellular calcium signals in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 43, 375-414.

Rose JKC, Bennett AB. 1999. Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: parallels between cell expansion and fruit ripening. *Trends in Plant Science* 4, 176-83.

Schottler M, Boland W. 1996. Biosynthesis of dodecano-4-lactone in ripening fruits: Crucial role of an epoxide-hydrolase in enantioselective generation of aroma components of the nectarine (*Prunus persica var. nucipersica*) and the strawberry (*Fragaria ananassa*). *Helvetica Chimica Acta* 79, 1488-96.

Schultz DJ, Craig R, Cox-Foster DL, Mumma RO, Medford JI. 1994. RNA isolation from recalcitrant plant tissue. *Plant Molecular Biology Reporter* 12, 310-16.

Schultz TF, Medina J, Hill A, Quatrano RS. 1998. 14-3-3 proteins are part of an abscisic acid-VIVIPAROUS1 (VP1) response complex in the Em promoter and interact with VP1 and EmBP1. *The Plant Cell* 10, 837-47.

Schwab W, Aharoni A, Raab T, Perez AG, Sanz C. 2001. Cytosolic aldolase is a ripening related enzyme in strawberry fruits (*Fragaria x ananassa*). *Phytochemistry* 56, 407-15.

Seymour GB, Taylor JE, Tucker GA. 1993. *Biochemistry of fruit ripening*. London: Chapman & Hall.

Stacy RAP, Nordeng TW, Culianez Macia FA, Aalen RB. 1999. The dormancy-related peroxiredoxin anti-oxidant, PER1, is localized to the nucleus of barley embryo and aleurone cells. *The Plant Journal* 19, 1-8.

Suutarinen J, Anakainen L, Autio K. 1998. Comparison of light microscopy and spatially resolved Fourier transform infrared (FT-IR) microscopy in the examination of cell wall components of strawberries. *Lebensmittel Wissenschaft and Technologielebensm* 31, 595-601.

Trainotti L, Spinello R, Piovan A, Spolaore S, Casadoro G. 2001. beta-galactosidases with a lectin-like domain are expressed in strawberry. *Journal of Experimental Botany* 52, 1635-45.

Trainotti L, Spolaore S, Pavanello A, Baldan B, Casadoro G. 1999. A novel E-type endo-beta-1,4-glucanase with a putative cellulose-binding domain is highly expressed in ripening strawberry fruits. *Plant Molecular Biology* 40, 323-32.

Vazquez de Aldana CR, Marton MJ, Hinnebusch AG. 1995. GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the eIF-2 alpha kinase GCN2 in amino acid-starved cells. *EMBO Journal* 3, 3184-99.

Vogel G, Aeschbacher RA, Muller J, Boller T, Wiemken A. 1998. Trehalose-6-phosphate phosphatases from *Arabidopsis thaliana*: Identification by functional complementation of the yeast tps2 mutant. *The Plant Journal* 13, 673-83.

Wang CL, Chin CK, Ho CT, Hwang CF, Polashock JJ, Martin CE. 1996. Changes of fatty acids and fatty acid-derived flavor compounds by expressing the yeast Delta-9 desaturase gene in tomato. *Journal of Agriculture and Food Chemistry* 44, 3399-402.

Wilkinson JQ, Lanahan MB, Conner TW, Klee HJ. 1995. Identification of mRNAs with enhanced expression in ripening strawberry fruit using polymerase chain reaction differential display. *Plant Molecular Biology* 27, 1097-108.

Young TE, Gallie DR. 2000. Regulation of programmed cell death in maize endosperm by abscisic acid. *Plant Molecular Biology* 42, 397-414.

Young TE, Gallie DR, DeMason DA. 1997. Ethylene-mediated programmed cell death during maize endosperm development of wild-type and shrunken2 genotypes. *Plant Physiology* 115, 737-51.

Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latche, Pech JC, 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, 589-600.

CHAPTER 5

Asaph Aharoni, Leopold C. P. Keizer, Harro J. Bouwmeester, Zhongkui Sun, Mayte Alvarez Huerta, Harrie A. Verhoeven, Jan Blaas, Adèle M. M. L. van Houwelingen, Ric C. H. De Vos, Hilko van der Voet, Ritsert C. Jansen, Monique Guis, Jos Mol, Ronald W. Davis, Mark Schena, Arjen J. van Tunen, and Ann P. O'Connell

Identification of the *SAAT* Gene Involved in Strawberry Flavor Biogenesis by Use of DNA Microarrays

The Plant Cell 12: 647-662 (2000)

IDENTIFICATION OF THE *SAAT* GENE INVOLVED IN STRAWBERRY FLAVOR BIOGENESIS BY USE OF DNA MICROARRAYS

Fruit flavour is a result of a complex mixture of numerous compounds. The formation of these compounds is closely correlated with the metabolic changes occurring during fruit maturation. Here we describe the use of DNA microarrays and appropriate statistical analyses to dissect a complex developmental process. In doing so, we have identified a novel strawberry alcohol acyltransferase (*SAAT*) gene, which plays a crucial role in flavour biogenesis in ripening fruit. Volatile esters are quantitatively and qualitatively the most important compounds providing fruity odors. Biochemical evidence for the involvement of the *SAAT* gene in the formation of fruity esters is provided by the characterisation of the recombinant protein expressed in *Escherichia coli*. The SAAT enzyme showed maximum activity with aliphatic medium chain alcohols, whose corresponding esters are major components of strawberry volatiles. It was capable of utilising short and medium chain, branched and aromatic acyl-coenzyme A molecules as co-substrate. The results suggest that the formation of volatile esters in fruit is subject to the availability of acyl-coenzyme A molecules and alcohol substrates, and it is dictated by the temporal expression pattern of the *SAAT* gene(s) and substrate specificity of the SAAT enzyme(s).

INTRODUCTION

Fruits in addition to their aesthetic qualities form an important part of our diet mainly as a source of energy, vitamins, and minerals and antioxidants. Despite the long history of genetic selection of fruit, we still lack valuable information on the molecular, cellular, and physiological events that control important processes such as flavour formation. Strawberry (*Fragaria* spp) is unusual because what is called the fruit actually originates from the expansion of the flower base (the receptacle) as a pseudocarp, with the real fruits (achenes) on the epidermal layer. During fruit development and maturation, both physical and morphological changes are often a result of changes in protein levels and activities, which may reflect shifts in overall mRNA abundance. The most pronounced changes involve alterations to fruit shape, size, texture, and pigmentation. This coincides with a rise in the soluble solids content and the production of natural aroma and flavour compounds (Perkins-Veazie, 1995).

In recent years, several groups have focused on identifying strawberry genes differentially expressed during ripening (Medina Escobar et al., 1997; Manning, 1998; Nam et al., 1999). The advent of molecular tools, such as cDNA microarray analysis, now adds a new dimension to gene expression studies. This type of analysis provides a powerful means to systematically study the expression profiles of large subsets of genes in a given tissue under specific physiological and environmental conditions. Combining the appropriate biochemical knowledge with gene expression data can provide indirect evidence for the elucidation of gene function. DNA microarray technology has almost exclusively been used to study gene expression in humans and yeast (Schena et al., 1996; DeRisi et al., 1997), with preliminary expression studies reported for the model plant Arabidopsis thaliana (Schena et al., 1995; Ruan et al., 1998). However, none to date has been reported for a commercial crop.

More than 300 compounds have been identified that can contribute to the complex process of aroma biosynthesis in strawberry (for review, see Maarse, 1991). The major components of strawberry flavour and aroma can be grouped into several chemical classes, which include acids, aldehydes, ketones, alcohols, esters, and lactones. Other contributing groups are sulfur compounds, acetals, furans, phenols, epoxides, and hydrocarbons (Zabetakis and Holden, 1997). Esters are one of the most important classes of volatile compounds in fruit flavour, and in strawberry alone more than a hundred esters types have been detected (Honakan and Hirvi, 1990). Esterification is the result of transacylation from acyl-coenzyme A (CoA) to an alcohol (Figure 1). The enzyme catalyzing the reaction is termed an alcohol acyltransferase (AAT), a key enzyme in aroma biochemistry. The influence of esters (isoamyl and ethyl acetate) on beer flavour renders AAT one of the most important enzymes in the fermentation process performed by micro-organisms. As such, it has been the subject of investigations with both yeast and fungi (Yamakawa

et al., 1978; Yamauchi et al., 1989; Yoshioka and Hashimoto, 1984; Malcorps and Dufour, 1992; Fujii et al., 1994, 1996).

In plants, AAT activity has been investigated in both flowers and fruit. Volatile esters are constituents of floral scent. The purification of the acetyl-CoA:benzylalcohol acetyl-transferase (BEAT) protein from flowers of *Clarkia breweri* and the isolation of the gene encoding it has been reported (Dudareva et al., 1998). BEAT has a high affinity for aromatic alcohols, such as benzyl alcohol and cinnamyl alcohol. In fruit, melon, banana, and strawberry, AAT proteins have been investigated using crude fruit extracts (Harada et al., 1985; Ueda et al., 1992; Perez et al., 1993, 1996; Olias et al., 1995). Ueda et al. (1992) concluded that the esters produced by crude strawberry fruit extracts had alcohol moieties reflecting the alcohols predominantly synthesised in the fruit and acid moieties reflecting the acyl-CoA specificity of the AAT enzyme. Olias et al. (1995) compared strawberry and banana proteins possessing AAT activity and found clear differences between the alcohol and acyl-CoA specificity of the two enzymes. The strawberry AAT enzyme had high activity with hexanol and with acetyl- or butyl-CoAs (but slightly less with the latter). The banana enzyme, on the other hand, had high activity with butanol and acetyl-CoA, but showed lower activity with butyl-CoA. A clear correlation could be observed between substrate preference of the enzymes and volatile esters present in both fruits. In this study, we report the cloning and characterisation of a fruit gene encoding an AAT capable of catalyzing the formation of volatile esters in strawberry fruit.

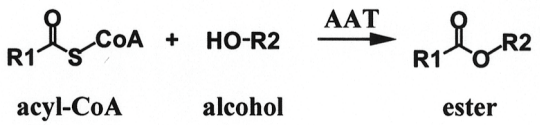


Figure 1 General scheme for the esterification reaction catalyzed by AAT. The enzyme AAT catalyzes the transfer of an acyl moiety from acyl-CoA onto the corresponding alcohol resulting in the formation of an ester.

RESULTS

Expression Analysis using cDNA Microarrays

Microarrays were used to quantitatively examine gene expression during strawberry fruit development. In total, 1701 cDNA clones (probes) from strawberry fruit and 480 cDNA clones from petunia corolla were picked at random from cDNA libraries, amplified by polymerase chain reaction (PCR), and arrayed in duplicate on chemically-modified microscope slides using a robotic printing device. The petunia cDNAs were arrayed together with the strawberry cDNAs to assess the specificity of the hybridization assay in the experiments and to identify genes from strawberry and petunia that are highly conserved at the nucleotide level. Three experiments comparing

strawberry fruit developmental stages green with red, white with red, and turning with red were performed. In each experiment, one mRNA population (target) was labelled with cyanine 3 (Cy3) and the other with cyanine 5 (Cy5). The labelled targets were then mixed and hybridized simultaneously to a microarray. To exclude artefacts, a reciprocal labelling experiment was performed with each pair of targets identical to the first experiment, except that the labels were exchanged. After hybridization, the fluorescence pattern of each microarray was recorded for the Cy3 and Cy5 fluorescent dyes and clones that exhibited differential fluorescence were chosen for further analysis.

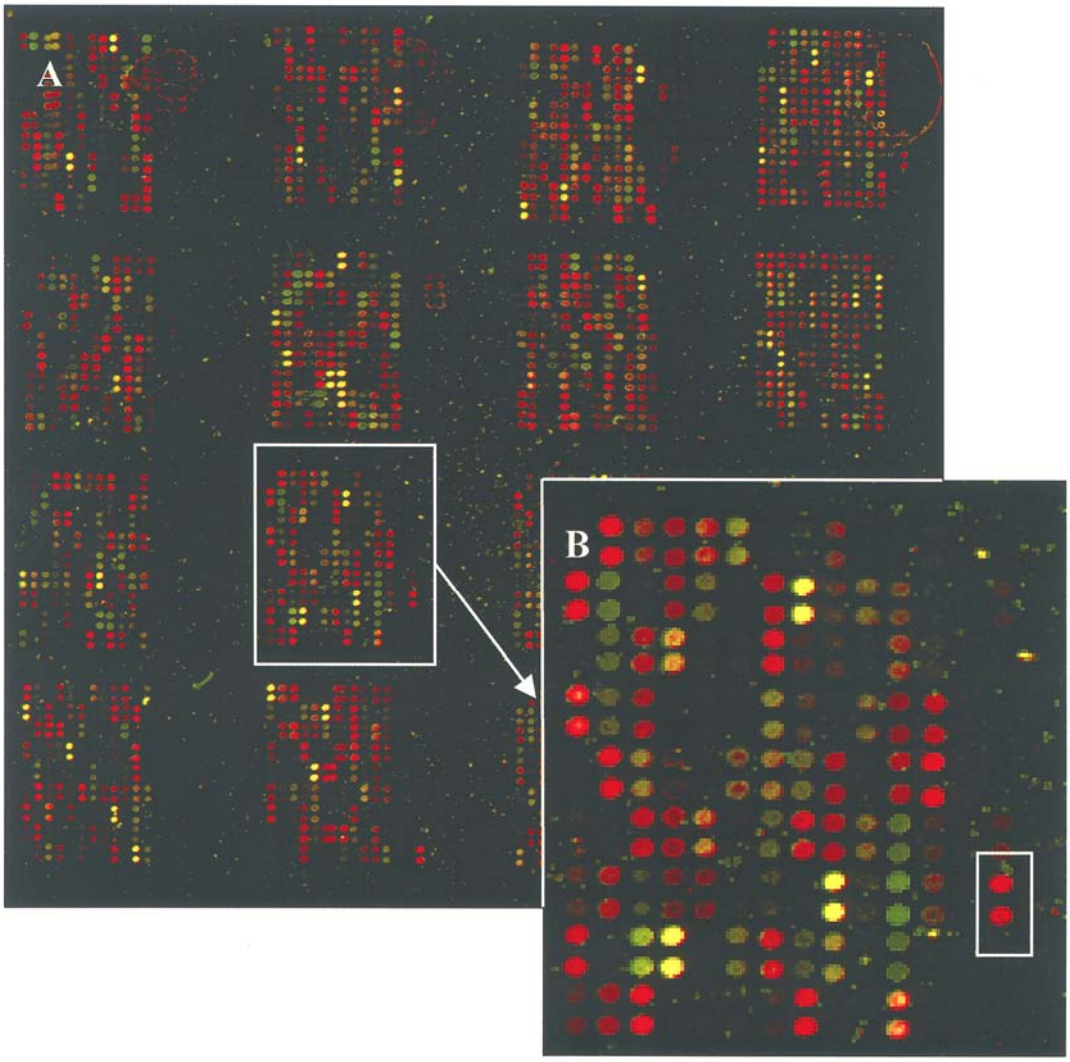


Figure 2 Strawberry and petunia cDNA microarrays. **(A)** The strawberry and petunia cDNAs were spotted in a 4 x 4 format using a 16-pin print head. In each of the 16 sub-arrays the first 12 columns from the left are strawberry probes (total 1701, in duplicate) and the four columns from the right are petunia probes (total 480, in duplicate). The array area is 17 x 17 mm. The image is a two-color overlay obtained with green stage target (fluorecently labeled with Cy5) and red stage target (fluorecently labeled with Cy3) co-hybridized with a single microarray. In the superimposed image the green stage target is represented as a green signal, and the red stage target as a red signal. Signal intensities provide an estimate of expression levels and green or red spot colors, correspond to higher transcript levels in the

green or red stage targets, respectively. Genes with no significant difference in expression between the two stages of development show an intermediate yellow or brown color. **(B)** Enlarged image of the sub-array boxed in (A). Several petunia probes hybridized strongly with targets prepared from strawberry red fruit stage mRNA. One such cDNA clone (boxed) is a member of the highly conserved polyubiquitin gene family.

An image of the microarray following hybridization, comparing the green and red stages is shown in Figure 2A and demonstrates the dynamic range of expression ratios between the probes arrayed. The microarray system exhibits a high specificity to the hybridized target from strawberry. In a few cases, petunia probes produced a strong signal when the array was hybridized with strawberry targets (Figure 2B). Three of the petunia clones showing strong hybridization with strawberry targets (indicated by an intense red signal) were members of the polyubiquitin gene family, which is highly conserved in eukaryotes (approximately 85% homology among unrelated plant species) (Belknap and Garbarino, 1996).

Detection of Differentially Expressed Genes

The expression data were transformed onto a log scale and evaluated by using statistical analysis-of-variance models. Of 1701 strawberry cDNA clones, a total of 247, 168, and 137 cDNAs (significant at single test [per probe] $P<0.05$: 126, 87, and, 76 cDNAs at $P<0.01$; 68, 42 and 27 cDNAs at $P<0.001$) were differentially expressed during the green, white and turning stages, respectively (the same clone can appear at more than one stage) when compared to the red stage. In total, 401 cDNA clones were identified as being differentially expressed. Only expression ratios above a certain value or below the inverse of this value indicate a statistically significant upregulation or down-regulation. For the three experiments green/red, white/red, and turning/red, the threshold ratio for detection was 3.07, 3.32, and 2.24, respectively (at single test with $P<0.05$). The maximum observed ratio was 22.5 in the comparison green/red (Figure 3A). The difference in the number of cDNA clones upregulated in the red stage (green/red [177], white/red [105], and turning/red [60]) and downregulated (green/red [70], white/red [63] and, green/red [77]), in the three experiments can be due to differences in physiological states studied, and the source of the probes (red stage).

Sequence alignment of the cDNA clones identified as differentially expressed by microarray analysis revealed that several shared a high degree of sequence similarity. Of 44 redundant cDNA clones (having at least another one similar cDNA clone represented on the microarray) identified as differentially expressed, two were represented each a total of 27 times. One of these clones (C58) showed homology to a ripening-induced protein (GenBank accession number AJ001445), the other (B7) to a putative metallothionein-like protein (GenBank accession number AJ001444). Both homologs were previously isolated from the wild strawberry (*F. vesca*), and their functions have yet to be determined (Nam et al., 1999). The redundancy of the two cDNAs on the

microarray allowed an evaluation of the precision of the assay (Figure 3B). Expression ratios in the various developmental stages of both groups of 27 similar cDNA clones correlated well. The two families of cDNAs show sequence identity in their overlapping regions.

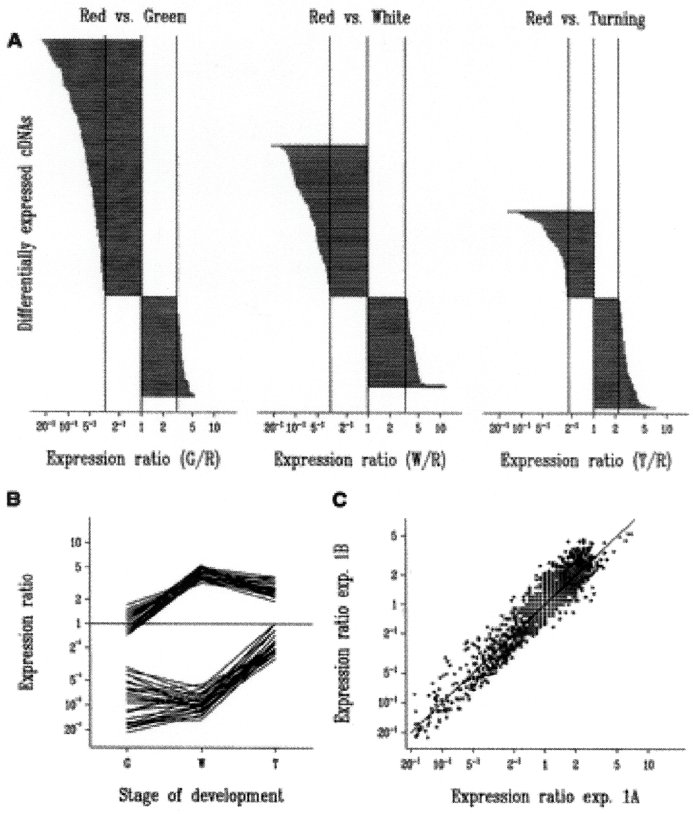


Figure 3 Evaluation of microarray experiments. **(A)** Expression ratios green (G), white (W) and turning (T) stage targets relative to red (R) stage target identified a total of 401 probes corresponding to differentially expressed cDNA clones in the three experiments. The ratios are shown on a logarithmic scale, and are ordered per experiment. The top series are cDNAs upregulated in the R stage (177, 105 and 60 cDNAs, respectively) and the bottom series are cDNAs downregulated in the R stage (70, 63 and 77 cDNAs, respectively). Vertical lines represent the least significant ratios at <0.05 (3.07, 3.32 and 2.24, respectively) and their reciprocals.**(B)** Expression ratios for a series of homologous cDNAs in the green, white and turning stages relative to the red stage (expression ratio marked as 1 for red stage). At top, 27 cDNAs showing homology to a metallothionein gene (including clone B7); at bottom, 27 cDNAs showing homology to an auxin induced gene (including clone C58). The GenBank accession numbers (nucleotide sequence) for strawberry cDNA clones B7 and C58 is AI795160 and AI795161, respectively.**(C)** Precision of the microarray method is illustrated by the scatter plot of 1701 expression ratios estimated in two partial replicates of the green/red experiment. In the first experiment, green fruit target was labeled with Cy5 and red fruit target was labeled with Cy3 and hybridized with a microarray. A second replicate microarray was then hybridized with the dyes reversed (experiment [exp.] designated 1A). In experiment 1B, the first hybridization of experiment 1A was repeated using a third replicate microarray and was analyzed in a "reversed" manner with the results of the second microarray from experiment 1A.

However, variability in signal may be due to variable probe (tethered nucleic acid) or target (free nucleic acid) length and/or genetic redundancy (gene family).

Variability or interarray differences were assessed by examining expression ratios of two microarrays hybridized with fluorescent targets used in one of the two green/red labeling experiments. Figure 3C shows a scatter plot of the expression ratios derived from both microarrays for all of the 1701 cDNAs. The fact that the ratios for all 1701 clones fall along the identity line reveals a low degree of variability between the two separate experiments (the variation between the two partial replicates is characterised by a coefficient of variation in the ratios of 21%). After correction for redundancy, 239 unique differentially expressed cDNA clones were classified on the basis of their BLAST (Altschul et al., 1990) search output. We have adapted a systematic way of categorising expressed sequence tags (ESTs) as described previously (Tamames et al., 1996). The cDNA clones were divided into different categories: information (DNA, RNA, and protein), energy (primary and secondary metabolism), communication (hormonal regulation, detoxification, signal, stress or defense etc.) and unknowns. More than 30% of the cDNA clones were defined as unknowns. These clones either had (1) significant homology to genes with unknown function only or (2) no match or low homology to other database sequences in the BLAST search. The contribution of clones related to energy pathways was striking. In the green/red experiment, of the 177 cDNA clones identified as being upregulated during the red stage, 53% was related to energy, 27% to communication, and 20% to information. A more thorough description of the strawberry microarray experiments (including sequence data) will be documented in a subsequent study.

Identification of the *SAAT* Gene using cDNA Microarray Technology and Detailed Expression Analysis

Fruit color is a clear indication of ripening, which in the majority of fruits is normally accompanied by the accumulation of flavor and aroma components (Seymour at al., 1993). We postulated that cDNA clones showing similar expression profiles to color genes may be involved in flavour formation. One such cDNA clone, *SAAT* (for strawberry AAT), showed sequence similarity to genes encoding enzymes with acyltransfer activity. In fruit from several plant species, as well as in other organisms such as yeast, biochemical evidence suggests that an AAT catalyses the final step in the synthesis of volatile esters (Harada et al., 1985; Ueda et al., 1976). Quantitative microarray expression analysis revealed that *SAAT* had a 16-fold higher expression level during the red stage of fruit development when compared to the green stage (Figures 4A and 4B). Detailed RNA gel blot analysis (Figure 4C) showed that *SAAT* is exclusively expressed in the receptacle tissue (fruit without achenes). *SAAT* expression was first detected during the white stage of fruit development, with maximal expression levels attained between the turning and red stages, as deduced by both microarray and RNA gel blot analyses.

Gas chromatograph–mass spectrometry (GC-MS) analysis of different stages of strawberry fruit development (Figure 5) showed that first detectable sign of release of volatile esters was during the pink stage (between the turning and red stages), with maximal levels attained during the dark red stage. The expression profile of a strawberry cDNA showing homology to an Arabidopsis gene encoding pyruvate decarboxylase (*PDC*) (GenBank accession number U71122) was also investigated, because it may play a role in providing precursors for the formation of ethyl esters (Figures 4B and 4C). The *PDC* gene expression profile correlates well with the *SAAT* expression profile during fruit development. However, *PDC* gene expression was not confined to the receptacle tissue. DNA gel blot analysis suggested that *SAAT* is not a member of a multigene family (data not shown).

Sequence Analysis of *SAAT*

The full-length *SAAT* cDNA clone is 1618 bp, encoding a polypeptide of 452 amino acid residues with a predicted molecular mass of 50.7 kD. Although a sequence homology search revealed low sequence identity (29% identical amino acids) between SAAT and its closest sequence homolog (F21J9.20: Arabidopsis genomic clone: GenBank accession number AC000103), it did contain several conserved sequences which identified it as belonging to a plant super-family of multifunctional acyltransferases (Figure 6). Members of this gene family have a catalytic reaction mechanism related to the ancient chloramphenicol 3-*O*-acetyltransferase (CAT) and dihydrolipoyl *S*-acetyltransferase (DHLAAT) class of enzymes (St Pierre et al., 1998). Several members of this superfamily whose protein sequences show similarity to strawberry SAAT were previously enzymatically characterized and demonstrated to have acyl transfer activity. The *Catharanthus roseus* gene deacetylvindoline 4-*O*-acetyltransfrase (*DAT*; GenBank accession number AF053307) encodes an enzyme catalyzing the biosynthesis of vindoline from acetyl-CoA and deacetylvindoline (St Pierre et al., 1998). Another homolog (19.4% identity at the amino acid level) is the acetyl-CoA:benzylalcohol acetyltransferase gene (*BEAT*; GenBank accession number AF043464) encoding the *Clarkia breweri* enzyme catalyzing the formation of the ester benzyl acetate from benzyl alcohol and acetyl-CoA (Dudareva et al., 1998).

Sequence alignment of the proteins encoded by the above mentioned genes with the SAAT protein sequence illustrates the main consensus sequences shared between these proteins (Figure 6). The H-Xaa-Xaa-Xaa-D (aa; amino acid) motif (corresponding to residues 156-160 in SAAT) is the most conserved consensus sequence present even in acyltransferases such as CAT and DHLAAT (Reed and Hackert, 1990), carnitine and choline acyltransferases (Brown et al., 1994) from non-plants systems. A second highly conserved motif is DFGWG (corresponding to residues 388-392 in SAAT), located near the C terminus. A third consensus sequence is located at the N terminus, L-S-Xaa-T-L-Xaa-Xaa-Xaa-Y-Xaa-Xaa-Xaa-G (corresponding to residues 66-78 in SAAT). This domain is completely identical between the SAAT and BEAT proteins. In addition,

the eight residues (LSETLTLY) of the third domain are present in the green alga *Chlorella vulgaris* acetyl-CoA carboxylase (GenBank accession number BAA57908), carboxyl transferase β subunit.

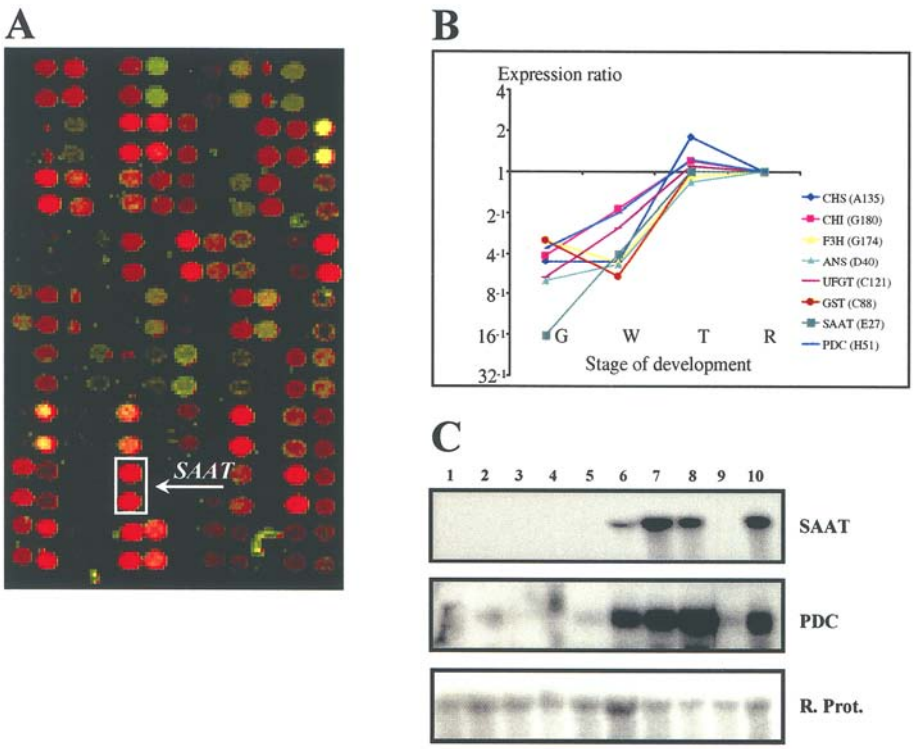


Figure 4 Analysis of *SAAT* and *PDC* gene expression in strawberry. **(A)** An overlay image of a subset of a microarray hybridized with samples originating from green and red strawberry fruit mRNA. As given for Figure 2, signal intensities provide an estimate of expression levels and green or red spot colors, correspond to higher transcript levels in the green or red stage samples, respectively. The *SAAT* cDNA appeared as an intense red signal (boxed) and is marked by an arrow (each cDNA was arrayed in duplicate). The physical size of the sub-array is approximately 4 x 2.5 mm. **(B)** Expression profiles of *SAAT*, *PDC,* and six strawberry pigmentation related genes during strawberry development as detected in the microarray experiments. The strawberry genes and their GenBank accession numbers are as follows: *CHS,* chalcone synthase (AI795154); *CHI*, chalcone flavanone isomerase (AI795155); *F3H, f*lavanone-3-β -hydroxylase (AI795156); *ANS, a*nthocyanidin synthase (AI795157); *UFGT*-UDP- glucose:flavonoid-3-*O*-glucosyltransferase (AI795158); *GST*, glutathione *S*-transferase (AI795159); *SAAT, s*trawberry AAT (AF193789); *PDC, p*yruvate decarboxylase (AF193791). G, green stage; W, white stage; T, turning stage; R, red stage. Expression ratios in green, white, and turning stages are relative to the R stage (expression ratio marked as 1 for R stage). **(C)** RNA gel blot analysis of the expression of *SAAT* and *PDC* in different tissues of strawberry. Lane 1, root; lane 2, petiole; lane 3, leaf; lane 4, flower; lane 5, green fruit; lane 6, white fruit; lane 7, turning fruit; lane 8, red fruit; lane 9, achenes; lane 10, dark red fruit. The blot was hybridized with the *SAAT* probe and then rehybridized with a strawberry cDNA probe showing homology to a gene encoding a ribosomal protein (R. Prot.).

Acetyl-CoA carboxylase is a biotinylated enzyme that catalyses the ATP-dependent formation of malonyl-CoA from acetyl-CoA and bicarbonate. This signifies the importance of this domain in reactions using acetyl-CoA as co-substrate.

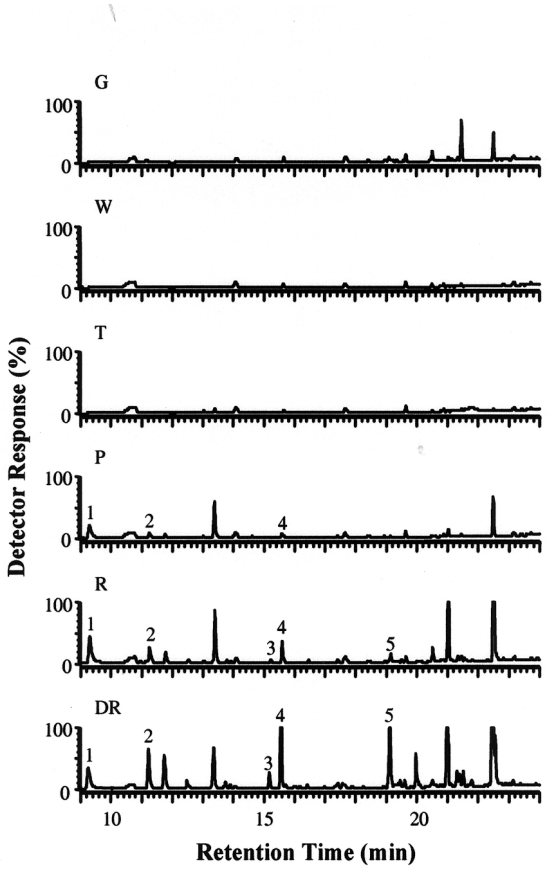


Figure 5 Volatile ester emission during strawberry fruit development. GC-MS chromatograms (detector response: 100% of $2x10^6$ total ion counts) of volatiles in vivo released by strawberry fruits (cv Elsanta) at different stages of development. Developmental stages: G, green; W, white; T, turning; P, pink; R, red; DR, dark red. The five main volatile esters detected are marked with numbers: 1-methyl hexanoate; 2-hexyl acetate; 3-hexyl butanoate; 4-octyl acetate; 5-octyl butyrate.

Functional Expression of the *SAAT* cDNA in *E. coli*.

The entire coding region of the *SAAT* cDNA cloned in frame with a polyhistidine affinity tag (His-tag; see methods) was used for heterologous expression studies in *E. coli*. Protein gel blot analysis revealed a single 58-kD band in the SAAT sample that could not be detected in the control samples (data not shown). The predicted molecular mass of the recombinant SAAT His-Tag protein is 54-kD (50.7 plus 3.8 kD), so the band is within the expected size range. Fractions of the SAAT and control samples eluting from the His-Tag columns were tested for alcohol acyltransferase activity. Acetyl-CoA and butanol were used primarily as substrates for the

analysis. Gas chromatography-mass spectrometry (GC-MS) profiles showed an additional peak arising from the SAAT fraction corresponding to the ester butyl acetate which could not be detected in the control fractions (Figure 7). This confirmed that the formation of butyl-acetate was a direct result of the activity of the SAAT recombinant protein.

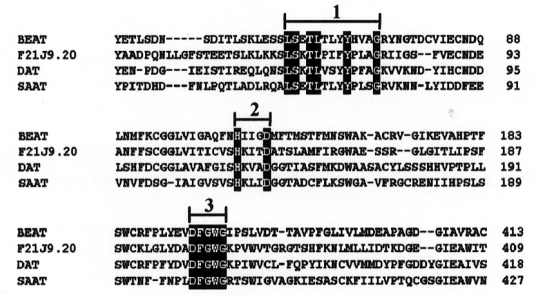


Figure 6 Protein sequence alignment of SAAT and three other related plant proteins. Alignment of SAAT to related protein sequences: F21J9.20, genomic sequence from Arabidopsis bacterial artificial chromosome F21J9, GenBank accession number AC000103; BEAT- *Clarkia breweri* acetyl-CoA:benzylalcohol acetyltransferase, GenBank accession number AF043464; DAT, *Catharanthus roseus* deacetylvindoline 4-*O*-acetyltransferase, GenBank accession number AF053307. The alignments delineate conserved motifs 1, L-S-Xaa-T-L-Xaa-Xaa-Xaa-Y-Xaa-Xaa-Xaa-G; 2, H-Xaa-Xaa-Xaa-D; 3, DFGWG.

Characterization of the SAAT Protein Enzymatic Activity

We assessed the substrate specificity of the strawberry SAAT recombinant protein in vitro by supplying a range of alcohols and acyl-CoAs, and subsequently analyzing the volatiles produced by both radioactivity detector–gas chromatography (radio-GC) and GC-MS. The enzyme activities with different alcohols and acetyl-CoA as co-substrate were assessed using liquid scintillation counting and radio-GC analysis using ^{14}C-acetyl-CoA as co-substrate (Figure 8 and Table 1). Incubations of SAAT with ^{14}C-labelled acetyl-CoA and 1-octanol produced ^{14}C-labeled 1-octylacetate (Figures 8A and 8B) and incubations with 1-hexanol produced ^{14}C-labelled 1-hexylacetate (Figures 8A and 8C).

The enzyme assays were shown to be linear with protein concentration and reaction time for up to 60 min for acetyl-CoA concentrations as low as 0.02 mM (in combination with 20 mM octanol) and hexanol concentrations as low as 2 mM (in combination with 0.1 mM acetyl-CoA). The enzyme exhibited a broad pH range (pH optimum around 8.3). With acetyl-CoA the enzyme accepted a broad range of alcohols as substrate (Table 1). Enzyme activity increased with increasing alcohol carbon chain length up to octanol and then declined. Clear differences in enzyme activity were detected between the four isomers of hexenol tested: *trans*-2-hexenol was a

better substrate than *cis*-2-hexenol, while *cis*-3-hexenol was a better substrate than *trans*-3-hexenol.

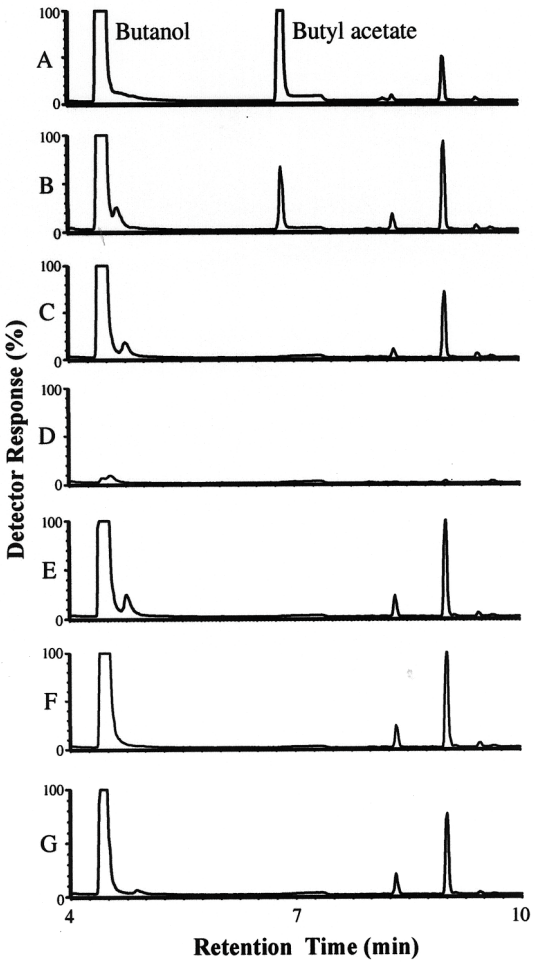


Figure 7 Verification of ester formation by the SAAT protein using GC-MS. GC-MS (detector response, 100% = $2x10^6$ total ion counts) of volatiles produced with the incubation conditions as described below. **(A)** Butanol and butyl acetate standards. **(B)** SAAT protein plus butanol plus acetyl-CoA.**(C)** As given for (B); protein absent. **(D)** As given for (B); butanol absent. **(E)** As given for (B); acetyl-CoA absent. **(F)** GFP protein plus butanol plus acetyl-CoA. **(G)** Empty pRSET B vector elute plus butanol plus acetyl-CoA. Other visible peaks are impurities from the butanol substrate.

The effect of the position of the hydroxy moiety of the alcohol on SAAT activity varied: 1-propanol was a better substrate than 2-propanol, however, 2-butanol was a better substrate than 1-butanol. SAAT also accepted the branched primary alcohol isoamylalcohol. Activity was also detected with aromatic (benzyl- and phenylethyl-) and cyclic (furfuryl-) alcohols though activities were much lower than with 1-octanol (4 to 10%). In contrast, no activity could be detected (both by GC-MS and using liquid scintillation counting) with the terpene alcohol linalool. The ability of

SAAT to use different acyl-CoAs as substrates was qualitatively determined by GC-MS. The SAAT recombinant enzyme was capable of using acyl-CoAs up to C10, branched acyl-CoAs as well as aromatic acyl-CoAs, in combination with 1-propanol or 1-butanol as alcohols (data not shown).

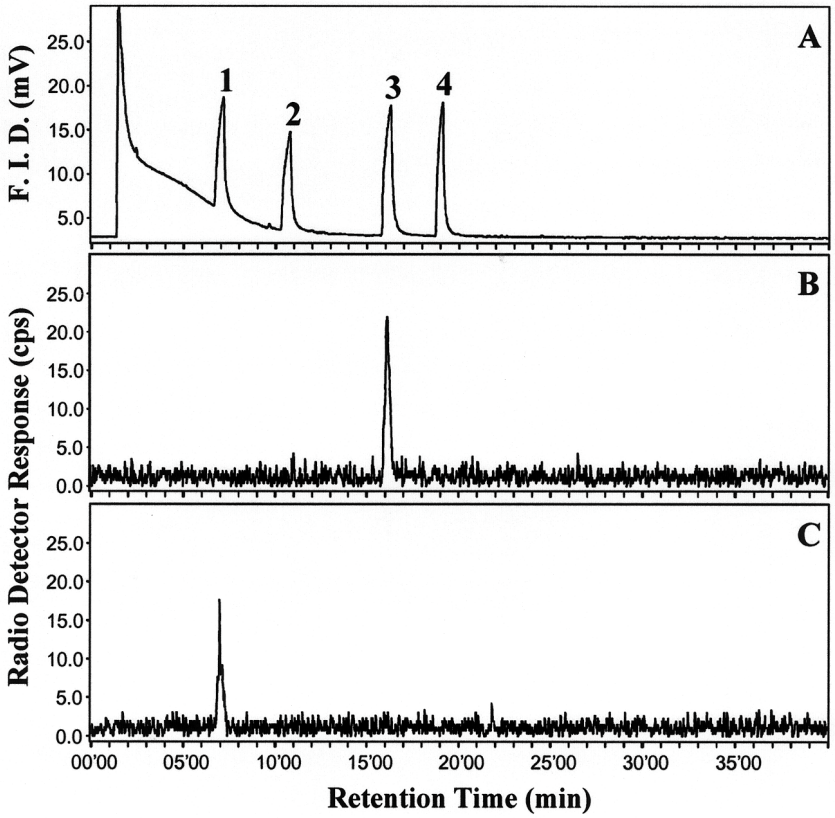


Figure 8 SAAT-catalyzed ester formation from [^{14}C]-acetyl-CoA. **(A)** Flame ionization detector (F.I.D.)-signal of unlabeled standards of peak 1**,** hexylacetate; peak 2, 1-hexanol; peak 3, octylacetate, and peak 4, 1-octanol. **(B)** and **(C)** Radiodetector signal of labeled products formed by the SAAT protein from 0.1 mM ^{14}C-acetyl-CoA and alcohols with **(B)** 2 mM 1-octanol or **(C)** 2 mM 1-hexanol.

The kinetic properties of SAAT were determined for acetyl-CoA and five different alcohols, using liquid scintillation counting. The substrate-activity relationships showed the typical saturation curve (Figure 9). The apparent K_m and V_{max} for acetyl-CoA were 104.2 µM and 4.5 nanomoles of product per hr per µg of protein, respectively (Figure 9A). The apparent values of K_m (in millimoles) and V_{max} (in nanomoles product per hr per µg of protein) for 1-butanol were 46.1 and 0.4, for 1-hexanol 8.9 and 0.6, for 1-octanol 5.7 and 2.6 (Figure 9B), for *trans*-2-hexenol 16.8 and 1.1 and for *cis*-2-hexenol 17.9 and 0.6, respectively (data not shown). The SAAT enzyme characteristics support the data presented in Table 1, where the enzyme activity increases with carbon chain length, from butanol *via* hexanol to octanol (with a concomitant increase in V_{max} and

decrease in K_m values). The difference in activity with the hexenol isomers appears to be caused by differences in V_{max} values rather than affinity.

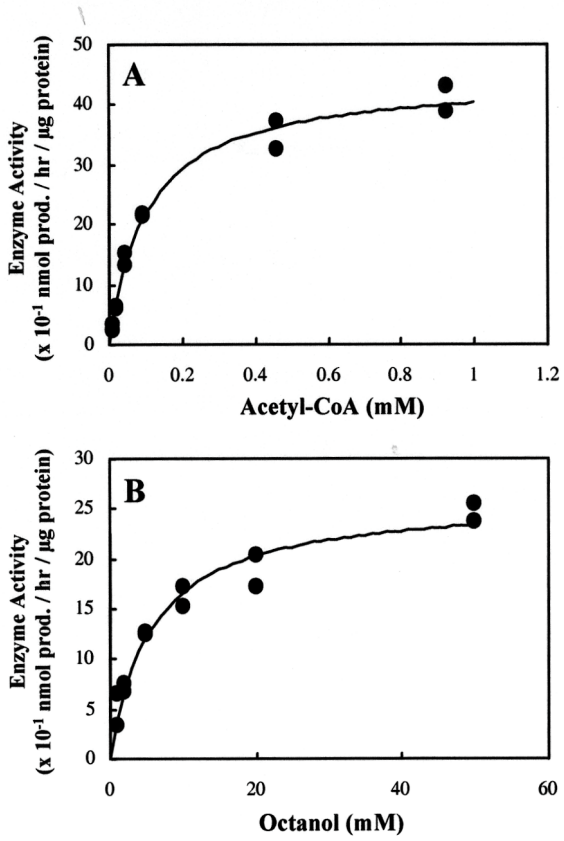


Figure 9 SAAT alcohol acyltransferase activity as a function of substrate concentration for the formation of octyl acetate. **(A)** For acetyl-CoA (in the presence of 20 mM 1-octanol). **(B)** For 1-octanol (in the presence of 0.1 mM acetyl-CoA). Equations for fitted curves are **(A)**, $V = 4.45[S]/(0.1042 + [S])$, $R^2 = 0.99$ and **(B)**, $V = 2.6\ [S]/(5.7 + [S])$, $R^2 = 0.96$. Data were obtained using liquid scintillation counting.

DISCUSSION

Accessing Genetic Information Using cDNA Microarrays

In this study we have reported on the successful application of DNA microarray technology to provide a lead to the identification of a novel flavour gene with subsequent verification of function. This demonstrates the capability of the technology to act as a powerful tool to link gene to function, which will no doubt aid future plant functional genomics research. Statistical analysis of experimental data confirmed the reproducibility of our microarray system. It allows for a 95% reliable detection of 4-5-fold difference in mRNA expression levels with a lower limit of 2-3-fold, as detected in other non-plant systems (Schena et al., 1996; DeRisi et al., 1997).

The use of simultaneous, two-color hybridization schemes increases the precision of the differential measurements by eliminating artefacts associated with comparing separate microarrays (Schena et al., 1995). Ratios obtained on a single microarray are nonetheless generated with two separate scans and therefore artefacts may still arise due to differences in the two channels that are not due to altered mRNA levels. The use of reciprocal labelling schemes, for the two targets in independent experiments helps correct for dye-related differences, sample contaminants, dust, different laser settings, non-linearities of photomultiplier tubes and other artefacts that can erroneously affect the fluorescent read-out of the two channels. Our microarray assays exhibit high specificity to the strawberry targets, as displayed by the lack of hybridization to the petunia clones, with the exception of the highly homologous polyubiquitin cDNA clones that produced a strong signal. Such homologous genes offer the possibility for future syntenic expression analysis to be performed between plant species.

The presence of different types of volatile esters in ripe fruit of many species raises several questions concerning the mechanisms of ester formation and the genetic controls governing the entire process. A strawberry gene (*SAAT*) isolated from the ripe red fruit was shown in the course of this study to encode an enzyme with AAT activity. The evidence for the identification of *SAAT* as a putative alcohol acyl transferase originated from microarray expression profiling data coupled with biochemical data concerning volatile ester release during fruit development. *SAAT* gene expression is fruit-specific commencing at the white stage of fruit ripening, preceding detectable volatile ester formation in strawberry (cv. Elsanta), reaching maximal levels between turning and red stages. This result correlates with AAT activities reported for several commercial strawberry cultivars during fruit development (Perez et al., 1996). The expression pattern supports the proposition that the absence of ester formation in immature stages of fruit ripening is due to a lack of ester producing enzymes (Yamishata et al., 1977). Pyruvate decarboxylase has been proposed to supply precursors for the formation of esters (Zhou et al., 1995) and the fact that its expression profile is similar to that of *SAAT* suggests that both maybe under the same genetic regulation.

Biochemical Evidence for the Involvement of *SAAT* in the Formation of Fruity Esters

Extensive in vitro enzyme assays revealed that a comprehensive array of acyl-CoA molecules and alcohols (both short and long chains with even and uneven carbon number, aromatic, cyclic and aliphatic, branched and un-branched, saturated and unsaturated) may serve as substrates for the SAAT enzyme in vivo. Esters identified by our analyses (1- Butyl acetate, 1- Hexyl acetate, 1-Octyl acetate) as major volatiles emitted in latter stages of fruit ripening were demonstrated to be genuine products of the recombinant SAAT enzyme. Perez et al. (1993, 1996) partially purified an AAT protein from strawberry (cv Chandler) fruits and tested AAT activity in several strawberry cultivars. The enzyme described by Perez et al. showed broad substrate specificity, with a preference for the medium chain alcohol hexanol and a lower activity with lower chain-length

alcohols (activity with ethanol 25% of that with hexanol). In our enzyme assays we also included longer chain alcohols and found an even higher activity with 1-heptanol and 1-octanol, but lower with 1-nonanol and 1-decanol, when compared to 1-hexanol. Perez et al. (1996) reported only 14% activity with 3-hexen-1-ol relative to 1-hexanol. However, it is unclear whether they used the *cis*- or the *trans*-isomer of 3-hexenol. The activity of the SAAT enzyme with *cis*-3-hexen-1-ol and with *trans*-3-hexen-1-ol was 46% and 14%, respectively, when compared with 1-hexanol (Table 1). Thus, the relative activity of AAT with 3-hexenol reported by Perez et al. agrees closely with the relative activity we found for *trans*-3-hexenol.

The SAAT enzyme was also capable of using the 10-carbon decanoyl-CoA as a co-substrate. This ability may account for the presence of this type of long chain esters in strawberry (Honkanen and Hirvi, 1990). Although the acyclic monoterpene alcohol linalool has been identified as a component of volatiles emitted by strawberry cultivar Elsanta (Ulrich et al., 1997), it was not used as substrate by the recombinant SAAT enzyme. This is in agreement with the absence of linalool esters in strawberry volatile profiles including the Elsanta cultivar (Honkanen and Hirvi, 1990).

The K_m value for acetyl-CoA found for SAAT (104 µM) was somewhat higher than that reported by Perez et al. (1996) for a partially purified strawberry AAT (65 µM) and by Harada et al. (1985) for a banana AAT (50 µM). The calculated K_m values of SAAT for some alcohols (especially the value for the best substrate 1-octanol: 5.7 mM) are comparable to the data for butanol (3 mM) reported by Perez et al. (1996). However, they are higher than the 0.4 mM for isoamylalcohol reported by Harada et al. (1985) for the banana AAT. Although our K_m data are comparable to the data for alcohol acyltransferases reported in the literature, they are approximately 1000-fold higher than the K_m values reported for enzymes with a high substrate specificity, such as amorpha-4,11-diene synthase (K_m 0.6 µM; Bouwmeester et al., 1999b) and (-)-alpha-gurjunene synthase (K_m 5.5 µM; Schmidt et al., 1999). Even cytochrome P-450 enzymes that catalyse the hydroxylation of (+)- and (-)-limonene to *trans*-carveol show K_m values between 10 and 20 µM (Karp et al., 1990; Bouwmeester et al., 1999a). All these terpene biosynthetic enzymes with low K_m values show high substrate specificity. Further research is required to find out whether the relatively high K_m values of AATs can be correlated with the relatively low substrate specificity of these enzymes.

The wide range of alcohols and acyl-CoAs accepted by the SAAT enzyme cannot completely explain the differences in ester composition between the developmental stages and different strawberry cultivars. An important factor is the availability of substrates, which may be dependent on the activity of catabolic pathways (e.g. lipid breakdown). Another factor, which may influence the ester composition, is the activity of esterases that perform the reverse reaction to AAT and has been described for fruits, including strawberry (Ueda and Ogata, 1976). Surprisingly, the molecular mass of the partially purified AAT enzyme from strawberry (Perez *et al.*, 1996) was 70 kD, in contrast to the 54 kD estimated for the SAAT recombinant enzyme. Other plant AAT enzymes, such as from banana and *Clarkia breweri* (Harada et al., 1985; Dudareva et al., 1998)

were also estimated to have a lower molecular mass than estimated for the partial purified strawberry enzyme of Perez *et al.* (40 kD and 58 kD respectively). In addition, the different K_m value found for 1-butanol (46.1 versus 3 mM reported by Perez et al., 1996) could indicate that the enzyme purified by Perez et al. is a different AAT than the one described here.

Ancestry of *SAAT*

The presence of several consensus sequences in the SAAT protein sequence assigns it to a super-family of multifunctional proteins responsible for coenzyme A-dependent acyl transfer (St-Pierre et al,. 1998). The H-Xaa-Xaa-Xaa-D motif was initially detected in mammalian systems (Reed and Hackert, 1990; Shaw and Leslie, 1991), and site-directed mutagenesis experiments with the rat carnitine palmitoyltransferase (CPTII) suggested that both the conserved histidine and the conserved aspartate residues are part of the protein catalytic site mediating acyl transfer and cleavage of free CoA (Brown et al., 1994). The AAT enzyme from yeast (ATF1; GenBank accession number D63449) is capable of reacting with acetyl-CoA and with various kinds of short chain alcohols (Yoshioka and Hashimoto, 1984). In spite of the similarity in catalytic activity between SAAT and ATF1, they show very low sequence identity (4% at the amino acid level). This suggests that there are no direct evolutionary relationships between the two enzymes. However, we identified two regions with sequence conservation. One is an H-Xaa-Xaa-Xaa-D motif (residues 191 to 195 in ATF1 and residues 57 to 61 in SAAT), the other a stretch of 12 amino acids located between two tryptophan residues (W-Xaa-Xaa-F-F-X-P-L-Xaa-F-Xaa-W; residues 330 to 341 in ATF1 and residues 380 to 391 in SAAT). The latter conserved sequence overlaps the D-F-G-W-G motif (previously described).

The isolation of other fruit alcohol acyltransferases will allow us to study the significant factors fundamental to the diversity of volatile ester profiles within and between fruit species and will pave the way in the future for natural flavour production. Furthermore, it will provide the means for improving flavour characteristics of commercial fruits lost through successive years of breeding and selection.

Table 1. Substrate specificity of the SAAT recombinant enzyme towards different types of alcohols

Alcohol	Carbon no.			
Methanol	C1:0	Methyl acetate	1.11 ± 0.26	yes
Ethanol	C2:0	Ethyl acetate	0.62 ± 0.10	yes
1-Propanol	C3:0	1-Propyl acetate	2.60 ± 0.26	yes
2-Propanol	C3:0	2-Propyl acetate	1.30 ± 0.15	yes
1-Butanol	C4:0	1-Butyl acetate	2.29 ± 0.25	yes
2-Butanol	C4:0	2-Butyl acetate	3.11 ± 0.02	yes
3-Methyl-1-butanol (isoamylalcohol)	C5:0	3-Methyl-1-butyl acetate (isoamyl acetate)	3.68 ± 0.25	yes
1-Hexanol	C6:0	1-Hexyl acetate	8.44 ± 0.37	yes
cis-2-Hexen-1-ol	C6:1	*cis*-2-Hexenyl acetate	6.05 ± 0.55	no
cis-3-Hexen-1-ol	C6:1	*cis*-3-Hexenyl acetate	4.06 ± 0.14	yes
trans-2-Hexen-1-ol	C6:1	*trans*-2-Hexenyl acetate	9.20 ± 0.65	yes
trans-3-Hexen-1-ol	C6:1	*trans*-3-Hexenyl acetate	1.25 ± 0.07	no
1-Heptanol	C7:0	Heptyl acetate	14.89 ± 4.12	no
1-Octanol	C8:0	1-Octyl acetate	16.36 ± 2.69	yes
1-Nonanol	C9:0	1-Nonyl acetate	14.00 ± 0.11	no
1-Decanol	C10:0	1-Decyl acetate	7.79 ± 0.10	yes
Furfurylalcohol	C5:2	Furfuryl acetate	0.72 ± 0.06	no
Benzylalcohol	C7:3	Benzyl acetate	0.68 ± 0.04	yes
2-Phenylethylalcohol	C8:3	2-Phenylethyl acetate	1.58 ± 0.12	yes
Linalool	C10:2	Linalyl acetate	ND[c]	no

[a] Comparison of esterification activity with different alcohols (20 mM), and using ^{14}C-acetyl-CoA (0.1 mM) as acyl donor. Activity (mean±SD, n = 2) is expressed as 10^{-1} nanomoles of product formed per hour per microgram of enzyme.

[b] The ester product was reported to occur in commercial strawberry varieties (data from Honkanen and Hirvi. 1990).

[c] ND, no detectable activity.

METHODS

RNA Isolation, cDNA Library Construction, Mass Excision and Sequence Analysis

Total RNA was isolated from strawberry (*Fragaria x ananassa*) cultivar Elsanta and from petunia (*Petunia hybrida)* variety W115 as described previously (Van Tunen et al., 1989; Schultz et al., 1994). Red ripe fruit tissue from strawberry, including the achenes, and open corolla tissue from petunia were used for the construction of cDNA libraries. Both cDNA libraries were constructed in the UNI–XR vector (Stratagene La Jolla, CA) and the inserts were directionally cloned. In total, 20×10^3 plaque-forming units from the non-amplified strawberry and petunia libraries were excised (ExAssist/SOLR system, Stratagene). High-quality plasmid DNA from 1701 strawberry and 480 petunia colonies picked randomly was extracted using the Qiagen BioROBOT 9600 (Qiagen, Chatsworth, CA). The 1100 strawberry and 480 petunia cDNAs were partially sequenced

from the 5' end before performing microarray experiments. Sequencing was performed using the Applied Biosystems (Foster City, CA) dye terminator cycle sequencing Ready Reaction kit and the Applied Biosystems 373 and 370A DNA sequencers. Comparison analysis of the sequences was conducted with the advanced basic local alignment search tool, BLAST X server (Altschul et al., 1990) and the National Center for Biotechnological Information (www.ncbi.nlm.nih.gov) non-redundant protein database. Software used for DNA and protein analysis were the Geneworks program (IntelliGenetics, Oxford, UK) and the DNASTAR program (DNASTAR Inc. Madison, WI).

Fluorescent Targets and Probes

Total RNA (up to 1.5 mg) was isolated from strawberry fruits at four selected stages of development: green fruit (medium size green berries), white fruit (no any sign of pigment), turning (half of each berry colored red) and red (full red firm berries). This RNA was used for mRNA preparation with a mRNA purification kit (Pharmacia Biotech). First-strand cDNA was prepared as follows: oligo(dT) 21mer (4 µg) was annealed to 5 µg of mRNA by heating the reaction to 65ºC for 3 min and then by transferring to 25ºC for 10 min. To the reaction (50 µL total), the following components were added (final concentrations or amounts): 1x first strand buffer, 1x DTT (0.1 M), 0.6 units per µL of ribonuclease block, 500 µM of dATP, dGTP and dTTP, 40 µM dCTP, 40 µM cyanine 3 (Cy3) –dCTP or Cy5 – dCTP, and 6 units per µL Superscript II RNase H – reverse transcriptase (Gibco). Reverse transcription was carried out at 37ºC for 2 hr, and samples were precipitated with ethanol and resuspended in 10 µL of TE (10 mM Tris-HCL and 1 mM EDTA, pH 8.0). cDNA/mRNA hybrids were denatured by boiling for 3 min, and the samples were chilled on ice. The RNA was degraded by adding 0.25 µL of 1N NaOH (10 min at 37ºC). Samples were neutralised by the addition of 2.5 µL of 1 M Tris-Cl, pH 6.8, and 2.0 µL of 1 M HCL and precipitated with ethanol. The pelleted cDNA was washed with 80% ethanol, dried, and dissolved in 6.5 µL of double distilled H_2O. After adding 2.5 µL of 20 X SSC (1 X SSC is 0.15 M NaCl and 0.015 M sodium citrate) and 1 µL of 2% SDS, the final concentration of the target was approximately 0.5 µg/µL.

A total of 1701 strawberry and 480 petunia cDNAs were amplified by polymerase chain reaction (PCR) using the T3 and T7 universal primers using the GeneAmp PCR system 9600 (Perkin Elmer, Foster City, CA). Redundant clones were not eliminated (approximately 30%). The primers contained a six-carbon amino modification (Isogen Bioscience BV, Maarssen, The Netherlands) at the 5' end. The PCR reaction (100 µL total volume) was prepared in a 96-well PCR plate (MicroAmp, Perkin Elmer) by mixing the following components: 10 x PCR buffer (15 mM $MgCl_2$), 0.2 mM each deoxynucleotide triosphosphates, 1 µM of each T3 and T7 modified primers, 2.5 units Taq DNA polymerase (Qiagen) and 10 ng of plasmid DNA template. The PCR program used 30 cycles of PCR (30 sec at 94ºC, 30 sec at 55ºC, and 60 sec at 72ºC). PCR products

were purified using the QIAquick PCR purification kit (Qiagen) and eluted in 100 µL of 0.1 x TE, pH 8.0. Samples were dried to completion, resuspended in 7.5 µL of 5 X SSC (approximately 1 mg/mL) and transferred to a 384-format plate to be subsequently used for spotting.

Arraying and Slide Processing

Amplified cDNA sequences were spotted onto silylated microscope slides (CEL Associates, Houston, TX) using a 16-pin printhead and a custom built arraying robot. Strawberry (1701) and petunia (480) cDNA clones were arrayed in duplicate (4362 total probes per array) on 1.7 cm^2 area. Each probe contained 0.5-1.0 ng of PCR product. After arraying, the slides were left overnight to dry. For slide processing before hybridization, slides were soaked twice in 0.2% SDS and twice in double-distilled H_2O successively for 2 min at room temperature with gentle agitation and transferred to 95°C – 100 °C for 2 min (DNA denaturation). After drying the slides for 5 min at room temperature, they were transferred into a sodium borohydrate solution (1 g $NaBH_4$ dissolved in 300 mL of PBS and 100 mL of 100% ethanol) for 5 min, rinsed three times in 0.2% SDS for 1 min and in double-distilled H_2O for 1 min and allowed to dry.

Hybridization, Scanning and Data Acquisition

Two and a half microliters of each fluorescent target was mixed and a total of 5.0 µL was applied to the microarray under a 22 x 22 mm cover slip. The microarray was placed in a custom-made hybridization chamber for 6 hr at 62ºC in a dry hot air incubator. Following hybridization, the array was washed by agitating it in 1 X SSC and 0.1% SDS and 0.1 X SSC and 0.1% SDS successively (5-min wash) at room temperature. The arrays were dried and scanned for fluorescence emission. Arrays were scanned using the ScanArray 3000 (General Scanning, Watertown, MA). A separate scan was conducted for each of the two fluorescent dyes used for hybridization. Using AIS software (Imaging research, Ontario, Canada), the integrated optical density of each individual probe on the array was measured. Extraction of data from each probe was done by using a grid to place a defined circle fitting to the size of the DNA spots and measuring the integrated optical density inside each circle.

Statistical Analysis of Reciprocal Two-Color Microarray Experiments

For statistical modeling, an appropriate scale-of-analysis was chosen in which factors under study act linearly and where uncontrolled error is Gaussian (normally) distributed with constant variance. Preliminary analysis indicated that the fluorescence measurements should be ^{e}log-transformed for proper statistical analysis of variance. The transformed data were averaged over the two neighbouring replicates and then regressed on experimental factors (i.e., namely dye [Cy3

or Cy5], mRNA source for example, green or red), and PCR amplicon (1701 cDNA clones) and on their two-way and three-way interactions, using standard software for analysis-of-variance (Genstat 5 Committee IACR-Rothamsted, UK).

The combinations Cy3-red and Cy5-green have been scored on microarray 1 and Cy5-red and Cy3-green on microarray 2 and the fluorescence values were obtained in four separate laser scans (i.e., statistically the factors dye, tissue, and microarray are [partially] confounded). y_{ijk} denotes the natural logarithm of the fluorescence measurement for dye *i* (with *i*=1, 2 for the dyes Cy3 and Cy5), target *j* (with *j*=1, 2 for green and red), and probe *k* (*k*=1 to 1701), already averaged over the two neighboring replicates. In the following text, dots denote averages, for example $y_{\cdot 1k}$ denotes the average over the two dyes. We used similar definitions for $y_{\cdot 2k}$, $y_{\cdot 1\cdot}$ and $y_{\cdot 2\cdot}$. On the two microarrays $y_{11k} - y_{22k}$ and $y_{21k} - y_{12k}$ are then measures of the upregulation of probe *k* in target green relative to target red. The average difference $y_{\cdot 1k} - y_{\cdot 2k} = \frac{1}{2}[(y_{11k} - y_{22k}) + (y_{21k} - y_{12k})]$ is not influenced by either dye or microarray related variability. However, this measure can be influenced by differences introduced by experimental artefacts specific to any of the two targets. The average expression difference between the two targets, $y_{\cdot 1\cdot} - y_{\cdot 2\cdot}$, originated partly from upregulation and down-regulation of subsets of the 1701 cDNAs. This difference between the targets was also due to various experimental artefacts. We searched for probe-specific relative upregulation or down-regulation differences averaged over the dye conditions. In statistical terms, the two-way target-by-probe interactions are tested against the lowest stratum error, the mean square of residuals mean squares (MS), estimated from the three-way dye-by-target-by-probe variation. Differences ($Diff_k$) were tested per probe (k) by using a standard two-sided *t*-test on $Diff_k \Big/ \sqrt{\frac{1}{2} MS}$, where the factor $\frac{1}{2}$ arises due to the implicit averaging over two fluorescence dyes, at P=0.05 (used in the analysis), P=0.01 and P=0.001 single-test (per probe) significance levels. The exponent of the estimated probe-specific differences on the logarithmic-scale provided the probe-specific cDNA ratios at the original scale. In the case of a genuine 4-5 fold difference, the *t*-test will be significant with a probability of 95% (power of the test).

RNA Gel-Blot Analyses

For RNA gel blot analysis total RNA was isolated from different strawberry tissues as described by Schultz et al. (1994). Ten µg of glyoxal (1.5 M) denaturated total RNA was electrophorized and transferred to a Hybond N^+ membrane (Amersham). After fixation (2 hr at 80°C) blots were hybridized as described by Angenent et al. (1992). The hybridization probes were made by random labelling oligonucleotide priming (Feinberg and Vogelstein, 1984) of the entire cDNAs. Blots were washed two times half an hour in 0.1 X SSC and 0.1% SDS at 65°C.

Analyses of Volatiles Released by Ripening Strawberry Fruits

Volatiles released by strawberry fruits in vivo were sampled using solid-phase microextraction (SPME). Intact fruits of about similar size at different stages of ripening were placed under a glass funnel (diameter 8 cm) closed with aluminum foil, thus creating a headspace. The needle of the SPME- device (Supelco Inc., Bellefonte, PA, USA) was inserted into the funnel through the foil, and volatiles were trapped by exposing a fused silica fiber coated with 100 µm polydimethylsiloxane to the strawberry headspace for 30 min. The SPME-trapped strawberry volatiles were analysed by gas chromatography mass spectrometry, as described by Verhoeven et al. (1997). Volatile compounds were identified by screening the National Institute of Standards and Technology library for comparable mass spectra and by comparison with authentic reference compounds (Aldrich).

Expression of SAAT in Escherichia coli.

The expression vector pRSET B (Invitrogen, Carlsbad, CA), was used for *SAAT* (for strawberry alcohol acyltransferase [AAT]) expression in *E. coli*. (Stratagene, BL21 Gold DE3 strain). The BamHI and HindIII restriction sites of the original pRSET B were primarily used for cloning the green fluorescent protein (*GFP*) coding sequence and this construct as well as the empty pRSET B vector served as controls for the experiments. Cloning the *GFP* gene to the pRSET B vector inserted an additional SalI restriction site at the 3', and together with the BamHI site located at the 5' end of the *GFP* gene served as sites for cloning the *SAAT* cDNA. The *SAAT* coding region was amplified using the forward primer 5'-CGGATCCGGAGAAAATTGAGGTCAG-3' and the reverse primer 5'-CGTCGACCATTGCACGAGCCACATAATC-3'. The primers added BamHI and SalI restriction sites at both ends of the PCR fragments for cloning to the pRSET B vector. The forward primer eliminated the native methionine ATG codon of *SAAT* and formed a fusion protein at the N terminus with a peptide that included an ATG translation initiation codon, a series of six histidine residues (His-Tag) and an Anti-Xpress epitope.

For small scale purifications (50 mL of bacterial culture), the His-Tag protein purification was performed using Ni-NTA spin columns under native conditions according to the manufacturer's instructions (Qiagen). Larger scale purifications (250 mL of bacterial culture) were conducted using B-PER 6XHis columns (Pierce, Rockford, IL). The soluble fraction of the lysate and the elute from the columns were analysed by SDS-PAGE (10%, stained with Coomassie Brilliant Blue R 250) and Western blotting (first antibody was anti-Xpress, Invitrogen and second antibody was GAR-AP, Boehringer). First elutes from the His-Tag columns were used for enzyme activity assays.

Enzymatic Characterisation of the Recombinant SAAT

The AAT activity of the *SAAT*-encoded protein was established in two ways:
(1) For the activity with different alcohols and acetyl-coenzyme A (CoA), ^{14}C-acetyl-CoA was used as substrate and the formation of (radiolabeled) products analysed using radioactivity detector-gas chromatography (radio-GC) and quantified using liquid-scintillation counting. Radio-GC was performed essentially as described previously (Bouwmeester et al., 1999c). Temperature programming was as follows: 70°C for 10 min, 10°C per min to 270°C, and a final time of 10 min. For quantification of the enzyme activity, 1.4 µg of SAAT protein was diluted to 100 µL with 50 mM Tris buffer (pH 8.3, to which ^{14}C-acetyl-CoA (0.1 mM at 0.1 Ci/mol in routine assays) and alcohol (20 mM in routine assays, 2 µL of hexane stock) were added. After incubation for 30 min at 30°C, the assays were left on ice for 15 min, and 700 µL of hexane was added. Assays were vortexed, and centrifuged, and a 600-µL portion of the hexane phase was removed for liquid scintillation counting. The ratio of hexane soluble radioactivity (esters) to the total radioactivity added as acetyl-CoA was used to calculate product formation.

(2) For the utility of different acyl-CoA molecules, assays using 325 µL of buffer (50 mM Tris-HCl, pH 8.0, containing 1 mM DTT), 25 µL of protein elute, and 25 µL of acyl-CoA (4 mM in buffer) were stirred in a glass vial. The enzyme reaction was started by the addition of 25 µL of alcohol (160 mM in buffer). After a 15-min incubation with continuous stirring (35°C), solid $CaCl_2$ was added (final concentration 5 M). Volatiles released into the vial headspace (at 35 °C with stirring) were subsequently trapped for 15 min by exposing a fused silica fiber coated with 100 µm polydimethylsiloxane to the headspace for 30 min. The SPME-trapped volatiles were analysed by gaschromatography-mass spectrometry, as described above.

ACKNOWLEDGEMENTS

We thank Maayan Aharoni for her contribution to data analysis; Nick Mace and Ken Smith for scanning the microarrays in the first experiments; Robert Hall, Michel Ebskamp, and especially Oscar Vorst and Andy Pereira for helpful discussions and comments on the manuscript. We thank Dr. Twan America for his help with bacterial expression. Gerie van der Heijden for his help with image analysis.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**, 403-410.

Angenent, G.C., Busscher, M., Franken, J., Mol, J.N., and van Tunen, A.J. (1992). Differential expression of two MADS box genes in wild-type and mutant petunia flowers. Plant Cell **4**, 983-993.

Belknap, W.R., and Garbarino, J.E. (1996). The role of ubiquitin in plant senescence and stress responses. Trends Plant Sci. **1**, 331-335.

Bouwmeester, H.J., Konings, M.C.J.M., Gershenzon, J., Karp, F., and Croteau, R. (1999a). Cytochrome P-450 dependent (+)-limonene hydroxylation in fruits of caraway. Phytochemistry **50**, 243-248.

Bouwmeester, H.J., Wallaart, T.E., Janssen, M.H.A., van Loo, B., Jansen, B.J.M., Posthumus, M.A., Schmidt, C.O., de Kraker, J.-W., König, W.A., and Franssen, M.C.R. (1999b). Partial purification and characterization of amorpha-4,11-diene synthase. The sesquiterpene synthase catalyzing the first probable step in the biosynthesis of artemisinin. Phytochemistry **52**, 843-854.

Bouwmeester, H.J., Verstappen, F.W.A., Posthumus M.A., and Dicke, M. (1999c). Spider mite-induced (3*S*)-(*E*)-nerolidol synthase activity in cucumber and lima bean. The first dedicated step in acyclic C11-homoterpene biosynthesis. Plant Physiol. **121**, 173-180.

Brown, N.F., Anderson, R.C., Caplan, S.L., Foster, D.W., and McGarry, J.D. (1994). Catalytically important domains of rat carnitine palmitoyltransferase II as determined by site-directed mutagenesis and chemical modification. Evidence for a critical histidine residue. J. Biol. Chem. **269**, 19157-19162.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278**, 680-686.

Dudareva, N., JC, D.A., Nam, K., Raguso, R.A., Pichersky, E., and Nam, K.H. (1998). Acetyl-CoA:benzylalcohol acetyltransferase - an enzyme involved in floral scent production in *Clarkia breweri* Plant J. **14**, 297-304.

Feinberg, A.P., and Vogelstein, B. (1984). "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Addendum. Anal. Biochem. **137**, 266-267.

Fujii, T., Nagasawa, N., Iwamatsu, A., Bogaki, T., Tamai, Y., and Hamachi, M. (1994). Molecular cloning, sequence analysis, and expression of the yeast alcohol acetyltransferase gene. Appl. Environ. Microbiol. **60**, 2786-2792.

Fujii, T., Yoshimoto, H., Nagasawa, N., Bogaki, T., Tamai, Y., and Hamachi, M. (1996). Nucleotide sequences of alcohol acetyltransferase genes from lager brewing yeast, *Saccharomyces carlsbergensis*. Yeast **12**, 593-598.

Harada, M., Ueda, Y., and Iwata, T. (1985). Purification and some properties of alcohol acetyltransferase from banana fruit. Plant Cell Physiol. **26**, 1067-1074.

Honkanen, E., and Hirvi, T. (1990). The flavour of berries. In Food Flavours, I.D. Morton and A.J. Macleod, eds (Amsterdam: Elsevier Scientific Publications), pp. 125-193.

Karp, F., Mihaliak, C.A., Harris, J.L., and Croteau, R. (1990). Monoterpene biosynthesis: specificity of the hydroxylations (-)-limonene by enzyme preparations from peppermint (*Mentha piperita*), spearmint (*Mentha spicata*), and perilla (*Perilla frutescens*) leaves. Arch. Biochem. and Biophys. **276**, 219-226.

Maarse, H. (1991). Volatile Compounds in Foods and Beverages. (New York: Marcel Dekker, Inc.).

Malcorps, P., and Dufour, J.P. (1992). Short-chain and medium-chain aliphatic-ester synthesis in *Saccharomyces cerevisiae*. Eur. J. Biochem. **210**, 1015-1022.

Manning, K. (1998). Isolation of a set of ripening-related genes from strawberry: their identification and possible relationship to fruit quality traits. Planta **205**, 622-631.

Medina Escobar, N., Cardenas, J., Valpuesta, V., Munoz Blanco, J., and Caballero, J.L. (1997). Cloning and characterization of cDNAs from genes differentially expressed during the strawberry fruit ripening process by a MAST-PCR-SBDS method. Anal. Biochem. **248**, 288-296.

Nam, Y.W., Tichit, L., Leperlier, M., Cuerq, B., Marty, I., and Lelievre, J.M. (1999). Isolation and characterization of mRNAs differentially expressed during ripening of wild strawberry (*Fragaria vesca* L.) fruits. Plant Mol. Biol. **39**, 629-636.

Olias, J.M., Sanz, C., Rios, J.J., and Perez, A.G. (1995). Substrate specificity of alcohol acyltransferase from strawberry and banana fruits. In Fruit Flavors. Biogenesis, characterization and authentication, R.L. Rouseff and M.M. Leahy, eds (Washington, DC: American chemical Society), Am. Chem. Soc. Symp. Ser. 596 pp.134-141.

Perez, A.G., Sanz, C., and Olias, J.M. (1993). Partial purification and some properties of alcohol acyltransferase from strawberry fruits. J. Agricult. Food Chem. **41**, 1462-1466.

Perez, A.G., Sanz, C., Olias, R., Rios, J.J., and Olias, J.M. (1996). Evolution of strawberry alcohol acyltransferase activity during fruit development and storage. J. Agricult. Food Chem. **44**, 3286-3290.

Perkins Veazie, P. (1995). Growth and ripening of strawberry fruit. Hort. Rev. **17**, 267-297.

Reed, L.J., and Hackert, M.L. (1990). Structure-function relationships in dihydrolipoamide acyltransferases. J. Biol. Chem. **265**, 8971-8974.

Ruan, Y., Gilmore, J., and Conner, T. (1998). Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. **15**, 821-833.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**, 467-470.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996). Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. USA **93**, 10614-10619.

Schmidt, C.O., Bouwmeester, H.J., Bülow, N., and König, W.A. (1999). Isolation, Characterization and mechanistic studies of (-)-alpha-gurjunene synthase from *Solidago canadensi*s. Arch. Biochem. Biophys. **364**, 167-177.

Schultz, D.J., Craig, R., Cox-Foster, D.L., Mumma, R.O., and Medford, J.I. (1994). RNA isolation from recalcitrant plant tissue. Plant Mol. Biol. Rep. **12**, 310-316.

Seymor, G. B., Taylor, J. E., and Tucker, G. A. (1993). Biochemistry of Fruit Ripening (Chapman & Hall, London).

Shaw, W.V., and Leslie, A.G. (1991). Chloramphenicol acetyltransferase. Annu. Rev. Biophys. Biophys. Chem. **20**, 363-386.

St Pierre, B., Laflamme, P., Alarco, A.M., Luca, V.d., and De Luca, V. (1998). The terminal *O*-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. Plant J. **14**, 703-713.

Tamames, J., Ouzounis, C., Sander, C., and Valencia, A. (1996). Genomes with distinct function composition. FEBS Lett. **389**, 96-101.

Van Tunen, A.J., Hartman, S.A., Mur, L.A., and Mol, J.N.M. (1989). Regulation of chalcone flavanone isomerase (*CHI*) gene expression in *Petunia hybrida*: the use of alternative promoters in corolla, anthers and pollen. Plant Mol. Biol. **12**, 539-551.

Ueda, Y., and Ogata, K. (1976). Studies of biosynthesis of esters in fruit volatiles. II. Esterification of added alcohol in separated cells from banana, strawberry and melon. J. Food Sci. Tech. Tokyo, **23**, 288-294.

Ueda, Y., Tsuda, A., Bai, J.H., Fujishita, N., and Chachin, K. (1992). Characteristic pattern of aroma ester formation from banana, melon, and strawberry with reference to the substrate specificity of ester synthetase and alcohol contents in pulp. J. Jpn. Soc. Food Sci. Technol. **39**, 183-187.

Ulrich, D., Hoberg, E., Rapp, A., and Kecke, S. (1997). Analysis of strawberry flavour - discrimination of aroma types by quantification of volatile compounds. Z. Lebensm. Unters. For. **205**, 218-223.

Verhoeven, H., Beuerle, T., and Schwab, W. (1997). Solid-phase microextraction: Artifact formation and its avoidance. Chromatographia **46**, 63-66.

Yamakawa, Y., Goto, S., and Yokotsuka, I. (1978). Fractionation and some properties of acetic-ester synthesizing enzyme from *Cladosporium cladosporioides* No.9. Agr. Biol. Chem. **42**, 269-274.

Yamauchi, H., Hasuo, T., Amachi, T., Akita, O., Hara, S., and Yoshizawa, K. (1989). Purification and characterization of acyl coenzyme A: Alcohol acyltransferase of *Neurospora* sp. Agr. Biol. Chem. **53**, 1551-1556.

Yamashita, I., Iino, K., Nemoto, Y., and Yoshikawa, S. (1977). Studies on flavor development in strawberries. 4. Biosynthesis of volatile alcohol and esters from aldehyde during ripening. J. Agricult. Food Chem. **25**, 1165-1168.

Yoshioka, K., and Hashimoto, N. (1984). Ester formation by brewers' yeast during sugar fermentation. Agric Biol Chem. **45**, 333-340.

Zabetakis, I., and Holden, M.A. (1997). Strawberry flavour: analysis and biosynthesis. J. Sci. Food Agr. **74**, 421-434.

Zhou, L., Ke, D., Kader, A.A., Zhou, L.L., and Ke, D.Y. (1995). Mode of controlled atmosphere action on ester biosynthesis of strawberries. Acta Agr. Universitatis Pekinensis **21**, 180-186.

CHAPTER 6

Asaph Aharoni, Ashok P. Giri, Francel W.A. Verstappen, Robert Sevenier, Harrie A. Verhoeven, Gerard H. Rouwendaal, Maarten A. Jongsma, Wilfried Schwab & Harro J. Bouwmeester

Gain and Loss of Flavour and Aroma Compounds During Crop Domestication

Submitted for publication

GAIN AND LOSS OF FLAVOUR AND AROMA COMPOUNDS DURING CROP DOMESTICATION

The blend of flavour compounds produced by fruits serve as "biological perfumes", used for the attraction of living creatures including human beings[1,2,3]. These volatile mixes include hundreds of metabolites, varying in their composition in different characteristic fruit flavours[4]. Surprisingly, the same compounds might often be involved in the protection of the fruit against spoilage by pathogens and insects[5,6]. The mechanism by which natural plant compounds such as flavours are gained and lost during evolution and domestication are largely unknown. Here, we reveal a process for the evolution of diversity in strawberry fruit flavour components based on a change in enzyme localization and alteration in gene expression profile. Through a change in subcellular localization the enzymes encountered new substrates and produced novel metabolites characteristic of the polyploid strawberry cultivars of today. We also identified the opposite situation in which the capacity to synthesise flavour molecules was lost during years of strawberry breeding. The results signify how products of evolution and natural diversity are created and selected during domestication of a crop species.

An array of natural plant products known as terpenoids contribute to the aroma of herbs such as mint[7], the scent of flowers like roses[8] and the flavour and aroma of fruit species including strawberry[9]. We detected a remarkable difference in terpene emission between the wild, diploid strawberry (*Fragaria x vesca*) and the octaploid, cultivated species (*Fragaria x ananassa*). The monoterpene linalool and the sesquiterpene nerolidol are the only compounds detected in the cultivated variety, whereas wild strawberry fruits emit only olefinic monoterpenes not detected in the cultivated species (Fig. 1a). Mono- and sesquiterpenes are synthesised by the action of terpene synthases from geranyl diphosphate (GDP) and farnesyl diphosphate (FDP) respectively[10] (Fig. 1b). Our aim was to unravel the molecular mechanisms which have altered the flavour profile of crops during domestication such as the loss of olefinic monoterpenes and gain of linalool and nerolidol in the cultivated strawberry.

Gene expression profiling using microarray technology identified a ripening-induced putative terpene synthase gene (*FaNES1*) that was expressed in the receptacle tissue of the cultivated strawberry[11] (Fig. 1c and 1d), but was barely detectable in the wild species (Fig. 1e). Analysis of the 5' end of the *FaNES1* cDNA revealed that the encoded protein contains two methionine residues (Met1 and Met2), located in frame and 31 amino acid residues apart, but with a stop-

codon in between (Fig. 2). Sequence analysis of PCR amplified genomic DNA fragments between Met1 and Met2 revealed several additional fragments from both strawberry species. Some of these were longer and were not interrupted by a stop-codon; others were shorter and contained a stop codon similar to the original *FaNES1* clone. The longer fragments contained the tandem arginine motif (RR), typical of a plastid targeting signal in monoterpene synthases[12] (Fig. 2a). In addition to *FaNES1*, we isolated the full-length cDNAs corresponding to the longer fragments from the cultivated and wild strawberries (*FaNES2* and *FvNES1*, respectively) using RACE-PCR (Fig. 2). Despite the clear difference at the N-termini, the proteins showed a high sequence conservation (95% identity).

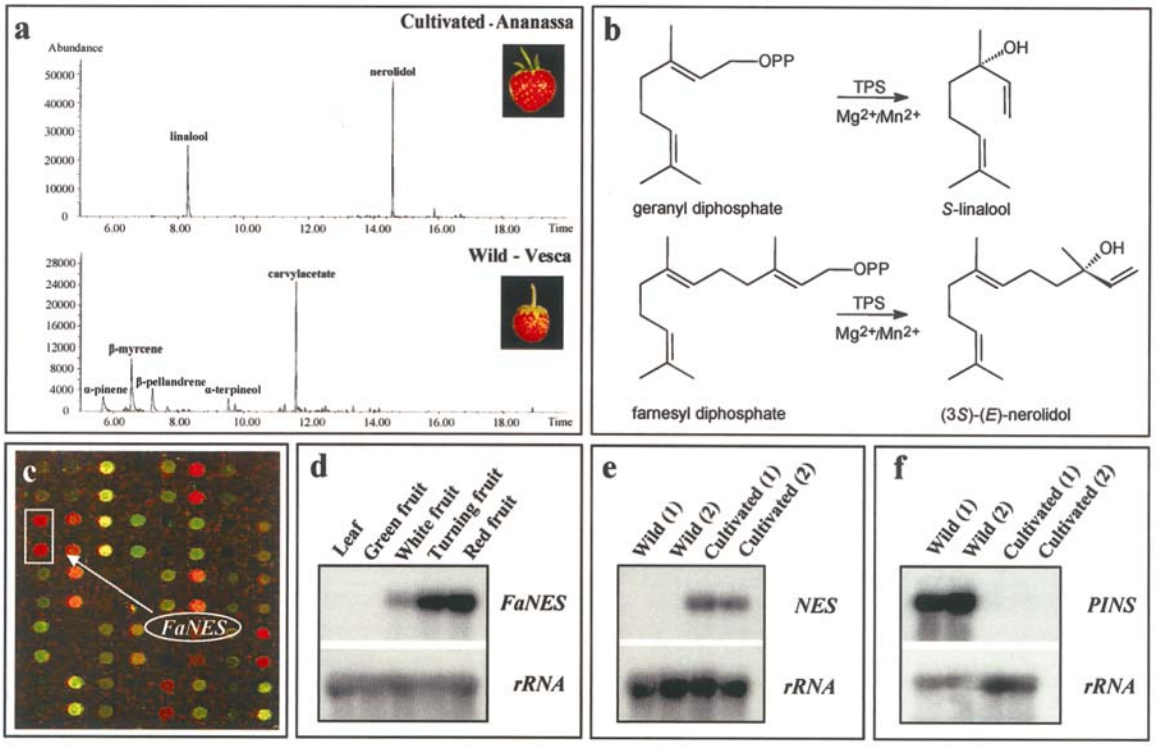


Figure 1 Terpenoid production and gene expression in wild and cultivated strawberry species. **a**, Terpenoids detected by headspace analysis of ripe fruits. **b**, Reactions catalyzed by terpene synthases (TPS) for the formation of the monoterpene alcohol linalool and the sesquiterpene alcohol nerolidol. **c**, Detection of *FaNES1* as receptacle associated using cDNA microarrays. Red and green signals represent higher gene expression in the receptacle and achene ("seeds") tissues, respectively (yellow signal is similar levels in both tissues). **d,** *FaNES* expression in tissues of the cultivated strawberry detected by RNA gel-blot. rRNA, ribosomal RNA. **e**, and **f,** Mirror images of *NES* and *PINS* expression in ripe fruits of wild and cultivated strawberry species detected by RNA gel-blots.

Heterologous expression in *E.coli* of *FaNES1* and *FaNES2* starting from either Met1, the RR motif (in *FaNES2*), or Met2 showed that in all cases the recombinant proteins converted both GDP to linalool as well as FDP to nerolidol. No activity could be detected with any of the recombinant proteins produced from fragments of the *FvNES1* cDNA. The 5' end fragment corresponding to the *FaNES1* gene was identified in more than 70 cultivated varieties (Fig. 4a) and this correlated in

all cases to the production of nerolidol and linalool (data not shown). On the other hand, neither this fragment nor nerolidol or linalool could be detected in any of the twelve wild varieties examined (Fig. 4a). RT-PCR experiments showed that *FaNES1* was the only gene in the cultivated species showing detectable expression in the ripe fruits (Fig. 4b). Thus, *FaNES1* is expressed at high levels but contains a deletion in its 5' region, whereas *FvNES1* and *FaNES2* are only expressed at very low levels in both wild and cultivated species.

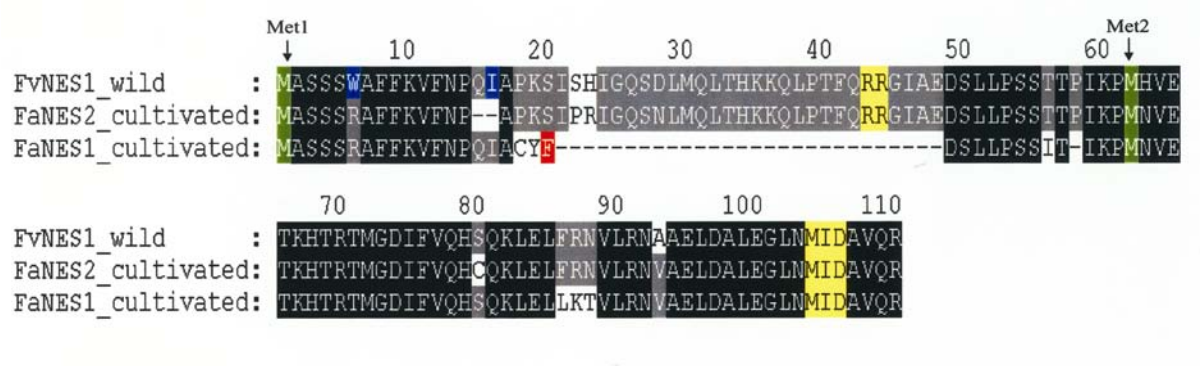


Figure 2 Deletion in the coding region of genes results in the gain of flavour compounds in cultivated strawberry fruit. The N-termini of the putative NES proteins up to the MID domain (marked in yellow). The stop codon in the *FaNES1* gene sequence, between Met1 and Met2 (shaded green), immediately follows the red shaded phenylalanine residue (F). The RR motif (shaded yellow) is indicative of a plastidic targeting signal in monoterpene synthases. Substitution (W6R) or removal (I16Δ) of the residues shaded blue resulted in a change of targeting (see Fig. 5). Black shading and gray shading represent residues identical in all three sequences or in 2 of them, respectively.

The *in-vivo* capacity of the FaNES1 enzyme to convert both GDP and FDP to linalool and nerolidol, respectively, was evaluated using transgenic Arabidopsis plants. The *FaNES1* gene was constitutively overexpressed and the heterologous protein was targeted to the plastids (see Methods). Transgenic plants expressing *FaNES1* produced and emitted both compounds but the level of linalool was about a 1000-fold higher than of nerolidol (Fig. 5). This difference was expected as plastids have been hypothesised to only contain the monoterpene precursor GDP, but also showed that some FDP must be present in plastids. The results demonstrate that the single protein encoded by *FaNES1,* a gene highly expressed during fruit ripening, is alone sufficient for the production of both linalool and nerolidol in cultivated strawberry varieties.

In addition, we examined the reverse situation in which the wild strawberry produces a range of monoterpenes that are not produced by the cultivated species. A monoterpene synthase was cloned from the wild species (*FvPINS;* see Methods) and heterologous expression in *E.coli* cells showed it to be a genuine monoterpene synthase. Its product profile (α-pinene as a main product and other monoterpenes such as β-phellandrene and β-myrcene as side products) also matched most of the components detected in the headspace of the wild strawberry (Fig. 1a). *FvPINS* expression in wild and cultivated strawberry was the mirror image of the *FaNES1* gene (Fig. 1e, f). Nucleic acid alignment of *FvPINS* and the corresponding coding region of *FaPINS,* which was

isolated from the cultivated strawberry, revealed only seven nucleic acid differences. The insertion of two tandem cytosine nucleotides in the middle of the coding region of *FaPINS* caused a frame shift, followed instantaneously by a UAA stop codon (Fig. 3). PCR on both wild and cultivated strawberry genomic DNA showed that one fourth of the fragments obtained from the cultivated strawberry showed the CC insertion while in the wild strawberry no fragments showed this frame shift mutation. Consequently, the capacity of the cultivated species to form multiple monoterpenes may have been lost due to the introduction of a deviant allele, which may cause the degradation of all mRNAs derived from the *FaPINS* gene. We suggest that RNA surveillance mechanisms such as Nonsense-Mediated Decay (NMD) might be responsible for the low expression of the *PINS* gene in the cultivated species. As a result of NMD, abnormal mRNAs containing premature translation termination codons are efficiently eliminated so that production of undesirable truncated proteins is avoided[13].

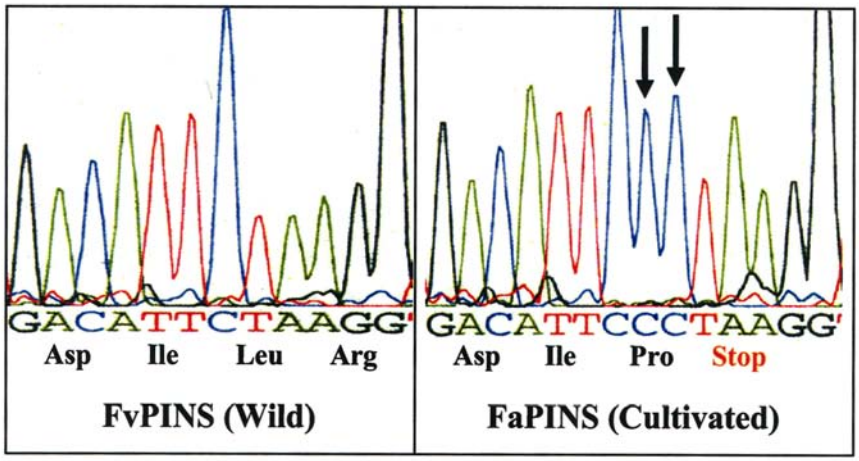


Figure 3 Insertion followed by a stop-codon in the coding region of genes result in the loss of flavour compounds in the cultivated strawberry. A two base pair insertion (CC, indicated by arrows) in the cultivated strawberry species results in a frame shift and an immediate stop-codon in the middle of the *FaPINS* gene coding region.

Starting from Met1, the putative FaNES1 protein contains a relatively large number of serine and alanine residues at the N- terminus (Fig. 2). These are characteristic of plastidic and mitochondrial localisation signals[14]. The current view is that the biosynthesis of monoterpenes occurs in the plastids and that of sesquiterpenes in the cytosol[15]. We have demonstrated that strawberry fruits produce monoterpenes (wild) or both mono- and sesquiterpenes (cultivated) (Fig. 1). However, ripe strawberries have been reported to contain mitochondria but no viable plastids[16,17] and the question therefore arises how the fruits might produce the typical plastidic monoterpenoids in the absence of these organelles.

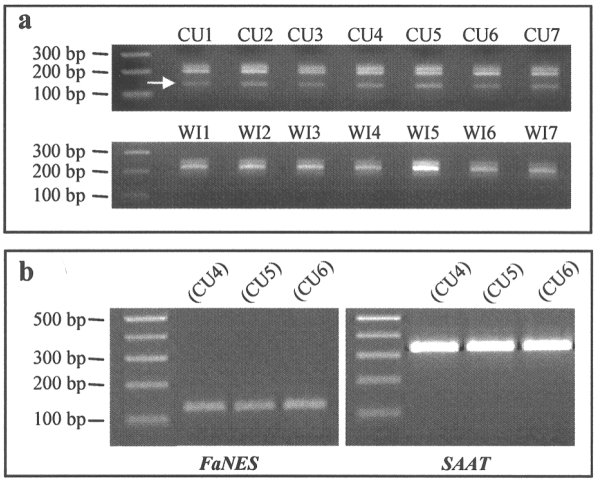


Figure 4 Correlation of *NES* genes presence, their expression and volatile product formation in wild and cultivated strawberry species. **a**, PCR on genomic DNA of 7 cultivated (CU1-CU7) and 7 wild strawberry species (WI1-WI7) using an oligonucleotide pair flanking the Met1 to Met2 region (see Fig. 2a). The arrow indicates a fragment of approximately 150 bp corresponding to the *FaNES1* gene. The higher bands correspond to other NES genes including ones with a proper targeting signal region **b,** RT-PCR (left) using RNA derived from fruits of cultivated varieties using similar oligonucleotides used in (**a**). The single band corresponds to the 150 bp fragment detected in (**a**). The strawberry alcohol acyltransferase gene (*SAAT;* Ref. 11) was used as control (right).

To test whether the 5' regions of the various cDNAs indeed encode a targeting signal we fused them to a Green Fluorescent Protein (GFP) reporter gene, transferred them to tobacco protoplasts and analysed transient GFP expression using confocal laser microscopy (Fig. 6). The results showed localised GFP fluorescence in chloroplasts of protoplasts transformed with the *FvNES1* 5' region (Fig. 6; C6-C8) and a dual targeting to both mitochondria and chloroplasts (mainly to mitochondria) for the *FaNES2* fragment (Fig. 6; C3-C5). In both cases the region between Met1 and Met2 alone was not sufficient for localization to the plastids or mitochondria and lead to expression in the cytosol. This region was however important in determining the specificity of localization (revealed using a hybrid construct; Fig. 6, C9). Fusions of the 5' part of *FaNES1* and GFP showed that the deletion and the stop-codon between Met1 and Met2 still allows translation (which will start from Met2) and results in a cytosolic localization of the protein (Fig. 6; C1-C2). Using a targeting signal prediction program (http://www.inra.fr/predotar/) the plastidic targeting of FvNES1 could be changed to the dual mitochondrial and plastidic targeting of *FaNES2* by a W6R substitution and I16Δ deletion (Fig. 6; C13 and Fig. 2). The results show that the N-termini of the proteins encoded by *FaNES2* and *FvNES1* are plastidic and/or mitochondrial targeting signals. Deletions and introduction of a premature stop-codon produce a transcript in which the targeting sequence is excluded from the coding region, and thus results in a cytosolic localisation as in the case of the *FaNES1* encoded protein.

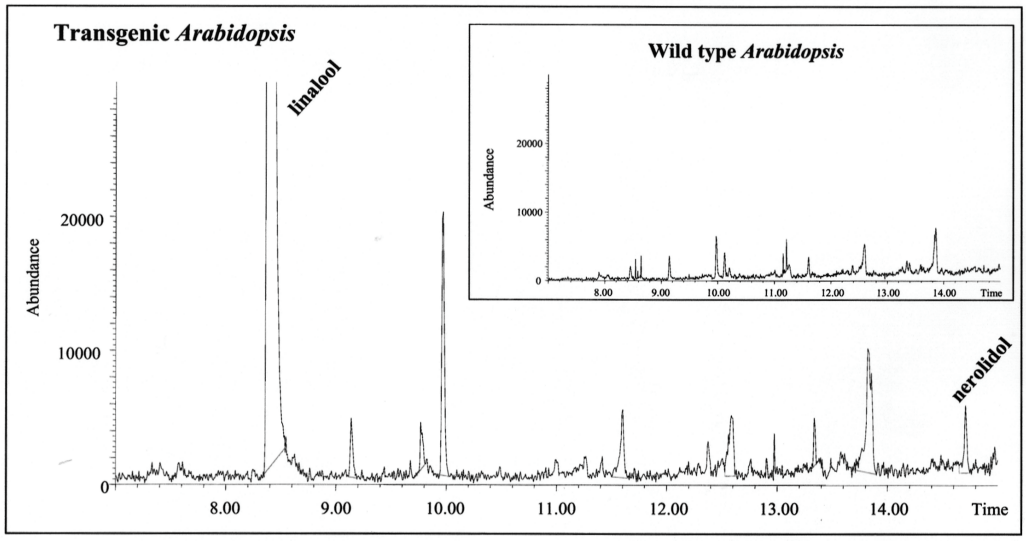
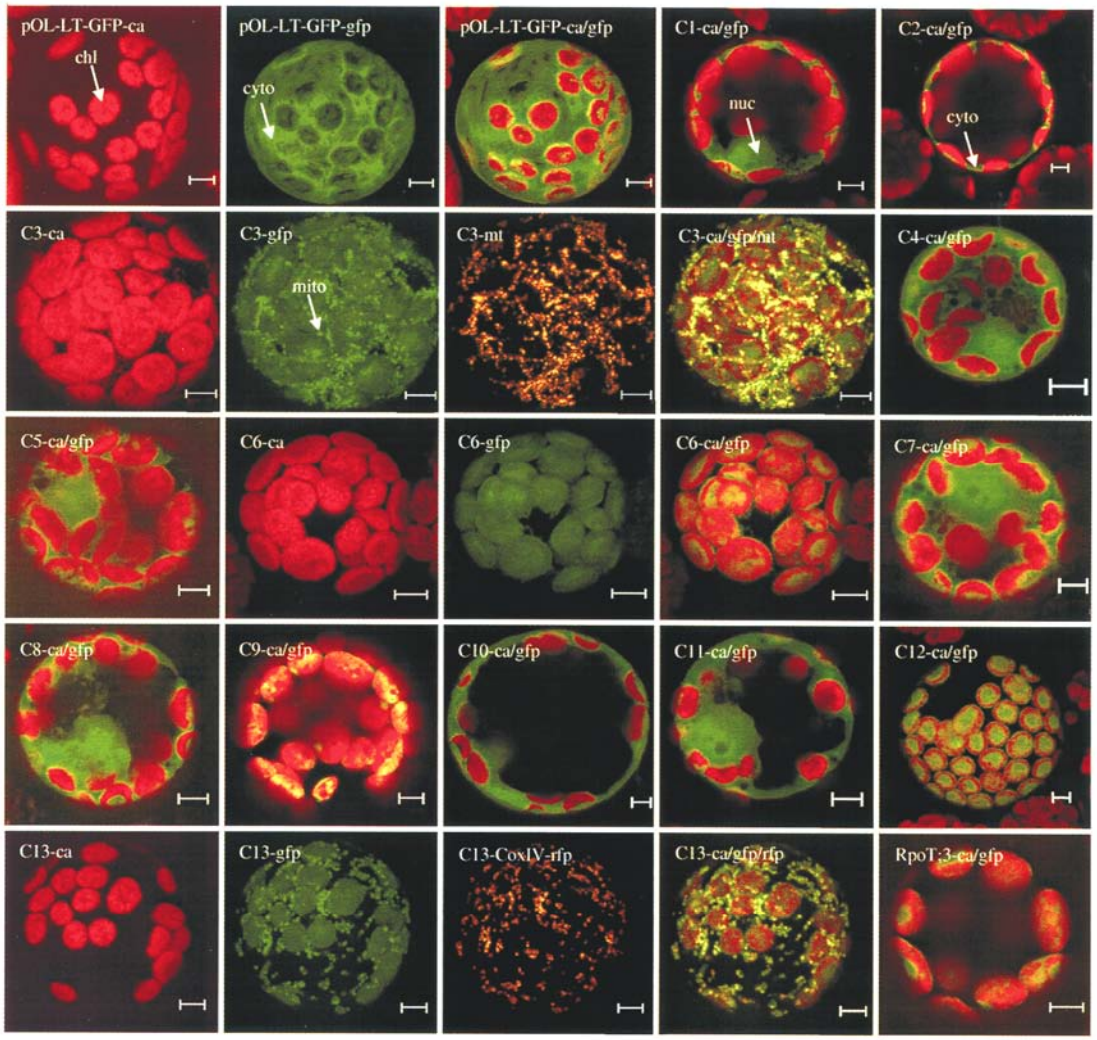


Figure 5 *In planta* simultaneous production of linalool and nerolidol by the FaNES1 enzyme. Headspace trapping and subsequent GC-MS analysis show emission of linalool and nerolidol from vegetative parts of the transgenic model plant Arabidopsis and their absence in the headspace profile of the wild type plant.

Our hypothesis that monoterpene biosynthesis in strawberry fruit occurs in the cytosolic compartment was strengthened by analysing the targeting capacity of the 5' region of the *FvPINS* gene. Two regions derived from the 5' end of the *FvPINS* gene (encoding 20 and 70 amino acid residues) did not lead to targeting and resulted in cytosolic GFP expression (Fig. 6; C10-C11). GFP fusion proteins of the 5' region of a monoterpene synthase active in young green lemon peel tissue (does contain plastids) (Lucker et al., *In press*), showed a clear plastidic subcellular localisation (Fig. 6; C12). Apparently, in strawberry the monoterpene precursor GDP is present in the cytosol in large enough amounts although its place of biosynthesis is unclear.

The present investigation of the gain and loss of flavour compounds in strawberry fruit has established a new molecular mechanism for the evolution of diversity in metabolic pathways. Ancestral genes (perhaps similar to *FaNES2* and *FvNES1*) most likely encoded plastidic and/or mitochondrial terpene synthases (both known sites for terpene formation in plants, Ref. 18). These genes may have been similar to higher terpene synthases such as diterpenes (C20) involved in primary metabolism, and proposed by Trapp and Croteau (Ref. 19) to serve as the ancestors of terpene synthases genes responsible for natural products biosynthesis (i.e. secondary metabolism). According to the latter report, duplication and divergence in structural and functional specialisation was accompanied by sequential intron loss (from 14 up to 6 introns), and the loss of specific motifs (*FaNES1* contains only 5 introns, the lowest number detected up to date in terpene synthases).


pOL-LT-GFP-ca
chl
pOL-LT-GFP-gfp
cyto
pOL-LT-GFP-ca/gfp
C1-ca/gfp
nuc
C2-ca/gfp
cyto
C3-ca
C3-gfp
mito
C3-mt
C3-ca/gfp/mt
C4-ca/gfp
C5-ca/gfp
C6-ca
C6-gfp
C6-ca/gfp
C7-ca/gfp
C8-ca/gfp
C9-ca/gfp
C10-ca/gfp
C11-ca/gfp
C12-ca/gfp
C13-ca
C13-gfp
C13-CoxIV-rfp
C13-ca/gfp/rfp
RpoT3-ca/gfp

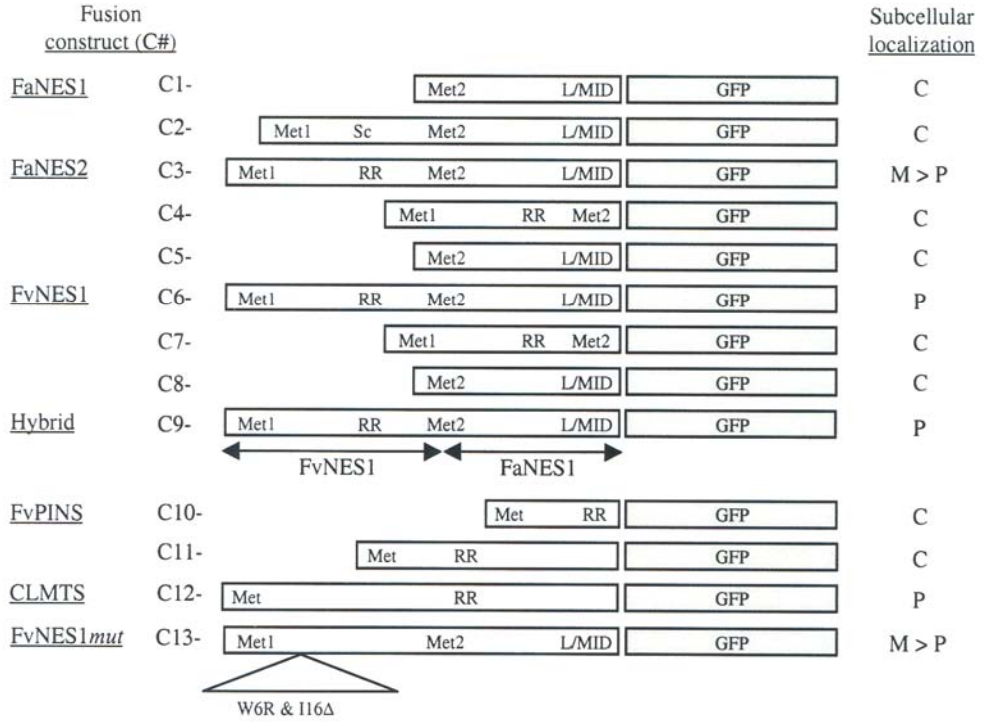

Fusion construct (C#)
Subcellular localization
FaNES1
C1-
Met2
L/MID
GFP
C
C2-
Met1
Sc
Met2
L/MID
GFP
C
FaNES2
C3-
Met1
RR
Met2
L/MID
GFP
M > P
C4-
Met1
RR
Met2
GFP
C
C5-
Met2
L/MID
GFP
C
FvNES1
C6-
Met1
RR
Met2
L/MID
GFP
P
C7-
Met1
RR
Met2
GFP
C
C8-
Met2
L/MID
GFP
C
Hybrid
C9-
Met1
RR
Met2
L/MID
GFP
P
FvNES1
FaNES1
FvPINS
C10-
Met
RR
GFP
C
C11-
Met
RR
GFP
C
CLMTS
C12-
Met
RR
GFP
P
FvNES1mut
C13-
Met1
Met2
L/MID
GFP
M > P
W6R & I16Δ

Figure 6 (previous page) Transient expressions of GFP fusions in wild-type tobacco protoplasts detected by confocal laser scanning microscopy. ca, chlorophyll auto-fluorescence detected in the red channel; gfp, green fluorescent protein fluorescence detected in the green channel; ca/gfp, combined red and green channels; mt, MitoTracker mitochondrial stain and rfp (red fluorescent protein) fluorescence both detected in the orange channel or merged with the green and red channels ca/gfp/mt and ca/gfp/rfp, respectively. pOL-LT-GFP is the original vector used for fusing the different strawberry gene fragments to GFP that is directing GFP expression to the cytosol and nucleoplasm (see Methods). The RpoT;3-gfp and the CoxIV-rfp constructs served as positive controls for targeting to chloroplasts and mitochondria, respectively (Ref. 27 and Ref. 29). Chloroplasts (chl, approximately 5µm size), mitochondria (mito, approximately 1µm size), cytosol (cyto) nucleoplasm (nuc) are pointed out by arrows. If for a single construct several images are shown they are derived from the same protoplast. At the bottom, schematic and localization results for each fusion construct. Met1 and Met2, methionine residues at the N-termini of the proteins (Fig. 2); MID, conserved motif in terpene synthases (Fig. 2); Sc, stop codon; C, cytosol; P, plastids; M > P, mitochondria more than plastids. CLMTS, GFP fusion of the 5' region of a typical monoterpene synthase from *Citrus Limon* (Lucker et al., *In press*). C13 is a mutated (*mut*) FvNES1 uses the same fragment as construct C6 but with a W6R substitution and IΔ16 deletion (see also Fig. 2). Scale bars, 5 µm.

In strawberry, evolutionary events occurring in parallel in several wild species native to different regions of the world modified the N-termini of the proteins. Deletions and insertion of stop codons in the coding regions removed the targeting signal region and led to a cytosolic subcellular localisation. This change in localisation enabled the enzymes to encounter their substrates but was not yet sufficient for the biosynthesis of high amounts of products. An additional change in the level of gene expression and tissue specificity (ripe fruit tissue) possibly by gene duplication during the increase in strawberry ploidy level resulted in the formation of the *FaNES1* gene. The selection of the *FaNES1* gene in the genome of certain strawberry species provided them with a strong selective advantage (either flavour or resistance or both), which was systematically passed on to the domesticated strawberries of today during years of breeding and selection. We suggest that the gain of linalool and nerolidol in the cultivated species under breeding selection pressure has lead to the loss of the ability to produce the olefinic monoterpenes of the wild strawberry, which are derived from the same substrate. However, it cannot be excluded that the loss of the olefinic monoterpenes is solely due to a negative selection pressure. In both possibilities a change in metabolite composition was accomplished by independent mechanisms initiated by insertions and deletions in coding regions of structural genes in the pathway. The remarkable diversity of more than 100,000 low molecular-mass natural products produced by plants[20] suggests that probably numerous other beneficial yet unknown evolutionary mechanisms have been adapted by plants for the production of specific metabolites.

METHODS

Volatile analysis

Greenhouse grown strawberry varieties from the Plant Research International breeding collection were used. For headspace analyses red ripe strawberry fruit were enclosed in 1L glass jars closed with a teflon-lined lid equipped with in and outlet. A vacuum pump was used to draw air through the glass jar at approximately 100 mL min^{-1}, with the incoming air being purified through a glass cartridge (140 x 4 mm) containing 150 mg Tenax TA (20/35 mesh, Alltech, Breda, the Netherlands). At the outlet the volatiles emitted by the detached fruits were trapped on a similar Tenax cartridge. Volatiles were sampled during 24 h. Cartridges were eluted using 3 x 1 mL of redistilled pentane-diethyl ether (4:1). Of these samples, 2 µL were analysed by GC-MS using an HP 5890 series II gas chromatograph equipped with an HP-5MS column (30 m x 0.25 mm i.d., 0.25 µm df) and an HP 5972A Mass Selective Detector as described by Bouwmeester et al. (Ref. 21). For characterisation of heterologous proteins after His-tag purification, 100 µL of the eluent was diluted to 1 mL with assay buffer containing 15 mM Mopso (pH 7.0), 10% (v/v) glycerol, 10 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM sodium ascorbate and 2 mM DTT. To the assay 20 µM [^{3}H]-GDP or [^{3}H]-FDP was added. After the addition of a 1-mL redistilled pentane overlay, the tubes were carefully mixed and incubated for 1 h at 30°C. The assays were extracted as described before (Ref. 22) and analysed using radio-GC on a Carlo-Erba 4160 Series gas chromatograph equipped with a RAGA-90 radioactivity detector (Raytest, Straubenhardt, Germany) and GC-MS[21,22].

Molecular techniques and expression in *E.coli* and *Arabidopsis*

Total cellular DNA isolation was performed as described by Marty et al., (Ref. 23). RNA isolation and expression analysis using either RNA gel blots or cDNA microarrays (different fruit tissues and leaves) were conducted as described earlier (Ref. 11). Full-length cDNAs were cloned using the SMART RACE cDNA Amplification Kit (Clontech) according to the manufacturer instructions with slight modifications minor to annealing temperatures (normally reduced by 5 to 10°C then recommended) or number of cycles (up to 35 cycles). The published fragment of the *FvPINS* gene (Ref. 24) was used for designing oligonucleotides and cloning the entire *FvPINS* and *FaPINS* cDNAs. PCR, restriction digests, plasmid DNA isolation and gel electrophoresis were performed using standard protocols. All fragments were purified out of gel using the GFX purification kit (Amersham). Cloning of PCR fragments was either done to the PCR SCRIPT (Stratagene) or pCR 4Blunt-TOPO (Invitrogen) vectors (for blunt-end products generated when using pfu polymerase) or to the pGEM-T Easy (Promega) vector (when A tailed PCR products were generated by the use of Taq polymerase). Sequencing was done using the ABI 310 capillary

sequencer according to the manufacturer instructions (ABI system, Perkin Elmer). Bacterial expression was performed as described by Aharoni et al. (Ref. 11). The region in between the two AUG codons in the *FvNES1* 5' end (encoding Met1 and Met2) was cloned in frame and upstream of the *FaNES1* cDNA (similar to the plastidic targeted C9 construct in Figure 6), transferred to the pBinPlus vector as described by Aharoni et al. (Ref. 25), which was used for obtaining transgenic Arabidopsis (ecotype Columbia) as described by Clough et al. (Ref. 26).

Green Fluorescent Protein (GFP) assays

The different fragments used for localisation analysis were fused upstream of and in-frame with the GFP gene in the cloning sites (SpeI and SalI) present in the cassette of pOL-LT-GFP-L64T65 (modified pOL-GFPS65C, Ref. 27) using *Xba*I and *Xho*I restriction sites (at the 5' and 3' end, respectively) introduced by PCR. Tobacco protoplasts (cv. SRI) were prepared and transformed using 50 µg plasmid DNA (of each construct also when two constructs were co-transformed) as described earlier (Ref. 28). MitoTracker Red CMXRos staining were conducted according to the manufacturer instructions (Molecular Probes, Eugene, OR, USA). Protoplasts were examined 24 h after transformation with a Carl Zeiss confocal laser scanning microscope (LSM 510) with an argon ion laser. The fluorescence of GFP (abs, 488 nm; em, 507 nm), was obtained using the 488 nm laserline and the emission signal was collected using a band-pass filter (BP 505-550). For imaging the 40x objective was used and the excitation intensity was set at 2 to 4% in most cases. The long-wavelength signal of the chlorophyll was collected using a long-pass filter (LP650). When the protoplasts were stained with MitoTracker Red (abs, 578 nm; em, 599 nm), images were taken via the multi excitation mode using the 488 and 543 nm laserlines sequentially. Both chlorophyll autofluorescence and MitoTracker Red emission were collected via a long-pass filter (LP650). For monitoring RFP (abs; 558 nm, em; 585 nm), excitation was at 543 nm and the emission signal was detected with a band-pass filter (BP 560-615). Optical sections were taken along the optical axis and projected into one image with the Zeiss LSM Image Browser (Carl Zeiss, Jena, Germany).

ACKNOWLEDGEMENTS

We thank Ian Small, Nemo Peeters, Maureen Hanson and Andreas Weihe for the GFP and RFP vectors. We thank Patrick Smit, Mark Hink, and Jan Willem Borst for assistance with CLSM, Andy Pereira, Robert Hall and Christopher Danpure for comments on the manuscript and Joost Lucker for providing the *CLMTS* cDNA and helpful discussions.

REFERENCES

1.**Pichersky, E. & Gang D. R.** Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. *Trends Plant Sci.* **5**, 439-445. (2000).

2.**Charron, C. S., Cantliffe, D. J. & Heath, R. R.** Volatile emissions from plants. *Horticultural Reviews* **17**, 43-72 (1995).

3.**Baldwin E. A., Scott, W. Shewmaker, C. K. & Schuch, W.** Flavor trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. *Hortscience* **35**, 1013-1022 (2000).

4.**Maarse, H.** (Marcel Dekker, Inc., New York).

5.**Seigler, D. S.** *Plant Secondary Metabolism* (Chapman and Hall, New York, 1999).

6.**Kubo, I., Muroi, H. & Himejima, M.** Antibacterial activity against *Streptococcus mutans* of mate tea flavor components. *J. Agric. Food Chem.* **41**, 107-111 (1993).

7.**Little, D. B. & Croteau, R. B.** in *Flavor chemistry* 239-253 (1999).

8**Helsper, J. P. F. G., J.A. Davies, H.J. Bouwmeester and M.H. van Kampen,.** Circadian rhythmicity in emission of volatile compounds by flowers of *Rosa hybrida*, cv 'Honesty'. *Planta* **207**, 88-95 (1998).

9.**MacLeod, A. J.** (ed.) *Food flavours. Part C. The flavour of fruits* (Elsevier Science Publishers, Amsterdam, 1990).

10.**Chappell, J.** Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 521-547 (1995).

11.**Aharoni, A. *et al.*** Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* **12**, 647-661 (2000).

12.**Williams, D. C., McGarvey, D. J., Katahira, E. J. & Croteau, R. B.** Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* 12213-12220 (1998).

13.**Isshiki, M., Yamamoto, Y., Satoh, H. & Shimamoto, K.** Nonsense-mediated decay of mutant waxy mRNA in rice. *Plant Physiol.* **125**, 1388-1395 (2001).

14.**Small, I., Wintz, H., Akashi, K. & Mireau, H.** Two birds with one stone: Genes that encode products targeted to two or more compartments. *Plant Mol. Biol.* **38**, 265-277 (1998).

15.**Phillips, M. A. & Croteau, R. B.** Resin-based defenses in conifers. *Trends in Plant Science* **4**, 184-190 (1999).

16.**Perkins-Veazie, P.** Growth and ripening of strawberry fruit. *Hort. Rev.* **17**, 267-297 (1995).

17.**Knee, M., Sargent, J. A. & Osborne, D. J.** Cell wall metabolism in developing strawberry fruits. *J. Exp. Bot.* **28**, 377-396 (1977).

18.**Bouvier, F., Suire, C., d' Harlingue, A., Backhaus, R. A. & Camara, B.** Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *Plant J.* **24**, 241-252 (2000).

19.**Trapp, S. C. & Croteau, R. B.** Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **158**, 811-832 (2001).

20.**Dixon, R. A.** Natural products and plant disease resistance. *Nature* **411**, 843-847 (2001).

21.**Bouwmeester, H. J., Verstappen, F. W. A., Posthumus, M. A. & Dicke, M.** Spider mite-induced (3*S*)-(*E*)-nerolidol synthase activity in cucumber and lima bean. The first dedicated step in acyclic C11-homoterpene biosynthesis. *Plant Physiol.* **121**, 173-180 (1999).

22.**Bouwmeester, H. J., Kodde, J., Verstappen, F. W. A., Altug, I. G., de Kraker, J-W. & Wallaart, T. E.** Isolation and characterization of two germacrene A synthase cDNA clones from chicory. *Plant Physiol.* **129,** 134-144 (2002).

23.**Marty, I. *et al.*** The cystathionine-gamma-synthase gene involved in methionine biosynthesis is highly expressed and auxin-repressed during wild strawberry (*Fragaria vesca L.*) fruit ripening. *Theo. App. Genet.* **100**, 1129-1136 (2000).

24.**Nam, Y. W. *et al.*** Isolation and characterisation of mRNAs differentially expressed during ripening of wild strawberry (Fragaria vesca L.) fruits. *Plant Mol. Biol.* **39**, 629-636 (1999).
25.**Aharoni, A. *et al.*** The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant J.* **28**, 319-332 (2001).
26.**Clough, S. J. & Bent, A. F.** Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735-743 (1998).
27.**Peeters, N. M. *et al.*** Duplication and quadruplication of *Arabidopsis thaliana* cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. *J. Mol. Evol.* **50**, 413-423 (2000).
28.**Negrutiu I, S. R., Potrykus I, Biasini G and Sala F.** Hybrid genes in the analysis of transformation conditions. *Plant Mol. Biol.* **8**, 363-373 (1987).
29.**Hedtke, B. *et al.*** Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo. *Plant J.* **17**, 557-561 (1999).

CHAPTER 7

Asaph Aharoni, C. H. Ric De Vos, Martina Wein, Zhongkui Sun, Raffaella Greco, Arthur Kroon, Joseph N.M. Mol and Ann P. O'Connell

The Strawberry *FaMYB1* Transcription Factor Suppresses Anthocyanin and Flavonol Accumulation in Transgenic Tobacco

The Plant Journal 28(3): 319-332 (2001)

THE STRAWBERRY *FaMYB1* TRANSCRIPTION FACTOR SUPPRESSES ANTHOCYANIN AND FLAVONOL ACCUMULATION IN TRANSGENIC TOBACCO

Fruit ripening is characterised by dramatic changes in gene expression, enzymatic activities and metabolism. Although the process of ripening has been studied extensively, we still lack valuable information on how the numerous metabolic pathways are regulated and co-ordinated. In this paper we describe the characterisation of *FaMYB1*, a ripening regulated strawberry gene member of the MYB family of transcription factors. Flowers of transgenic tobacco lines overexpressing *FaMYB1* showed a severe reduction in pigmentation. A reduction in the level of cyanidin 3-rutinoside (an anthocyanin) and of quercetin-glycosides (flavonols) was observed. Expression of late flavonoid biosynthesis genes and their enzyme activities were aversely affected by *FaMYB1* overexpression. Two-hybrid assays in yeast showed that FaMYB1 could interact with other known anthocyanin regulators, but it does not act as a transcriptional activator. Interestingly, the C-terminus of FaMYB1 contains the motif pdLNL$^{D}/_{E}$Lxi$^{G}/_{S}$. This motif is contained in a region recently proposed to be involved in the repression of transcription by AtMYB4, an *Arabidopsis* MYB protein. Our results suggest that *FaMYB1* may play a key role in regulating the biosynthesis of anthocyanins and flavonols in strawberry. It may act to repress transcription in order to balance the levels of anthocyanin pigments produced at the latter stages of strawberry fruit maturation and/or to regulate metabolite levels in various branches of the flavonoid biosynthetic pathway.

INTRODUCTION

Several characteristics of fleshy fruit development and maturation might be attributed to changes in the levels of certain phenolic compounds, amongst them flavonoids (Seymour *et al.,* 1993). In strawberry for example, astringency associated with early fruit development is due to the presence of tannins, whilst anthocyanins are responsible for the characteristic red color associated with late ripening (Cheng *et al.,* 1991). In cultivated strawberry, the glucosylated anthocyanin pelargonidin (pelargonidin 3-glucoside) is the main anthocyanin present in ripe fruit (approx. 88%), along with other pelargonidin-glycosides and cyanidin 3-glucoside (Perkins-Veazie, 1995). During ripening, in addition to anthocyanins, other flavonoid derivatives such as the flavonols quercetin and kaempferol are also produced (Hakkinen *et al.,* 1999). The biological function of these compounds in fruit has been mainly credited to their role in attracting fruit-eating animals and protecting against harmful ultraviolet light and pathogens (Shirley, 1996).

The flavonoid biosynthesis pathway is a branch of the large phenylpropanoid pathway, which includes the biosynthesis routes to compounds such as lignins and phenolic acids (and their derivatives). The precursors for the synthesis of all flavonoids are malonyl-CoA and p-coumaroyl-CoA (Dixon and Steele, 1999). Virtually all genes encoding enzymes for the biosynthesis of flavonols and anthocyanins have been identified (Mol *et al.,* 1998), including those modifying the basic skeleton (e.g. glycosylation, methylation, acylation and hydroxylation). In many plant species the structural genes of the flavonoid biosynthetic pathway appear to be largely regulated at the transcriptional level. Amongst the regulatory genes identified to date, members of the MYB family of transcription factors are chiefly represented (Jin and Martin, 1999). From the analysis of more than a hundred R2R3-MYB protein sequences from different plant species (mainly *Arabidopsis thaliana*), a number of conserved motifs have been identified (Kranz *et al.,* 1998; Riechmann and Ratcliffe, 2000). On the basis of these motifs, the MYB proteins have been classified into 22 subgroups. Some of these identified motifs may represent activation domains, whilst others may serve as repression domains or even domains for interaction with other transcription factors. The majority of MYB genes with assigned functions have been predicted to be transcriptional activators (Larkin *et al.,* 1993; Urao *et al.,* 1993; Martin and Paz-Ares, 1997; Glover *et al.,* 1998).

Transcriptional regulation of gene expression is not solely mediated by activators but also by the action of repressors and, in some cases, a transcription factor may perform both activities (Coffman *et al.,* 1997). Temporal and spatial interactions between different combinations of activators and repressors give rise to a wide spectrum of expression patterns (i.e. “combinatorial control”). Studying real transcriptional repression is complicated, due to effects caused by a dominant negative form of proteins (Schwechheimer and Bevan, 1998). Early studies on MYB transcription factors in maize reported on *C1-I* as a dominant inhibitor allele of the *C1* gene, which

is normally required for the synthesis of anthocyanin in the aleurone tissue (Paz-Ares *et al.,* 1990). Transient expression assays showed that changes in both the C-terminus and in the DNA binding domain were important for the *C1-I* inhibitory effect (Goff *et al.,* 1991). Transient expression assays with another *C1* homolog from maize, termed *Zm38*, showed that constructs retaining either the entire *Zm38* cDNA or a combination of the *Zm38* 3' region and the 5' region of the *C1* cDNA confer complete inhibition of the co-transformed wild type *C1* construct (Franken *et al.,* 1994). The mode of action of these putative repressors was suggested to be either competition for binding sites in target promoters or the formation of mixed dimers of C1-I and Zm38 with C1. Interestingly, aligning Zm38 C-terminus amino acid sequence reveals that it is structurally related to two other members of the subgroup 4 of MYB proteins (AmMYB308 and AtMYB4), recently reported to function as transcriptional repressors (Tamagnone *et al.,* 1998; Jin *et al.,* 2000). Overexpression of the *Antirrhinum AmMYB308* gene in tobacco caused an inhibition of hydroxycinnamic acid and monolignol accumulation, by reducing expression of genes encoding enzymes that are part of their biosynthetic pathway. An *Arabidopsis* knockout mutant of *AtMYB4*, the proposed orthologous gene of *AmMYB308*, showed increase in sinapate ester accumulation, which resulted in enhanced UV-B irradiation tolerance. Sinapate esters are utilised as sunscreens in *Arabidopsis* leaves. Under certain environmental conditions, when protection from irradiation is not essential (e.g. dark), AtMYB4 reduces the accumulation of sinapate esters by repressing the expression of the gene encoding the key enzyme in their biosynthesis (cinnamate 4-hydroxylase, *C4H*).

In an attempt to study the genetic regulation of strawberry ripening, we have cloned and characterised the *FaMYB1* gene. *FaMYB1* encodes a R2R3 MYB protein homolog, which is primarily expressed in the red ripe strawberry fruit. Heterologous expression in tobacco primarily affects the metabolism of anthocyanins and flavonols and suggests a role for *FaMYB1* as a transcriptional regulator of late flavonoid biosynthesis genes.

RESULTS

Strawberry FaMYB1 is a Short R2R3 MYB Protein

Using a random EST (Expressed Sequence Tags) sequencing approach in which 1100 cDNAs derived from a strawberry red fruit cDNA library were partially sequenced (Aharoni *et al.,* 2000), we identified and isolated two strawberry transcripts putatively encoding R2R3 MYB proteins (*FaMYB1* and *FaMYB2*). *FaMYB1* appeared to be a full-length cDNA (1135 bp) encoding a protein of 187 amino acids. *FaMYB2* was a truncated cDNA (732 bp) that when translated was nearly identical to the putative translation product of *FaMYB1* except for two base pair substitutions, which either did not alter the amino acid composition or produced a conserved

amino acid substitution. Both transcripts showed large differences in their 3' untranslated regions (UTR), primarily due to a deletion of 143 bp in *FaMYB2*. We assumed that both clones were derived from different alleles within the genome of the octaploid variety used for their isolation (c.v. *Elsanta*).

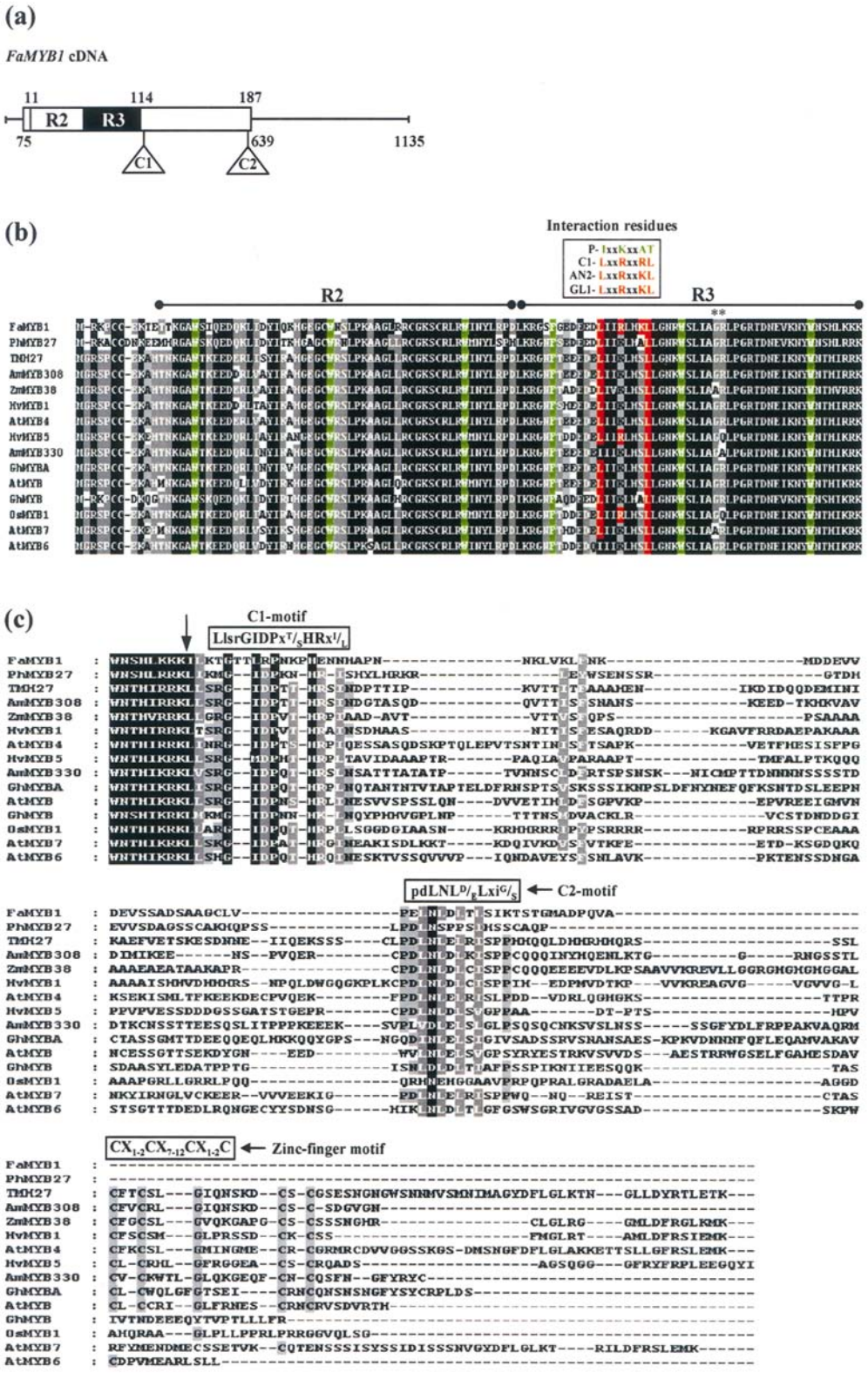


Figure 1 (previous page) *FaMYB1* cDNA structure and homology to plant R2-R3 MYB proteins. **(a)** Schematic drawing of *FaMYB1* cDNA showing the coding region (rectangle) and the 3' and 5' untranslated regions (lines). R2 and R3 are the two repeats of the MYB DNA binding domain. C1 and C2 are two conserved motifs located outside the MYB domain and identified in other members of subgroup 4 MYB proteins (Kranz *et al.,* 1998). The numbers in the upper and lower part correspond to amino acid and nucleic acid, respectively. **(b)** Amino acid sequence alignment of *FaMYB1* and other R2-R3 MYB protein homologs in the MYB domain. The line above indicates the R2 and R3 MYB repeats and the regularly spaced tryptophan (W) residues (or phenylalanine, F) are shaded in green. Black background indicates 100% conservation and gray or light gray represent ≥ 80% and ≥ 60% conservation, respectively. The four amino acid residues required for the interaction with the maize bHLH protein R (Grotewold *et al.,* 2000) present in the MYB proteins C1, AN2 and GL1 but not in the P MYB protein are shown (box) and marked in all proteins (shaded in red). Additional two amino acids important for interaction in plant cells are marked above with asterisks (residues GR), both are present in C1, AN2 and GL1. Abbreviations: At, *Arabidopsis thaliana*; Fa, *Fragaria anannasa*; Ph, *Petunia hybrida*; TM, Tomato; Am, *Antirrhinum majus*; Zm, *Zea mays*; Hv, *Hordum vulgare*; Gh, *Gossypium hirsutum*; Os, *Oryza sativa*. GenBank accession numbers or reference for the sequences: FaMYB1, AF401220; PhMYB27, Mur, 1995; TMH27, X95296; AmMYB308, JQ0960; ZmMYB38, X78846; HvMYB1, X78845; AtMYB4, AF062860; HvMYB5, X70880; AmMYB330, JQ0957; GhMYBA, L04497; AtMYB, genomic clone T22J18.19; GhMYB (GhMYB6), AF034134; OsMYB1, D88617; AtMYB7, U26937; AtMYB6, U26936. **(c)** Amino acid sequence alignment of *FaMYB1* and other R2-R3 MYB protein homologs outside the MYB domain. The arrow indicates the end of the R3 repeat in the MYB domain. Black background indicates 100% conservation and gray or light gray represent ≥ 80% and ≥ 60% conservation, respectively. The motifs C1, C2 and a zinc finger are indicated above their location in the protein sequences, as described by Kranz *et al.* (1998) for members of MYB subgroup 4. Upper or lower case letters indicate amino acids found in all or more than 50% of the subgroup genes respectively. If two amino acids are found at the same position, both are given and they are separated by a slash.

FaMYB1 encodes a relatively short MYB protein (187 amino acids compared to an average of 230 to 300 amino acids for most other plant MYB proteins). It has however, a relatively long 3' UTR (496 bp) compared to its predicted open reading frame (ORF, 564 bp, see Figure 1a). We confirmed that the protein encoded by *FaMYB1* was not a truncated R2R3 MYB protein but a functional entity by itself, by performing reverse transcription PCR (RT-PCR) and 3' RACE experiments (data not shown).

The FaMYB1 protein sequence showed greatest homology to GhMYB6, a cotton ovule protein (GenBank accession no. AF034134), with 60.1% overall identity and 84.5% identity in the R2R3 DNA binding domain (Figure 1b and 1c). As for other MYB proteins, FaMYB1 shared very little homology to other MYB proteins in the C-terminus region. It however contained two conserved motifs in either side of its short C-terminus (C1 and C2 in Figure 1a and 1c), which relates it to several other MYB proteins previously clustered as members of subgroup 4 (Kranz *et al.,* 1998). Unlike a number of its homologs, FaMYB1 does not contain a zinc-finger motif in the C-terminus (Figure 1c). Interestingly, some FaMYB1 homologs were previously suggested to act as transcriptional repressors or weak activators (AmMYB308 and AmMYB330, Tamagnone *et al.,* 1998; Zm38, Franken *et al.,* 1994; C1-I, Paz-Ares *et al.,* 1990; PhMYB27, Mur, 1995 and

AtMYB4, Jin *et al.,* 2000). Additional evidence was recently provided by Jin *et al,* (2000) that the motif pdLNL$^{D}/_{E}$Lxi$^{G}/_{S}$ (C2 motif in *FaMYB1*, see Figure 1a and 1c) forms part of the region involved in the repression of transcription.

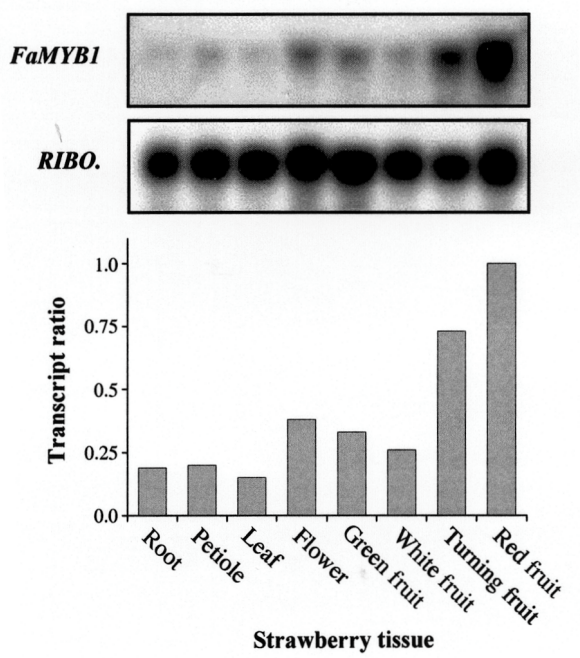


Figure 2 *FaMYB1* gene expression in strawberry.RNA gel blot analysis of the expression of *FaMYB1* in different tissues of strawberry. The blot was hybridized with a fragment containing the 3' region outside the conserved MYB domain of *FaMYB1* and then re-hybridized for loading control with a strawberry cDNA probe showing homology to a gene encoding a ribosomal protein (*RIBO*.). Quantification of the results was performed by measuring the OD/mm^2 (using the TINA software, Raytest, Straubenhardt, Germany) of the bands and calculating the ratios between the band in the gene specific gel blot and its corresponding band in the loading control blot (transcript ratio).

Recently, Grotewold *et al*. (2000) reported on the identification of amino acid residues in the maize transcriptional activator C1, which specify the interaction between the MYB domain of C1 and the N-terminal of the basic helix-loop-helix (bHLH) protein R. C1 depends on the interaction with R for its regulatory function in anthocyanin biosynthesis. With yeast two-hybrid experiments, the authors identified four residues in the first helix (all 4 residues must be present) and additional two residues in the second helix of the R3 repeat of C1, which are necessary for the interaction with R in plant cells (Figure 1b). The maize MYB protein P, closely related to C1, is not dependent on R for its activity and does not have the necessary interacting residues. Apart from C1, two additional MYB transcription factors, GL1 and AN2, previously shown to interact in yeast two-hybrid experiments with R, contain the required residues (Grotewold *et al.,* 2000). Interestingly, FaMYB1 was the only protein amongst its closest homologs, which contained the six required amino acid residues (Figure 1b).

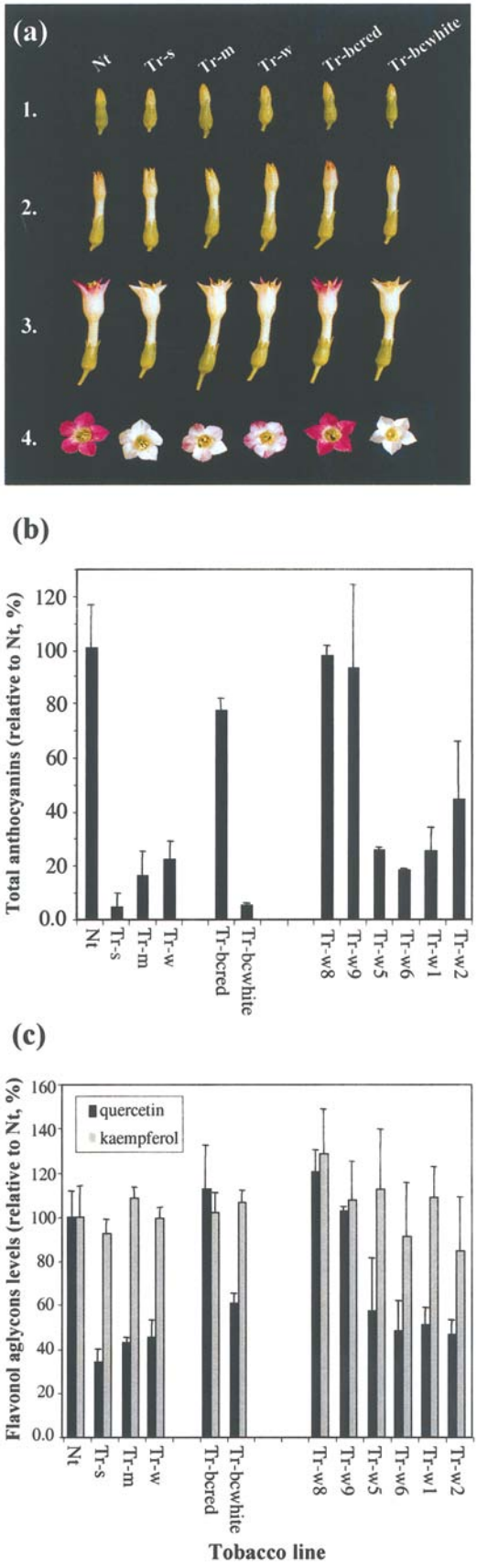


Figure 3 Changes in tobacco flowers induced by expression of strawberry *FaMYB1*. **(a)** Non-transformed (Nt), three *FaMYB1* expressing lines exhibiting different phenotypic characteristics (strong, Tr-s; medium, Tr-m; weak, Tr-w),

and a red (Tr-bcred) and a white (Tr-bcwhite) backcross line [Tr-s line was backcrossed with a non-transformed (Nt) wild type line] at four different flowering stages (1-4). Flowering plants showed a clear phenotypic change in petal pigmentation, already visible from early stages of flower development (prior to anthesis). In several plants alterations in petal color resulted in a complete loss of pigmentation. In other cases color patterning was observed in which the main veins of the flower became red whilst the rest of the flower appeared white. **(b)** HPLC analysis of anthocyanin levels in *FaMYB1* and non-transformed lines. Non-hydrolyzed extracts of petals from stage 4 (see above) tobacco flowers were used. Tr-w-1, 2, 5, 6, 8, 9 lines are the progeny of the weak phenotype line (Tr-w) selected visually for analysis according to their flower pigmentation. **(c)** *FaMYB1* affects total quercetin levels in primary transformants and progeny. HPLC results of hydrolyzed extracts from petals (stage 4 tobacco flowers) of primary transformants (Tr-s, Tr-m, Tr-w), a non-transformed line (Nt), two backcross lines (Tr-bcred and Tr-bcwhite) and six lines from the progeny of the selfed weak phenotype primary transformant (Tr-w-1, 2, 5, 6, 8, 9). Bars indicate the levels (means and s.d., n=3 plants) of the flavonols quercetin and kaempferol relative (in percentage) to the mean levels in the non-transformed line (Nt).

FaMYB1 is Highly Expressed in the Red Ripe Strawberry Fruit

FaMYB1 expression in both vegetative and reproductive strawberry plant tissues was analysed by RNA gel blot (Figure 2). Hybridization was performed using a probe derived from the 3' region of the *FaMYB1* cDNA (nucleotides 618 to 1005), in order to avoid cross hybridization with other members of the MYB gene family in strawberry. *FaMYB1* expression commenced at the turning stage and reached the highest levels in the red ripe fruit stage. Low expression levels were detected in the flower and green fruit tissues. We also checked the expression of *FaMYB1* in the achenes (seeds) and in the dark red over-ripe stage of fruit development (data not shown). No expression could be detected in achenes. However, a strong expression similar to the one shown for the red ripe fruit tissue was observed in the over-ripe fruit tissue.

Tobacco Plants Overexpressing *FaMYB1* and their Progeny are Affected in Flower Pigmentation

In order to ascertain a putative function for the strawberry *FaMYB1*, we used tobacco as a model system for heterologous expression experiments. *FaMYB1* was expressed under the control of an enhanced 35S-CaMV promoter. Twenty-five primary transformant lines and three control wild type lines were grown to maturity. One third of the flowering transformed lines showed clear phenotypic changes in petal pigmentation, already visible at early stages of flower development (prior to anthesis, see Figure 3a). In several lines, alterations in petal color resulted in a complete loss of pigmentation (severe phenotype). In other cases a distinct color patterning was observed in which often the main veins of the flower remained red and a few regions were pigmented with pale red color, whilst the remainder of the flower appeared white (medium phenotype). Apart from the reduction in petal pigmentation, the weak red colouring normally encountered in the distal part

of the mature stamen filaments of wild type tobacco was no longer visible in several transformed lines. Additionally, these lines there showed premature browning of the mature flowers.

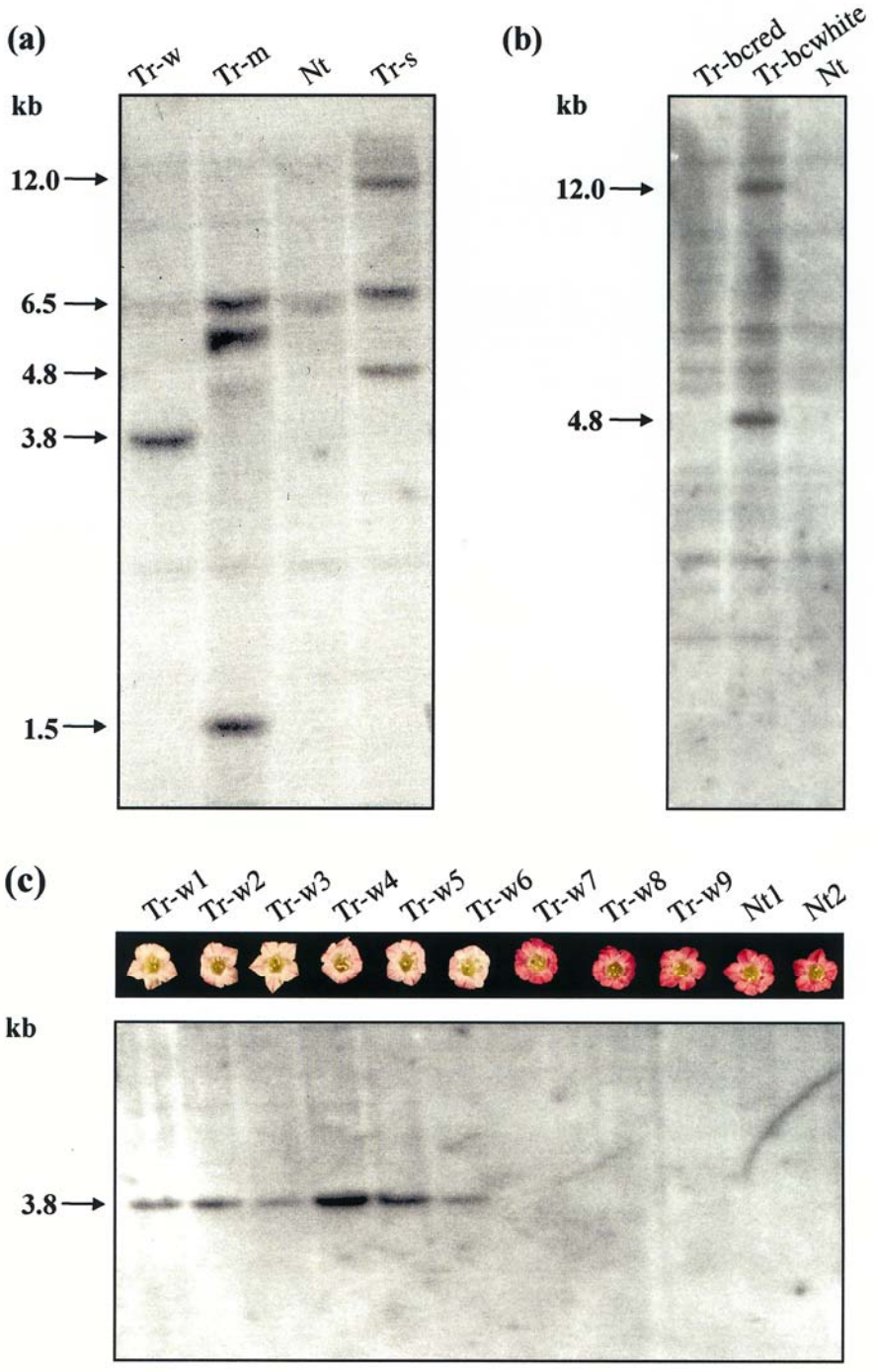


Figure 4 (previous page) Correlation between T-DNA insertions and the *FaMYB1* phenotype. **(a)** DNA gel blot of primary transformants and a non-transformed line of genomic DNA digested with *Xba*l, which cuts in the T-DNA on one side of the *FaMYB1* fragment. The *FaMYB1* coding region fragment was used as a probe. Different amounts of T-DNA copies were detected in the *FaMYB1* lines, while no T-DNA was present in the non-transformed line. Numbers at the left indicate the lengths of marker DNA fragments in kilobases. **(b)** Gel blot of genomic DNA digested with *Xba*l from two backcross lines (Tr-s crossed with non-transformed) showing the wild type phenotype (Tr-bcred) and the *FaMYB1* phenotype (Tr-bcwhite) and a line from a progeny of selfed non-transformed line (Nt). The *FaMYB1* coding region fragment was used as a probe. Numbers at the left indicate the lengths of marker DNA fragments in kilobases. **(c)** Gel blot of genomic DNA digested with *Xba*l from nine lines (Tr-w, 1-9) out of a selfed progeny of the primary transformant showing the weak phenotype (Tr-w) and two lines from a selfed progeny of a non-transformed line (Nt1 and 2). Tr-w contains a single T-DNA insertion as can be seen in (a). The *FaMYB1* coding region fragment was used as a probe. Tr-w, 1-9 lines were selected visually for analysis according to their flower pigmentation phenotype.

To examine the inheritance of the phenotype, two lines showing different phenotypic phenomena (strong, Tr-s; weak, Tr-w) were selected. The Tr-s line was backcrossed with a non-transformed wild type line (Nt), and the Tr-w line and one Nt line (as a control) were self-fertilised. A similar phenotypic effect on flower color was also detected in the progeny (Figure 3a). We analysed the changes in anthocyanin levels and their correlation with the severity of the *FaMYB1* phenotype using HPLC. Flowers of control tobacco lines contain the anthocyanin cyanidin-3-rutinoside as their main pigment component (> 90% of all anthocyanins), formed through the flavonoid biosynthetic pathway (Figure 7). We observed clear reductions in the levels of cyanidin-3-rutinoside in: a) flowers of primary transformed lines (strong, Tr-s; medium, Tr-m and weak, Tr-w); b) flowers of Tr-s backcrossed progeny (white petal line, Tr-bcwhite) and c) flowers of Tr-w selfed progeny (Tr-w 1,2,5,6) (Figure 3b). The levels of reduction correlated very well with the phenotypic response observed by eye (Figure 4c).

HPLC Analyses Reveal Specific Reduction in the Flavonol Quercetin in the Petals of *FaMYB1* Expressing Lines

In addition to anthocyanins, tobacco flowers accumulate the flavonols quercetin and kaempferol, in the form of various glycosides. These flavonols may function as co-pigments for the anthocyanins and also as sunscreens. Together with the anthocyanins, they share the compound dihydrokaempferol as a common precursor (Figure 7).

HPLC analyses of flavonol-glycosides revealed a clear reduction in the levels of quercetin-glycosides in the primary transformant lines and in the backcrossed and selfed progeny displaying the *FaMYB1* phenotype, whilst levels of kaempferol-glycosides were not altered (data not shown). We did not detect any new flavonol glycoside in the lines showing the *FaMYB1* phenotype. Flowers of control and transgenic lines and their progeny were also compared for total kaempferol

and quercetin levels, after acid hydrolysis of extracts (Figure 3c). A marked reduction in quercetin levels was observed in the primary transformant lines (Tr-s, Tr-m, Tr-w), and in the backcrossed progeny of Tr-s (Tr-bcwhite) and selfed progeny of Tr-w (Tr-w 1,2,5,6), whilst kaempferol levels remained unchanged. Reduction in quercetin levels correlated well with the reduction in anthocyanin levels in these lines (Figure 3b).

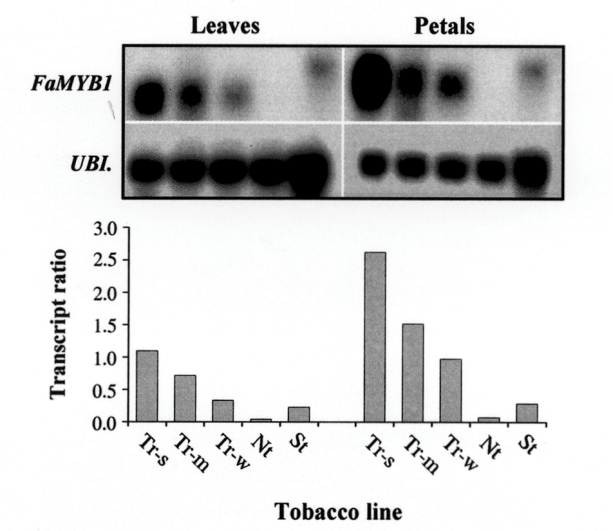


Figure 5 Correlation between *FaMYB1* expression in transformants and severity of the phenotype. Total RNA from leaves and petals (stage 4 tobacco flowers) of primary transformants (Tr-s, Tr-m and Tr-w), a non-transformed line (Nt) and ripe fruit stage strawberry tissue (St) were used for RNA gel blot analysis. The blot was hybridized with *FaMYB1* probe and then rehybridized for loading control with a petunia cDNA probe showing homology to a gene encoding a ubiquitin (UBI.) Quantification of the results was performed by measuring the OD/mm^2 (using the TINA software, Raytest, Straubenhardt, Germany) of the bands and calculating the ratios between the band in the gene specific gel blot and its corresponding band in the loading control blot (transcript ratio).

The Phenotype Correlates with *FaMYB1* T-DNA Insertions

DNA gel blot analysis was performed to confirm the transgenicity (i.e. to determine the number of T-DNA inserts) of the primary transformant lines and the backcrossed and selfed progeny of Tr-s line and Tr-w line, respectively. The Tr-w line showing the weak phenotype had a single insertion, whilst the lines showing a medium (Tr-m) phenotype and a strong (Tr-s) phenotype contained at least two (possibly three in the case of Tr-s) *FaMYB1* T-DNA insertions. Weak bands were observed in the non-transformed line (Nt), but the same bands appeared in all samples (e.g. the 6.5 kb band, see Figure 4a) and were probably due to cross-hybridisation with endogenous tobacco MYB genes.

The Tr-s backcrossed progeny line (Tr-bcred), showing full red colour petals, and the Tr-bcwhite line, showing completely white petals, were subjected to DNA gel blot analysis (Figure

4b). A clear association between the presence of the two bands corresponding to the two *FaMYB1* insertions in the Tr-s parent (12.0 kb and 4.8 kb) and petal color is shown.

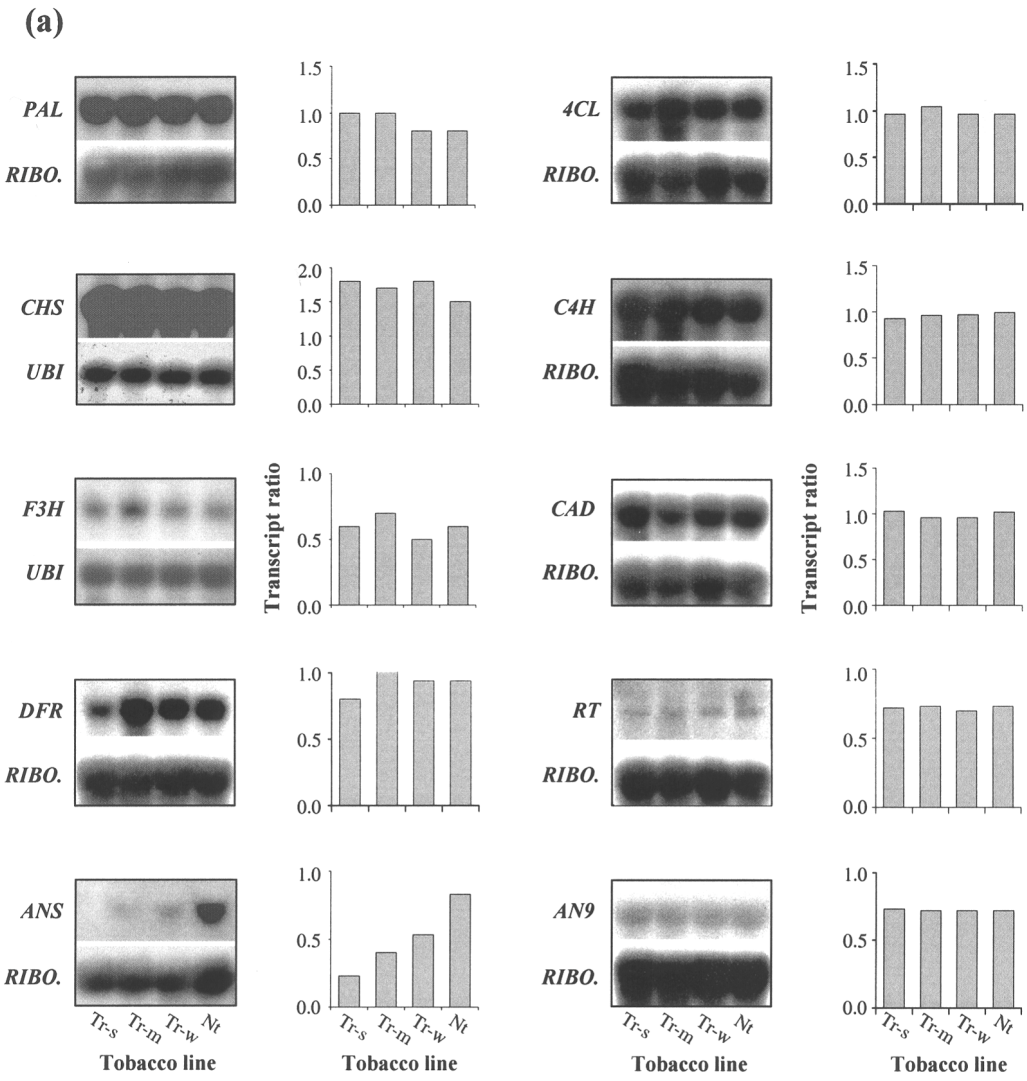


Figure 6. (previous page) Effect of *FaMYB1* overexpression on gene expression and enzymatic activity in different branches of phenylpropanoid metabolism. **(a)** Total RNA from petals (stage 4 tobacco flowers) of primary transformants (Tr-s, Tr-m and Tr-w), a non-transformed line (Nt) was used for RNA gel blot. The blots were hybridised with different probes corresponding to genes encoding the enzymes of flavonoid metabolism [CHS, F3H, DFR, ANS, GST (AN9), RT], general phenylpropanoid metabolism (PAL, C4H, 4CL) and lignin metabolism (CAD). Each of the blots was checked for RNA amounts by rehybridising with either a petunia cDNA probe showing homology to a gene encoding ubiquitin (*UBI.*) or a potato ribosomal RNA probe (*RIBO.*). Quantification of the results was performed by measuring the OD/mm^2 (using the TINA software, Raytest, Straubenhardt, Germany) of the bands and calculating the transcript ratio between the band in the gene specific gel blot and its corresponding band in the loading control blot (transcript ratio). **(b)** Flavonoid-UDP-glucose transferase enzyme activity is decreased in flowers of *FaMYB1* expressing tobacco lines. On each experimental day enzyme extracts were simultaneously prepared from flowers of control lines and *FaMYB1*-transformed lines (6 flowers per extract, 2 replicates) and tested for specific UDP-glucose transferase activity with the flavonoid substrates indicated. For each flavonoid tested, bars indicate glucosyl transferase activities in transformed lines relative to the activities in control lines (Nt = 100%) and represent means ± SD of different independent experiments (quercetin: n=3; kaempferol: n=2; cyanidin: n=1).

The segregation of the phenotype was also investigated in the Tr-w selfed progeny (Tr-w 1-9), each containing a single *FaMYB1* insertion. As was the case for the backcrossed progeny, the DNA gel blot results of the Tr-w progeny correlated well with the phenotype and demonstrated the dominant effect of *FaMYB1* (Figure 4c).

FaMYB1 expression levels in petals and leaves of primary transformed lines were investigated by RNA gel blot analysis (Figure 5). Relatively high *FaMYB1* expression levels were found in lines with the strong phenotype, and relatively low expression levels were observed in lines with the weak phenotype. We could therefore conclude that the strength of the *FaMYB1* phenotype in the transgenic lines was dependent on the relative expression levels of the introduced *FaMYB1* gene.

Gene Expression and Enzymatic Activities of Late Flavonoid Biosynthetic Pathway are Affected in the *FaMYB1* Transgenic Tobacco Lines

The effect of *FaMYB1* expression on various genes and enzymes of flavonoid, general phenylpropanoid and lignin metabolism was examined in flowers by RNA gel blots and enzymatic activity assays. Expression of chalcone synthase (*CHS*) and flavanone 3-hydroxylase (*F3H*) genes, encoding enzymes active in the upper part of the flavonoid biosynthetic pathway (up to the formation of dihydrokaempferol), were unaffected by *FaMYB1* overexpression (Figure 6a and Figure 7). In addition, no alteration of the expression of genes encoding enzymes related to general phenylpropanoid metabolism (phenylalanine ammonia-lyase, *PAL*; cinnamate 4-hydroxylase, C4H; 4-coumaroyl-CoA ligase, 4CL) and lignin metabolism (cinnamyl alcohol dehydrogenase, CAD) was observed.

In contrast, the gene encoding anthocyanidin synthase (ANS), an enzyme from the lower end of the flavonoid pathway, was significantly affected in its expression when compared with the non-transformed controls (Figure 6a). The affect on another gene from the lower part of the pathway encoding dihydroflavonol 4-reductase (DFR) is uncertain since a modest reduction in *DFR* expression was only observed in the Tr-s line.

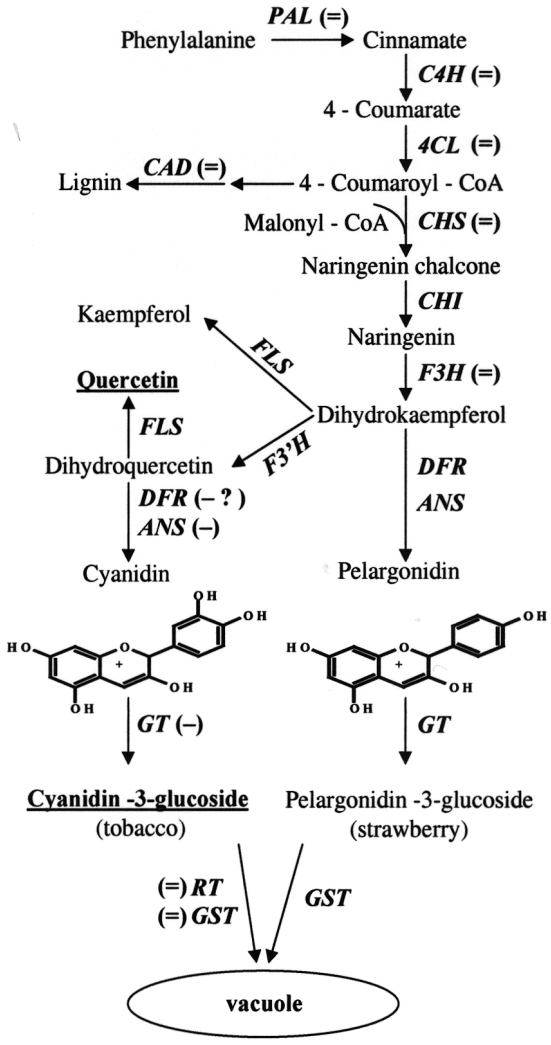


Figure 7 Effect of *FaMYB1* overexpression on metabolic flax in flavonoid metabolism. Cyanidin and pelargonidin are the main anthocyanin groups formed in tobacco flower and the cultivated strawberry fruit, respectively. Both are modified and accumulate in the vacuole. The enzymes catalyzing the reactions are: PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; CHS, chalcone synthase; CHI, chalcone flavanone isomerase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; GT, glucosyl transferase; RT, rhamnosyl transferase; GST (AN9), glutathione *S*-transferase; FLS, flavonol synthase; F3'H, flavonoid 3'-hydroxylase. Genes and enzymes showing reduced expression or activity in the *FaMYB1* overexpressing lines are marked with (-) and those that did not show any change in expression with (=). Metabolites with reduced levels in the *FaMYB1* overexpressing lines are depicted in bold and underlined.

Enzyme assays conducted on protein extracts from flowers of the Tr-s and the wild type tobacco (Nt) lines were performed in order to examine the effect on flavonoid-UDP-glucosyl transferase (GT) enzyme activity (Figure 6b). GT is also active at the lower end of the flavonoid pathway and it showed a significant and reproducible reduction in specific activity (3 to 4 folds) in the Tr-s line, compared to the NT line. This reduction in activity was observed when quercetin, kaempferol or cyanidin were used as substrates. The result implied a possible repression of *GT* gene expression. Gene expression of UDP-rhamnosyl transferase (*RT*), encoding another sugar transfer enzyme, was not altered in the *FaMYB1* expressing lines. The RT enzyme adds a rhamnose group to the anthocyanidin glucoside molecule produced by GT, to generate anthocyanidin rutinosides.

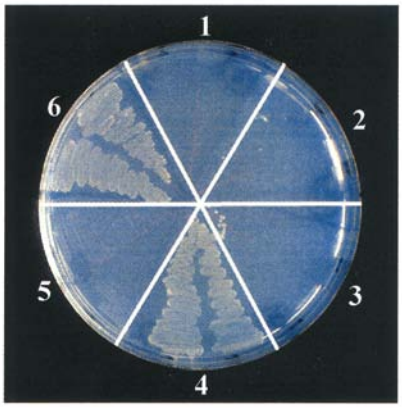
Another crucial step in anthocyanin metabolism that might be affected by *FaMYB1* is the export of anthocyanins from their site of synthesis in the cytoplasm to their site of permanent storage in the vacuole. Glutathione *S*-transferases (GSTs) are assumed to be involved in this activity, either by conjugating anthocyanin to the tripeptide GSH (γ-Glu-Cys-Gly) or by binding the anthocyanins and "escorting" them for sequestration without conjugate formation (Mueller *et al.,* 2000). Since anthocyanin-GSH conjugates were never demonstrated (both in-vitro and in-vivo), we used the common substrate for most GSTs 1-chloro 2,4-dinitrobenzene (CDNB) for analysing GST activity. The results did not show any difference in CDNB glutathionation between *FaMYB1* overexpressing lines and the controls. RNA gel blot analysis using the petunia flavonoid glutathione *S*-transferase homolog (*An9*) as a probe, showed no significant changes in *GST* expression in the *FaMYB1* transgenic lines.

Our results suggest that the main effect of *FaMYB1* overexpression is the repression of the expression of genes at the lower end of the flavonoid biosynthetic pathway, more directly related to the biosynthesis of anthocyanins and the flavonol quercetin.

FaMYB1 Interacts with Known Regulators of Anthocyanin Biosynthesis and does not Contain a Functional Activation Domain in Yeast Two-Hybrid System

In order to examine whether FaMYB1 can interact with known regulators of anthocyanin metabolism and/or act as a transcriptional activator, we performed two-hybrid assays in yeast. The yeast strain PJ69–4A was transformed with the *FaMYB1* gene coding region fused to either the Gal4 activation domain (Gal4AD) or the Gal4 DNA binding domain (Gal4BD). The coding regions of four *petunia* genes encoding known regulators of anthocyanin, namely JAF13 [Quattrocchio *et al.,* 1998; basic helix-loop-helix (bHLH) protein], AN1 (Spelt *et al.,* 2000; bHLH protein), AN2 (Quattrocchio *et al.,* 1998; MYB domain protein) and AN11 (De Vetten *et al.,* 1997; WD40 repeat protein) were fused to the Gal4AD and used for co-transformation with the Gal4BD-FaMYB1 construct. Co-transformation with the Gal4BD-FaMYB1 construct and the Gal4AD was performed, in order to test whether Gal4BD-FaMYB1 contains an activation domain.

The Gal4AD-FaMYB1 and Gal4BD were co-transformed as a control, to check whether FaMYB1 can bind to the GAL4 promoter. The results indicated that in yeast FaMYB1 does not contain a functional activation domain (Figure 8). However, FaMYB1 did interact with the two bHLH proteins (JAF13 and AN1), whilst no binding was observed between FaMYB1 and the MYB protein AN2 and the WD40 protein AN11 (Figure 8). The AN1 and JAF13 proteins were previously demonstrated to be factors necessary for anthocyanin synthesis in petunia tissues. This interaction validates the structural integrity of the FaMYB1 fusion and thus confirms that FaMYB1 was properly expressed. The Gal4AD-FaMYB1 experiment showed that FaMYB1 itself did not activate the Gal4 promoter.



1	Gal4BD	Gal4AD::FaMYB1$_{1-187}$
2	Gal4BD::FaMYB1$_{1-187}$	Gal4AD
3	Gal4BD::FaMYB1$_{1-187}$	Gal4AD::AN1$_{1-337}$
4	Gal4BD::FaMYB1$_{1-187}$	Gal4AD::JAF13$_{1-234}$
5	Gal4BD::FaMYB1$_{1-187}$	Gal4AD::AN2$_{1-255}$
6	Gal4BD::FaMYB1$_{1-187}$	Gal4AD::AN1$_{1-234}$

Figure 8 FaMYB1 interacts with known anthocyanin regulators in yeast two-hybrid assays. Results of the yeast two-hybrid assays on selection plate lacking histidine with double transformed yeast strain PJ96-4A are shown. Two independent transformants were used for each of the six construct combinations (shown below). Construct combinations 1 and 2 were performed in order to check whether FaMYB1 can bind to the GAL4 promoter and if FaMYB1 contains a functional activation domain in yeast, respectively.

DISCUSSION

In this paper we have shown that constitutive overexpression of *FaMYB1*, a MYB transcription factor isolated from red ripe strawberry fruit, in tobacco resulted in severe phenotypic changes in the transformed lines. HPLC analysis revealed a dramatic decrease in the levels of anthocyanins and of the flavonol quercetin in the flowers. The phenotype was inherited as a dominant trait to the progeny and its severity correlated well with *FaMYB1* transgene expression levels. This suggested that the phenotype was in direct correlation with the cellular concentration of *FaMYB1*.

The expression of genes encoding enzymes catalyzing reactions in the upper part of the flavonoid pathway, general phenylpropanoid and lignin metabolism were unaffected by *FaMYB1*

overexpression. However, expression of the *ANS* gene and activity of flavonoid-UDP-glucose transferase enzyme which suggests a possible reduction in expression of the *GT* gene, both active in the lower end of the flavonoid pathway were significantly reduced in all transgenics showing the phenotype .

The specific reduction in quercetin compared to kaempferol in the *FaMYB1* overexpressing tobacco lines suggests a reduction in the activity of the enzyme flavonoid 3'-hydroxylase (F3'H). F3'H catalyses the reaction forming dihydroquercetin, which is the direct precursor of both quercetin and cyanidin (see Figure 7). Our attempts to analyse F3'H gene expression and enzyme activity proved unsuccessful and as a result we could not verify whether F3'H is affected in the *FaMYB1* expressing plants. These results imply that the strawberry transcription factor FaMYB1 affects several metabolic steps at the lower end of the flavonoid biosynthetic pathway, by reducing mRNA levels of genes encoding enzymes involved in anthocyanin and flavonol biosynthesis.

Besides the phenotypic effect in the FaMYB1 overexpressing plants, several other lines of reasoning have lead us to believe that *FaMYB1* may function in strawberry as a transcriptional regulator of genes related to flavonoid metabolism. Many MYB proteins from different plant species have previously been shown to take part in the regulation of the phenylpropanoid metabolism and branch pathways, like the flavonoid biosynthetic pathway (Weisshaar and Jenkins, 1998). In the flavonoid pathway itself, MYB proteins have been identified as key regulators of the metabolism of different groups of compounds such as phenolic acids, flavonols and the pigments anthocyanin and phlobaphenes (Martin and Paz-Ares, 1997). *FaMYB1* showed the highest expression levels during the ripening stage of strawberry fruit, when anthocyanin levels reach their maximum. The importance of anthocyanin biosynthesis at this stage of development was further supported by the fact that out of 1100 EST's from a red ripe fruit library, we identified six putative genes encoding enzymes involved in the metabolism of anthocyanins (Aharoni *et al.,* 2000). Identifying a putative regulator for the flavonoid pathway (represented twice in our EST collection) was therefore not unexpected.

Additional evidence for the association of FaMYB1 with anthocyanin metabolism arose from two-hybrid assays in yeast, which indicated that FaMYB1 could interact with known bHLH anthocyanin regulators previously identified in *petunia*, AN1 and JAF13. Both AN1 and JAF13 are involved in the regulation of the lower part of the flavonoid pathway, although they are functionally and evolutionary distinct (different affinities for partner proteins and/or target DNA sequences) (Spelt *et al.,* 2000). The presence of all the six amino acid residues required for the interaction with the known bHLH anthocyanin regulator R in the R3 MYB repeat of FaMYB1, in particular the glycine-arginine pair (GR), strengthens the possibility that the interaction observed in yeast two-hybrid system in this study may occur *in planta*. However, the interaction with the R regulator might not always associate a protein to a role in flavonoid metabolism, as in the case of the *Arabidopsis* GL1 protein that interacts with R, but does not play a role in flavonoid biosynthesis (Payne *et al.,* 2000).

DNA binding transcriptional repressors act by a variety of mechanisms, including repression mediated by passive steric hindrance mechanisms or active repression (Hanna-Rose and Hansen, 1996). Active repression is thought to involve inhibitory protein-protein interactions with components of the basal transcriptional machinery (e.g. direct repression model) or positive transcriptional regulators (e.g. quenching model), or possibly induction of inactive chromatin structure at the regulated promoter. It might also involve recruitment of additional proteins as co-activators and co-repressors. In contrast to activators, each repressor generally contains a single small repression domain. Each type of repression motif might represent a unique interaction surface for contacting a particular target within the basal or regulatory transcriptional machinery, thus resulting in repression.

Information on repression mechanisms in plants is limited (Schwechheimer and Bevan, 1998) and the only R2R3 MYB protein identified as a repressor, containing a putative repressor domain, was just recently reported (AtMYB4; Jin *et al.,* 2000). AtMYB4 was suggested to act both by direct repression and by competition with activators on binding motifs located on the promoters of target genes. Both the C-terminus of the protein and a region in it containing the peptide sequence NLELRISLPDDV were shown to be required for the repression activity. The same motif (pdLNL$^{D}/_{E}$Lxi$^{G}/_{S}$) is also present in the C-terminus of FaMYB1 and other members of the MYB proteins sub-group 4 (Kranz *et al.,* 1998).

Structural homology between MYB proteins from different plant species might point to general similarity in the pathways regulated by them and in the type of regulation (activation or repression). This was shown recently for the *Arabidopsis* AtMYB75 (termed PAP1; Borevitz *et al.,* 2000) that was demonstrated to act as activator of anthocyanin biosynthesis genes, as predicted earlier from its sequence similarity to known anthocyanin biosynthesis regulators from other species (both at the MYB domain and the C-termini) (Quattrocchio *et al.,* 1999; Jin *et al.,* 1999). The overall structural homology between FaMYB1 and other MYB proteins from different plant species containing the pdLNL$^{D}/_{E}$Lxi$^{G}/_{S}$ motif and the fact that several of them were proposed to act as transcriptional repressors, suggests that FaMYB1 might function as a true repressor in strawberry.

Several mechanisms could be proposed to explain how FaMYB1 acts to reduce transcription of late flavonoid biosynthesis genes in tobacco. One possibility is that FaMYB1 is a repressor, acting similarly to AtMYB4 and recognizing its normal target genes, which are different than the ones recognized by AtMYB4 (i.e. different target site selectivity). The comparison between the DNA binding domains of FaMYB1 and AtMYB4 (Figure 1b) shows that 90 residues out of 104 are identical or contain conserved amino acid substitutions and 14 have non-conserved substitutions, of which two are amongst the four amino acids identified as necessary for protein interaction. The remaining 12 residues might be important for the difference in target site selectivity between the FaMYB1 and AtMYB4 proteins. In this case *ANS*, *GT* and possibly *DFR* might be the primary target sites for FaMYB1. Another possibility could be that FaMYB1, in high concentrations, might

bind non-specifically to target sites without capability to activate gene transcription and therefore act by hindering the function of a tobacco regulator. Repression might also be a non-direct effect in which FaMYB1 binds nonspecifically to another tobacco transcription factor, that would be inhibited in its function (Tamagnone *et al.,* 1998). Another indirect effect would be the titration of an endogenous bHLH protein, which in similarity to the R protein from maize, normally interacts with an endogenous tobacco MYB counterpart and controls expression of anthocyanin biosynthesis genes (Glover *et al.,* 1998; Payne *et al.,* 1999).

Microarray gene expression data showed that transcripts of *ANS* and *GT* accumulate co-ordinately during strawberry ripening and attain their maximum at the turning stage of development (Aharoni *et al.,* 2000). No significant difference in expression of these genes between the turning and red stages of development could be detected. Repression of anthocyanin and flavonol-related genes would be an efficient way to control metabolic flux throughout the phenylpropanoid pathway during strawberry fruit development. Reducing the levels of anthocyanins in late strawberry fruit ripening (when fruit are entirely red), would be beneficial in negating their potential toxic effects on the plant cell (Mueller *et al.,* 2000) and would substantially reduce carbon and energy consumption needed for their biosynthesis and transport into the vacuole.

Co-ordinated reduction in expression of anthocyanin pathway genes upon late development was demonstrated in flowers of *petunia* (Mur, 1995). In corollas of red colored petunia flowers, flavonoid gene expression starts at a very early stage and declines around anthesis. Interestingly, in the same tissue the expression of *PhMYB27*, encoding a MYB protein homolog, shows high levels of expression when flavonoid gene expression ceases. Apart from being a relatively short protein (184 amino acids) with no obvious acidic activation domain, PhMYB27 shows homology to other members of the MYB subgroup 4 proteins (Mur, 1995). It also contains a complete LlsrGIDPx$^{T}/_{S}$HRx$^{I}/_{L}$ motif and a partial pdLNL$^{D}/_{E}$Lxi$^{G}/_{S}$ motif outside the MYB domain. *PhMYB27* was shown to be expressed in cells that accumulate flavonols or anthocyanins including leaves, seed coat, seedlings and different floral tissues. Further evidence that PhMYB27 is involved in the regulation (possibly in repression) of flavonoid biosynthesis was supported by the fact that it is controlled by the anthocyanin regulatory loci *An1*, *An2*, *An4* and *An11*. Interestingly, AN1 (shown in this study to interact in yeast with FaMYB1) directly activates the expression of *dfrA* gene and of *PhMYB27* (Spelt *et al.,* 2000).

It is possible that FaMYB1 and PhMYB27 may have a similar function that is they may act as transcriptional repressors of late flavonoid biosynthesis genes in strawberry fruit and *petunia* flowers, respectively. One future way to verify this would be to overexpress or downregulate *FaMYB1* in strawberry and *PhMYB27* in a red colored petunia variety.

It might be also of interest to analyze the putative repression motif (present in MYB subgroup 4, including in *FaMYB1* and PhMYB27; pdLNL$^{D}/_{E}$Lxi$^{G}/_{S}$) by introducing it to the C-terminus of a known anthocyanin transcriptional activator, and analyzing the overexpression effect in transgenic plants.

EXPERIMENTAL PROCEDURES

Plant Material and Tobacco Transformation

We used the strawberry (*Fragaria x anannasa*) cultivar *Elsanta* and tobacco (*Nicotiana tabacum*) cultivar *SR1*. Transformation of tobacco was performed as described by Horsch *et al.,* (1985) using the *Agrobacterium tumefaciens* strain AGL0 (Lazo *et al.,* 1991). Plants were kept in a containment greenhouse with a 16 h photoperiod and a 21°C/17°C day/night temperature.

FaMYB1 Overexpression Construct

A 736 bp fragment including the *FaMYB1* ORF was inserted in a sense orientation to the pFLAP10 sub-cloning vector using *Xba*l and *Nco*l restriction sites introduced to the fragment by PCR at the 5' and 3' ends, respectively. The fragment was inserted between a double 35S-CaMV promoter and a nopaline synthase terminator. From the pFLAP10 vector the fragment was excised with *Pac*l and *Asc*l restriction digestions and introduced to the pBin+ binary vector containing a kanamycin resistance gene inside the T-DNA for selection of transformants (Engelen *et al.,* 1995).

Isolation of Nucleic Acids from Strawberry and Tobacco and Gel Blot Analyses

Total RNA was isolated from different strawberry tissues as described by Schultz *et al.* (1994) and from tobacco leaf and petal material as described by Verwoerd *et al.* (1989). For the isolation of genomic DNA from tobacco we used the protocol described by Doyle and Doyle (1990). For DNA gel blot analyses, aliquots of 5 μg tobacco genomic DNA were digested with *Xba*l (cuts once outside the *FaMYB1* fragment in the T-DNA) and separated on a 0.7% TAE (0.04 M Tris-acetate pH = 8.0 and 1 mM EDTA pH 8.0) agarose gel. The DNA was then denatured in 0.4 M NaOH for 30 min and transferred to a Hybond N^{+} membrane (Amersham) in 0.4 M NaOH. After fixation (2 h at 80°C) blots were hybridised as described by Angenent *et al.* (1992). The hybridization probes were made by random labelling oligonucleotide priming (Feinberg and Vogelstein, 1984) of the entire *FaMYB1* cDNA. Washing was performed under low stringency conditions [52° C with 2 times half an hour 2 X SSC (1 X SSC is 0.15 M NaCl, 0.015 M sodium citrate) and 0.1% SDS] and further under stringent conditions (65°C with 2 times 30 min 0.1 X SSC and 0.1% SDS). RNA

gel blot analysis (strawberry and tobacco) were performed as described previously (Aharoni *et al.,* 2000), except that the region outside the *FaMYB1* MYB domain was used as a probe in order to avoid cross-hybridisation. For analysing expression of different flavonoid genes we used cDNAs of tobacco *C4H* and *4CL*, petunia *PAL*, *An9*, *RT, F3'H*, *ANS*, *DFR*, tomato *F3H* and strawberry *CHS*, *CAD* as probes.

Enzymatic Activity Assays

For each experiment, enzyme extracts were prepared by isolating and pooling petal limbs from 6 flowers [developmental stages 2, 3, and 4 (see Figure 3a); 2 flowers per stage per plant, in triplicate]. For flavonoid-UDP-glucosyl transferase activity extracts were prepared by grinding flower tissue in ice-cold 0.2 M potassium phosphate buffer (pH 7.5) containing 10 mM DTT, 0.5 g DOWEX 1 X 2-100 anion exchanger and quartz sand. After centrifugation (10 min, 14000 g), enzyme reactions were started by mixing up to 150 µl of supernatant with 25 µl of 4 mM flavonoid substrate (dissolved in ethanol) and 25 µl of 12 mM UDPG. Reactions were incubated for 30 min at 30°C and stopped by adding 800 µl of a chloroform/methanol (2:1) mixture. Flavonoid glycosides in the aqueous layer were detected by HPLC (see below). For glutathione-S-transferase activity petal tissues were grinded with quartz sand in 5 ml of 0.1 M sodium phosphate (pH 7.4) containing 1 mM EDTA, and 0.25 g polyphenylpolypyrolidone. After centrifugation (10 min, 17500 g) extracts were desalted using Sephadex G-25. Enzyme activity was recorded at 30° C, in a 1 ml cuvet with 1 mM GSH and 1 mM chloro-2,4-dinitrobenzene, at pH 6.5.

HPLC Analyses of Flavonoids

Petals limbs of mature flowers (at least 3 per line) were used. Flavonoids were determined as aglycons or as their glycosides, by preparing hydrolyzed and non-hydrolyzed extracts, respectively. Hydrolyzed extracts were prepared by heating 0.15 g of petal tissue in 2 ml of an acid aqueous methanol (MeOH, HPLC quality) solution, consisting of 50% MeOH, 0.16% ascorbic acid, 0.16% t-butylhydroquinon (TBHQ) and 1.2 M HCl, at 90°C for 1 h. After hydrolysis, the extracts were diluted with 2 ml of 100% MeOH and sonicated for 5 min. Non-hydrolyzed extracts were prepared by soliciting 0.15 g of petal tissue in 1.5 ml of 75% MeOH, 0.1% trifluoroacetic acid (TFA) and 0.1% TBHQ for 15 min. All extracts were filtered over a 0.2 µm Teflon filter before analysis by HPLC. Flavonoids were separated on a 150 X 3.9 mm NovaPak C18 column (Waters Chomatography), using isocratic conditions (25% acetonitril in 0.1% TFA; flow rate 0.9 ml/min) for flavonol aglycons, and a gradient of 5 to 50% acetonitril in 0.1% TFA for anthocyanins and flavonol glycosides, and detected with a photodiode array detector (Waters 996).

Yeast Two-Hybrid Screen

The PJ69-4a yeast strain was used and it possesses the genotype: MATa trp-901 leu2-3, 112 ura3-52 his3-200 gal4D gal80D GAL-ADE2 LYS::GAL1-HIS3 met::GAL7-lac Z (James *et al.,* 1996). All yeast transformations were performed according to the Lithium Acetate method (Gietz and Schiestl, 1995). The ORFs of *AN1, AN2, AN11 and JAF13* genes (containing the *EcoR*l and *Xho*l restriction sites generated by PCR) were introduced in frame with the GAL4 activation or binding domains in the pAD and pBD vectors respectively (Stratagene). The *FaMYB1* ORF was introduced to the same vectors as described above using an *Mfe*l and *Xho*l restriction sites introduced by PCR at the 5' and 3' ends respectively. The yeast transformants were tested for interaction/activation on media without histidine or histidine and adenine.

Sequence Analyses

Sequencing was performed using the Applied Biosystems (Foster City, CA) dye terminator cycle sequencing Ready Reaction kit and the Applied Biosystems 310 DNA sequencer. Comparison and analyses of the sequences was conducted with the advanced basic local alignment search tool, BLAST (Altschul *et al.,* 1990) and the National Center for Biotechnological Information (www.ncbi.nlm.nih.gov) non-redundant protein database. Software used for DNA and protein analysis was the DNASTAR program (DNASTAR Inc. Madison, WI).

ACKNOWLEDGEMENTS

We are grateful to David Weiss, Cathie Martin and Francesca Quattrocchio for providing the cDNA clones; Arnoud Bovy for providing the pFLAP10 vector; P. James for the PJ69-4A yeast strain; Dirk Bosch for comments and advices on the manuscript; and Geert Scholten for his care of the plants.

REFERENCES

Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J., Sun, Z.K., Alvarez Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A., De Vos, R.C.H., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J. and AP, O'Çonnell. (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* **12**, 647-661.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Angenent, G.C., Busscher, M., Franken, J., Mol, J.N. and van Tunen, A.J. (1992) Differential expression of two MADS box genes in wild-type and mutant petunia flowers. *Plant Cell* **4**, 983-993.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A. and Lamb, C. (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **12**, 2383-2393.

Cheng, G.W. and Breen, P.J. (1991) Activity of phenylalanine ammonia-lyase (PAL) and concentrations of anthocyanins and phenolics in developing strawberry fruit. *J. Am. Soc. Hortic. Sci.* **116**, 865-869.

Coffman, J.A., Kirchhamer, C.V., Harrington, M.G. and Davidson, E.H. (1997) SpMyb functions as an intramodular repressor to regulate spatial expression of *Cylla* in sea urchin embryos. *Development* **124**, 4717-4727.

De Vetten, N., Quattrocchio, F., Mol, J. and Koes, R. (1997) The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev.* **11**, 1422-1434.

Dixon, R.A. and Steele, C.L. (1999) Flavonoids and isoflavonoids - a gold mine for metabolic engineering. *Trends Plant Sci.* **4**, 394-400.

Doyle, J.J. and Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**, 13-15.

Engelen, F.A.v., Molthoff, J.W., Conner, A.J., Nap, J.P., Pereira, A. and Stiekema, W.J. (1995) pBINPLUS: an improved plant transformation vector based on pBIN19. *Transgenic Res.* **4**, 288-290.

Feinberg, A.P. and Vogelstein, B. (1984) "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Addendum. *Anal. Biochem.* **137**, 266-267.

Franken, P., Schrell, S., Peterson, P.A., Saedler, H. and Wienand, U. (1994) Molecular analysis of protein domain function encoded by the myb-homologous maize genes C1, Zm 1 and Zm 38. *Plant J.* **6**, 21-30.

Gietz, R.D. and Schietzl, R.H. (1995) Transforming yeast with DNA. *Methods Mol. Cell Biol.* **5**, 255-269.

Glover, B.J., PerezRodriguez, M. and Martin, C. (1998) Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. *Development* **125**, 3497-3508.

Grotewold, E., Sainz, M.B., Tagliani, L., Hernandez, J.M., Bowen, B. and Chandler, V.L. (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. Proc. Natl. Acad. Sci. USA **97**, 13579-13584.

Goff, S.A., Cone, K.C. and Fromm, M.E. (1991) Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes Dev.* **5**, 298-309.

Hakkinen, S., Heinonen, M., Karenlampi, S., Mykkanen, H., Ruuskanen, J. and Torronen, R. (1999) Screening of selected flavonoids and phenolic acids in 19 berries. *Food Res. Int.* **32**, 345-353.

HannaRose, W. and Hansen, U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* **12**, 229-234.

Horsch, R.B., Fry, J.E., Hoffmann, N.L., Eichholtz, D., Rogers, S.G. and Fraley, R.T. (1985) A simple and general method for transferring genes into plants. *Science* **227**, 1229-1231.

James, P., Halladay, J. and Craig, E.A. (1996) Genomic libraries and a host strain designed for highly efficient two hybrid selection in yeast. *Genetics* **144**, 1425-1436.

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., Tonelli, C., Weisshaar, B. and Martin, C. (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO J.* **19**, 6150-6161.

Jin, H. and Martin, C. (1999) Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.* **41**, 577-585.

Kranz, H.D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R.C., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., Smeekens, S., Tonelli, C., Paz Ares, J., Weisshaar, B. and Jin, H.L. (1998) Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana. *Plant J.* **16**, 263-276.

Larkin, J.C., Oppenheimer, D.G., Pollock, S. and Marks, M.D. (1993) Arabidopsis GLABROUS1 gene requires downstream sequences for function. *Plant Cell* **5**, 1739-1748.

Lazo, G.R., Stein, P.A. and Ludwig, R.A. (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. *Bio/Technology* **9**, 963-967.

Martin, C. and Paz Ares, J. (1997) MYB transcription factors in plants. *Trends Genet.* **13**, 67-73.

Mol, J., Grotewold, E. and Koes, R. (1998) How genes paint flowers and seeds. *Trends Plant Sci.* **3**, 212-217.

Mueller, L.A., Goodman, C.D., Silady, R.A. and Walbot, V. (2000) AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiol.* **123**, 1561-1570.

Mur, L. (1995) Characterization of members of the *myb* gene family of transcription factors from *petunia hybrida*. PhD dissertation (Amsterdam, The Netherlands: Vrije Universiteit).

Payne, T., Clement, J., Arnold, D. and Lloyd, A. (1999) Heterologous myb genes distinct from GL1 enhance trichome production when overexpressed in Nicotiana tabacum. *Development* **126**, 671-682.

Payne, T., Zhang, F. and Lloyd, A.M. (2000) *GL3* encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1. *Genetics* **156**, 1349-1362

Paz-Ares, J., Ghosal, D. and Saedler, H. (1990) Molecular analysis of the C1-I allele from Zea mays: A dominant mutant of the regulatory C1 locus. *EMBO. J.* **9**, 315-321.

Perkins Veazie, P. (1995) Growth and ripening of strawberry fruit. *Hort. Rev.* **17**, 267-297.

Quattrocchio, F., Wing John, F., Van Der Woude, K., Mol Joseph, N.M. and Koes, R. (1998) Analysis of bHLH and MYB domain proteins: Species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. *Plant J.* **13**, 475-488.

Quattrocchio, F., Wing, J., Woude, K.v.d., Souer, E., Vetten, N.d., Mol, J., Koes, R., van der Woude, K. and de Vetten, N. (1999) Molecular analysis of the anthocyanin2 gene of Petunia and its role in the evolution of flower color. *Plant Cell* **11**, 1433-1444.

Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol.* **3**, 423-434.

Schultz, D.J., Craig, R., Cox-Foster, D.L., Mumma, R.O. and Medford, J.I. (1994) RNA isolation from recalcitrant plant tissue. *Plant Mol. Biol. Rep.* **12**, 310-316.

Schwechheimer, C. and Bevan, M. (1998) The regulation of transcription factor activity in plants. *Trends Plant Sci.* **3**, 378-383.

Seymour, G.B., Taylor, J.E. and Tucker, G.A. (1993) *Biochemistry of fruit ripening*. London: Chapman & Hall.

Shirley, B.W. (1996) Flavonoid biosynthesis: 'new' functions for an 'old' pathway. *Trends Plant Sci.* **1**, 377-382.

Spelt, C., Quattrocchio, F., Mol, J.N.M. and Koes, R. (2000) Anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *Plant cell* **12**, 1619-1631.

Tamagnone, L., Merida, A., Parr, A., Mackay, S., Culianez Macia, F.A., Roberts, K. and Martin, C. (1998) The AmMYB308 and AmMYB330 transcription factors from Antirrhinum regulated phenylpropanoid and lignin biosynthesis in transgenic tobacco. *Plant Cell* **10**, 135-154.

Urao, T., Yamaguchi Shinozaki, K., Urao, S. and Shinozaki, K. (1993) An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell* **5**, 1529-1539.

Verwoerd, T.C., Dekker, B.M.M. and Hoekema, A. (1989) A small-scale procedure for the rapid isolation of plant RNAs. *Nucl. Acids Res.* **17**, 2362.

Weisshaar, B. and Jenkins, G.I. (1998) Phenylpropanoid biosynthesis and its regulation. *Curr. Opin. Plant Biol.* **1**, 251-257.

CHAPTER 8

Asaph Aharoni, C.H. Ric de Vos, Harrie A. Verhoeven, Chris A. Maliepaard, Gary Kruppa, Raoul Bino, and Dayan Goodenowe

Nontargeted Metabolome Analysis by Use of Fourier Transform Ion Cyclotron Mass Spectrometry

OMICS: A Journal of Integrative Biology 6 (3) *In Press* (2002)

NONTARGETED METABOLOME ANALYSIS BY USE OF FOURIER TRANSFORM ION CYCLOTRON MASS SPECTROMETRY

Advanced functional genomic tools now allow the parallel and high-throughput analyses of gene and protein expression. Although this information is crucial to our understanding of gene function, it offers insufficient insight into phenotypic changes associated with metabolism. Here we introduce a high-capacity Fourier Transform Ion Cyclotron Mass Spectrometry (FTMS)-based method, capable of non-targeted metabolic analysis and suitable for rapid screening of similarities and dissimilarities in large collections of biological samples (e.g. plant mutant populations). *Separation* of the metabolites was achieved solely by ultra-high mass resolution, *Identification* of the putative metabolite or class of metabolites to which it belongs was achieved by determining the elemental composition of the metabolite based upon the accurate mass determination; and *Relative Quantitation* was achieved by comparing the absolute intensities of each mass using internal calibration. Crude plant extracts were introduced via direct (continuous flow) injection and ionized by either electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) in both positive or negative ionization modes. We first analyzed four consecutive stages of strawberry fruit development and identified changes in the levels of a large range of masses corresponding to known fruit metabolites. The data also revealed novel information on the metabolic transition from immature to ripe fruit. In another set of experiments, the method was used to track changes in metabolic profiles of tobacco flowers overexpressing a strawberry MYB transcription factor and altered in petal color. Only nine masses appeared different between transgenic and control plants, among which was the mass corresponding to cyanidin-3-rhamnoglucoside, the main flower pigment. The results demonstrate the feasibility and utility of the FTMS approach for a non-targeted and rapid metabolic "fingerprinting" which will greatly speed up current efforts to study the metabolome and derive gene function in any biological system.

INTRODUCTION

High-throughput technologies are now being established to perform large-scale analyses of gene and protein expression (Brown and Botstein, 1999; Chee et al., 1996; Brenner et al., 2000; Dutt and Lee, 2000), DNA-protein and protein-protein interactions (Uetz et al., 2000; Ren et al., 2001), protein activity (Martzen et al., 1999) and for generating large collections of mutants (Weigel et al., 2000; Winzeler et al., 1999; Parinov and Sundaresan, 2000). In genetic based strategies, mutant populations are often phenotypically screened for either morphological differences or for changes in specific metabolites. It is therefore likely that a large set of metabolic phenotypes is overlooked (i.e. "silent" mutants). Consequently, comprehensive and non-targeted, metabolite phenotyping technologies are required to bring understanding to the large amounts of data being collected (Trethewey et al., 1999; Fiehn et al., 2001).

Separation, identification, and quantification are the most fundamental requirements of all analytical methods. The primary challenge for non-targeted metabolome analysis is to meet these requirements simultaneously for all metabolites in the metabolome. The second and perhaps more difficult challenge, is to be able to meet these requirements with sufficient throughput and reproducibility such that it can be used along side other profiling methods. At present most of the analytical chemistry tools used for biological studies are optimized for the analysis of specific compounds or classes of compounds. Normally, selective extraction, separation and detection methods are deliberately utilized in order to simplify matters by removing unwanted metabolites to provide high quality data on the remaining compounds of interest. Although comprehensive metabolic profiling approaches may have to compromise on the quality of the data obtained for the individual compounds in the profile, the information obtained should attempt to meet the above requirements (Glassbrook and Ryals, 2001). Recently, important steps towards developing such an approach were reported for plants employing gas-chromatography/mass spectrometry (GC/MS)(Roessner et al., 2000; Fiehn et al., 2000). Fiehn et al. used *Arabidopsis thaliana* leaf extracts, and automatically quantified 326 distinct compounds, assigning a chemical structure to half of them. The GC/MS approach was also used for studying metabolism in potato tuber tissues derived from either transgenic plants, or plants exposed to different environmental conditions (Roessner et al, 2001a; Roessner et al., 2001b). They used statistical methods such as hierarchical cluster analysis and principal components analysis for assigning metabolic clusters to the individual plant systems examined, and determine the distance between them in term of metabolic activity.

This report introduces a new approach to perform metabolic analysis in any given biological sample. Non-targeted metabolic analysis was performed on strawberry fruit and tobacco flower extracts using Fourier Transform Ion Cyclotron Mass Spectrometry (FTMS). Since metabolites of different empirical formulas have different masses, very high mass resolution (>100,000) is required to resolve them. Currently, FTMS is the only MS system capable of routinely achieving

this level of resolution with a sufficiently fast data acquisition rate (100-1000 amu scan/sec.) to allow multiple scans in a relatively short time frame (1-2 min.). By the use of FTMS *Separation* of the metabolites was achieved solely by ultra-high mass resolution, eliminating the need for time consuming chromatography and derivatization. *Identification* of the putative metabolite or class of metabolites to which it belongs was achieved by determining the elemental composition of the metabolite based upon the accurate mass determination; and *Relative Quantitation* was achieved by comparing the absolute intensities of each mass using internal calibration.

In this study we describe the use of the FTMS approach for metabolic screening of differentially expressed metabolites. In the first experiment we analyzed the ripening shift in strawberry fruit during 4 consecutive steps of development. This experiment was performed in order to evaluate the capacity of the system to detect changes in masses of a large array of different plant primary metabolites (e.g. amino acids, fatty acids, carbohydrates) and secondary metabolites (e.g. flavonoids and terpenoids). Fruit development and maturation is a complex biological process that provides a good experimental system to investigate coordinated alterations to metabolic profiles. Shortly after fertilization, fleshy fruit such as strawberry enter a phase involving the expansion of the organ and the formation of seeds. Later, when seeds are set, the process of fruit maturation commences which provides a “biological bribe” to aid seeds dispersal (Grierson, 1998). In order to become attractive the fruit undergoes a dramatic shift in metabolism. This switch, referred to as ripening, correlates with the biosynthesis of a large set of different compounds such as flavour and aroma components and their precursors, production or degradation of acids and sugars which will influence sweetness, degradation of chlorophyll and accumulation of other pigments such as anthocyanins, and the accumulation of anti-pathogenic compounds and UV protectants such as various phenolic compounds.

In a second experiment we tested whether dissimilarities in plant metabolic profiles could be detected by the use of the FTMS system. In general, non-targeted metabolic analysis of transgenic plants that show down regulation or over-expression of genes like transcription factors might be most helpful in determining gene function and understanding regulatory networks controlling specific metabolic pathways. Metabolic profiles of transgenic tobacco plants affected in their flower color due to the overexpression of a strawberry transcription factor were compared to those of control plants. The results validate the capability of screening by FTMS to detect changes in levels of numerous types of metabolites in samples retaining altered metabolic phenotype. We therefore anticipate that the method presented here will be an important tool for future functional genomics research.

RESULTS

Comprehensive Metabolic Analysis With FTMS

In a first set of experiments Fourier Transform Ion Cyclotron Mass Spectrometry (FTMS) was used to profile four consecutive steps in strawberry fruit development (green, white, turning, red; Fig. 1A). To obtain a large metabolite distribution, two different extraction conditions, either 50/50 MeOH/0.1% formic acid (M50) or 100% acetonitrile (AN) were used. Samples were directly injected into the mass spectrometer. The metabolites in the extracts were ionized by four different modes of analysis [+/- electrospray (ESI) or +/- atmospheric pressure chemical ionization (APCI)] and then separated from each other based upon their masses. Figure 1B shows an example of the very high mass resolution (>100000) achieved during the FTMS analysis.

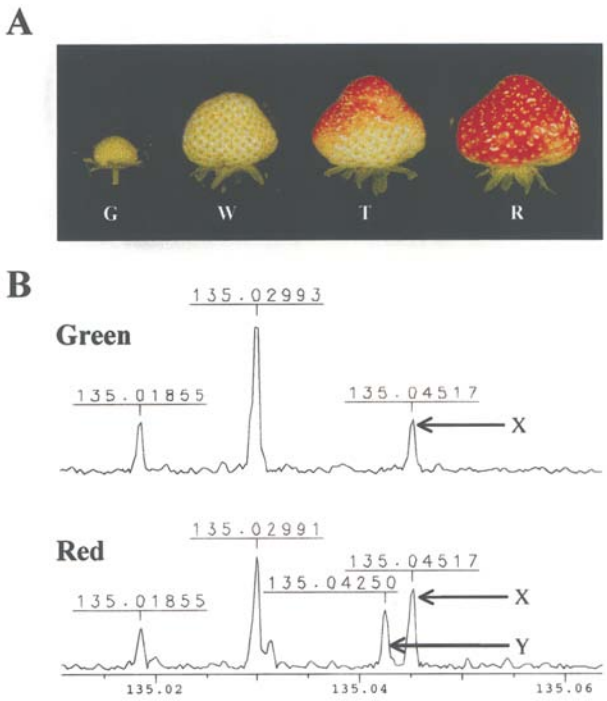


Figure 1 Non-targeted metabolic analysis in strawberry. **A,** Four consecutive stages of strawberry fruit development (G- green, W- white, T- turning, R- red) were subjected to metabolic analysis using FTMS. Similar fruit samples were used earlier to perform gene expression analysis using cDNA microarrays (Aharoni et al., 2000). **B,** An example of high resolution (>100,000) separation of very close mass peaks in data obtained from the analysis of green and red stages of fruit development. Peaks marked with an X have the same mass while peak Y is different by 3 ppm.

Using the strawberry fruit extracts, 32 samples were analyzed (4 stages of fruit development, 2 extraction methods and ionized with 4 different modes) and a unique ^{12}C mass was detected 13,412 times. Analysis of one extract ionized by a single mode resulted on average in the detection of 419 ^{12}C masses (Table 1). Further mining of the total peak demonstrated that extraction

conditions play a key role in metabolite detection. Only 16% to 36% of the masses detected were found in both the M50 and AN-extractions (Table 1). After accounting for the overlap in similar masses amongst the two extractions, a total of 5844 unique ^{12}C masses were obtained from extracts of the 4 different developmental fruit stages, ionized in the 4 ionization modes (Table 1). The 5884 masses detected include the redundancy in between the ionization modes, which was not calculated and estimated in the range of 10% (D.B. Goodenowe, unpublished data),

After the metabolites were separated, we assigned a putative identity to each of the masses observed by determining all possible empirical formulae capable of creating its accurate mass within a given error range of ± 1.0 ppm. This process is displayed in Figure 2. A very high degree of mass accuracy is required for this approach to have any reasonable utility. For example, for the observed mass 433.11270, only one potential empirical formula (naturally possible and within an error of 1.0 ppm) could be identified ($C_{21}H_{20}O_{10}[+H]^+$), however there are 27 possible empirical formulae within an error of 10.0 ppm**.** In this study out of the over 5000 unique monoisotopic (^{12}C), singly charged masses detected, a single empirical formula could be assigned to over half of them (data not shown). For example in the red stage extract, 55% of the masses were assigned a single empirical formula, 10% 2 formulae and 35% 3 or more formulae. The empirical formulae were then used to screen a commercial database representing 159,000 natural products (Chapman and Hall, Dictionary of Natural Products) for a possible identification of the metabolite.

TABLE 1. NUMBER OF MASSES OBSERVED ACROSS STAGES OF FRUIT DEVELOPMENT, IONIZATION-MODES AND EXTRACTIONS

Overall mass peaks (including redundancy)										**Unique mass peaks[c]**			
	Green		**White**		**Turning**		**Red**		**Total**	**Only**		**Both (%)[b]**	**Total**
	M50	**AN**	**M50**	**AN**	**M50**	**AN**	**M50**	**AN**	**M50 + AN**	**M50**	**AN**	**M50 + AN**	
ESI +	454	471	504	426	513	339	419	493	3619	578	520	374 (25%)	1472
ESI-	513	285	315	247	269	319	348	193	2489	608	454	208 (16%)	1270
APCI+	446	340	672	470	528	399	587	902	4344	368	595	536 (36%)	1499
APCI-	519	546	211	435	260	296	233	460	2960	492	800	311 (19%)	1603
Mean	483	411	426	395	393	338	397	512		512	592	357 (24%)	1461
Total[a]	1932	1642	1702	1578	1570	1353	1587	2048	**13,412**	2046	2369	1429 (24%)	**5844[c]**

[a]Total (non-unique) masses.

[b]Percentage of unique masses appearing in both extractions.

[c]Total unique masses across extractions, but not across ionization modes (not calculated). Thus, 5844 are the amount of unique mass peaks detected in the four stages analysed, by 2 extraction methods (M50- methanol/0.1% formic acid (50%/50%), AN- 100% acetonitrile) and 4 ionization modes (ESI +/-, electrospray ionization or APCI +/-, atmospheric pressure chemical ionization), but includes the redundancy between the ionization modes (estimated as 10%, D.B. Goodenowe, unpublished data).

We performed a replicate extraction and subsequent FTMS-analysis of the red stage strawberry fruit sample, in each of the 8 ionization/extraction method combinations. The replicates were used for an analysis of variance of the mass peaks, using the 2log intensity values (Fig. 3). Overall (in all combinations), 95% of all the differences between the red stage replicates were below 1.96-

fold, 90% of the differences below 1.72-fold, 75% below 1.41-fold, 50% below 1.21-fold and 25% below 1.09-fold. The standard deviation for variation between the replicates over all mass peaks was 1.27-fold, average difference between replicates was 1.32-fold, and the median, that is less influenced by outlying values, 1.21-fold. These relatively small deviations between replicates demonstrate the high reproducibility of both extraction procedure and FTMS analysis for the mass peaks detected.

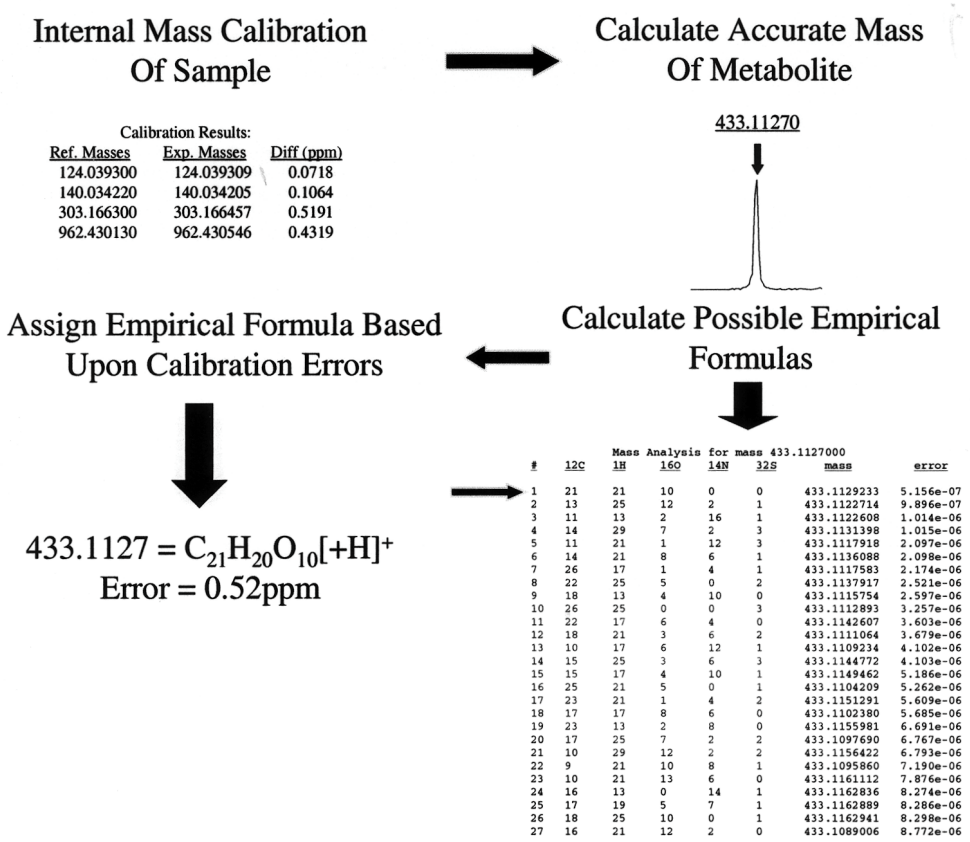


Figure 2 The procedure of assigning an empirical formula to separated mass peaks. We assigned an identity to each of the metabolites observed by determining all possible empirical formulas capable of creating its accurate mass within a given error range of ± 1.0 ppm (between the calculated mass of the formula and the experimental mass). Any mass peak that contained a chemically meaningful combination of Carbon, Hydrogen, Nitrogen, Oxygen, Phosphorus and Sulfur, with a single charge was accepted. In the example, having a mass accuracy of 1.0 ppm or less yields 2 possible formulae (only the first one is chemically meaningful), but a mass accuracy of 10.0 ppm or less results in >27 possible formulae.

Metabolic Changes in Strawberry Fruit Development

The non-targeted mass profiling data collected by FTMS analysis provided us a general picture of metabolic changes occurring during the complex process of fruit maturation (from the early green stage to the red ripe fruit). It also demonstrated the capability of such an approach to detect metabolites belonging to different types of chemical groups. Each mass in the dataset, including its

corresponding peak intensity, was used to determine differential expression of a putative metabolite in the 4 stages analysed (see Table 2). Due to the high mass accuracies we could simultaneously assign a putative identity and/or a classification to a certain chemical group to a large number of differentially expressed-masses. Generally, a possible metabolite identity could be more easily assigned to mass peaks up to m/z=300. Metabolites represented in the Table 2 were either showing a single hit in the natural product database search or were reported earlier to be produced in strawberry and other fruit or identified by us in strawberry. For example, in the case of $C_6H_{12}O_6$, glucose and fructose are depicted since they are both known to accumulate to high levels in strawberry fruit during ripening, and not other hexose sugars with the same empirical formula. The simultaneous detection of a large number of known strawberry metabolites, some of which are previously reported to accumulate during fruit maturation, in addition to other more common metabolites, within a single sample and upon a short analysis time demonstrates the power of this new approach.

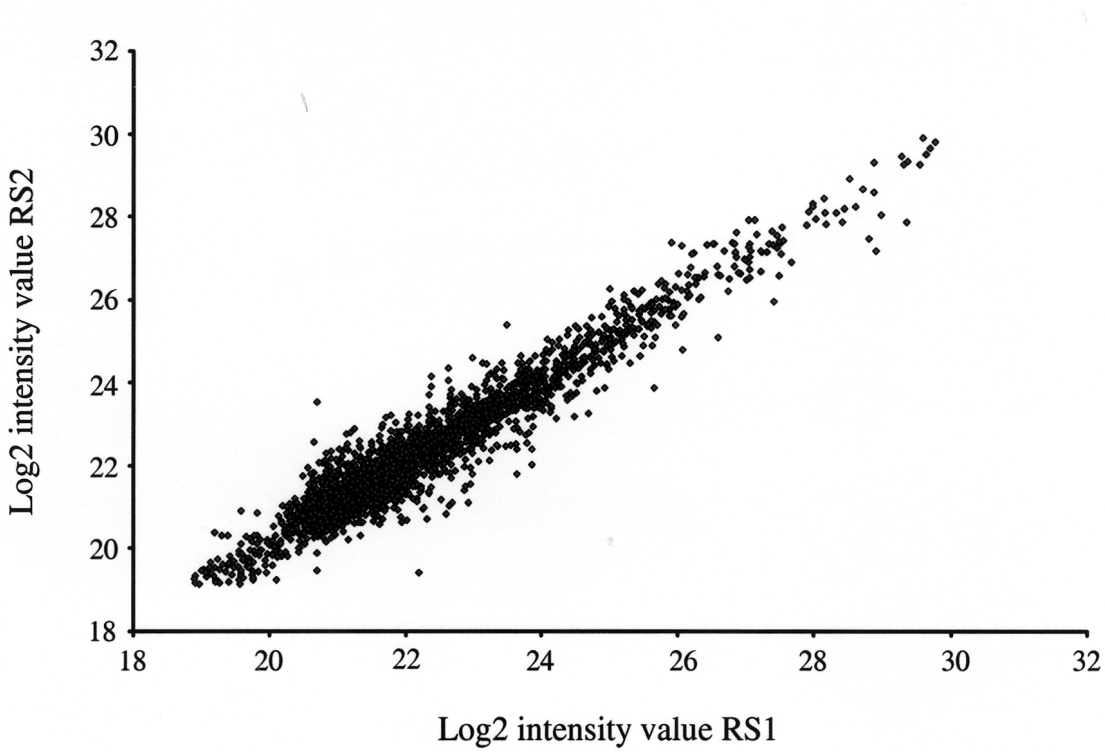


Figure 3 Reproducibility of the FTMS analysis method. Duplicate methanol/formic acid as well as acetonitrile extracts were prepared from red stage strawberries and subsequently analyzed by FTMS in each of the 4 ionization modes. Data represent areas of intensity data of mass peaks that were present (i.e. had a value above the background level) in both replicates.

Similar to other fruit, the switch from initial fruit growth to ripening in strawberry involves dramatic changes in the metabolic profiles. We divided the different metabolic profiles into 3 main groups: one set of putative metabolites showing a decrease in their levels from green to red stage

of fruit development, one showing an increase in levels from green to red and a third showing intermediate patterns of expression (Table 2, groups 1, 3 and 2, respectively).

TABLE 2. EXPRESSION PROFILES OF PUTATIVE METABOLITES IDENTIFIED DURING STRAWBERRY FRUIT DEVELOPMENT[a]

OM	IM	EX	R/G	T/G	W/G	EF	Putative Metabolite	CG
Group 1: Early expression						**Profile a**		
300.9990	AP-	AN	0.03	0.03	0.03	$C_{14}H_6O_8$	Ellagic acid	P. Acid
137.0243	AP-	AN	0.09	0.08	0.09	$C_7H_6O_3$	Hydroxybenzoic acid	P. Acid
175.1188	ES+	50	0.06	0.06	0.10	$C_6H_{14}N_4O_2$	Arginine	A. acid
167.0349	AP-	AN	0.08	0.05	0.06	$C_8H_8O_4$	Vanillic acid	P. Acid
133.0142	ES-	AN	0.04	0.04	0.04	$C_4H_6O_5$	Malic acid	O. acid
2890717	ES-	50	0.04	0.07	0.12	$C_{15}H_{14}O_6$	Catechin (epi)/leucopelargonidin	Phenolic
179.0349	AP-	50	0.16	0.14	0.11	$C_9H_8O_4$	Caffeate	P. Acid
175.0764	AP-	AN	0.13	0.13	0.12	$C_{11}H_{12}O_2$	Ethyl cinnamate	Ester
464.0882	ES-	50	0.13	0.13	0.13	$C_{21}H_{20}O_{12}$	Quercetin glucoside	Flavonoid
161.0608	AP-	AN	0.19	0.1	0.14	$C_{10}H_{10}O_2$	Methyl cinnamate	Ester
289.0707	AP+	50	0.11	0.15	0.28	$C_{15}H_{12}O_6$	Dihydrokaempferol	Flavonoid
305.0664	AP-	AN	0.37	0.37	0.37	$C_{15}H_{14}O_7$	Leucocyanidin	Flavonoid
193.0640	ES+	AN	0.22	0.22	0.22	$C_6H_{12}N_2O_3S$	*N*-Alanylcysteine	A. Acid d.
155.0338	AP+	50	0.31	0.26	0.25	$C_7H_6O_4$	Gentisic/Protocatechuic acid	P. Acid
231.0297	AP-	50	0.25	0.25	0.25	$C_{12}H_8O_5$	Malusfuran	Furan
101.0961	AP+	AN	0.25	0.25	0.25	$C_6H_{12}O$	(E)2-hexenol / hexanal	Alcohol
331.0671	AP-	50	0.23	0.23	0.23	$C_{13}H_{16}O_{10}$	Glucogallin	Phenolic
220.1180	ES+	AN	0.27	0.27	0.27	$C_9H_{17}NO_5$	Pantothenic acid	Vitamin
141.0182	ES+	AN	0.49	0.48	0.46	$C_6H_4O_4$	Dihydroxybenzoquinone	Phenolic
305.0654	AP+	50	0.42	0.42	0.42	$C_{15}H_{12}O_7$	Dihydroquercetin	Flavonoid
						Profile b		
144.1746	ES+	50	0.2	0.33	0.68	$C_9H_{21}N$	Heptylamine	Amine
193.0706	AP+	50	0.24	0.34	0.64	$C_7H_{12}O_6$	Quinic acid	O. Acid
						Profile c		
193.0342	AP+	AN	0.01	0.78	0.81	$C_6H_8O_7$	Citrate / Isocitrate	O. acid
175.0236	AP+	50	0.10	0.74	1.63	$C_6H_6O_6$	Aconitate / Dehydroascorbic acid	O. acid
147.0288	AP+	AN	0.18	0.99	0.64	$C_5H_6O_5$	Alpha-ketoglutarate	O. acid
210.0607	AP+	AN	0.19	0.76	0.67	$C_6H_{11}NO_7$	xylo-5-Hexulosonic acid	S. Acid
138.0550	ES+	50	0.43	0.64	1.26	$C_7H_7NO_2$	Anthranilate	Acid
210.1277	AP+	AN	0.30	0.85	0.88	$C_{15}H_{15}N$	*N*-Propyl carbazole	Alkaloid
Group 2: Intermediate expression						**Profile d**		
166.0863	ES+	50	0.20	0.75	0.15	$C_9H_{11}NO_2$	Phenylalanine	A. acid
132.1019	ES+	AN	0.41	0.71	0.29	$C_6H_{13}NO_2$	Leucine	A. acid
148.0604	ES+	50	0.36	0.56	0.36	$C_5H_9NO_4$	Glutamate	A. acid
118.0862	ES+	50	0.30	0.67	0.38	$C_5H_{11}NO_2$	Valine	A. acid
						Profile e		
123.0451	AP-	AN	0.9	0.9	0.11	$C_7H_8O_2$	3-methylcatechol	Phenol
147.0764	ES+	50	0.61	0.87	0.23	$C_5H_{10}N_2O_3$	Glutamine	A. acid
134.0448	AP+	50	0.64	0.72	0.44	$C_4H_7NO_4$	Aspartate	A. acid
109.0284	AP+	AN	1.18	0.62	0.33	$C_6H_4O_2$	1,4-Benzoquinone	Phenolic
175.0248	ES-	50	0.90	0.53	0.45	$C_6H_8O_6$	Ascorbic Acid	O. acid
						Profile f		
123.0440	AP+	AN	2.08	0.43	0.30	$C_7H_6O_2$	Benzoic acid	P. acid
271.2631	AP+	AN	2.01	0.37	0.33	$C_{17}H_{34}O_2$	Methyl hexadecanoate	Ester
Group 3: Late expression						**Profile g**		
433.1127	ES+	50	255.00	12.90	1.00	$C_{21}H_{21}O_{10}$	Pelargonidin-3-glucoside	Flavonoid
211.0601	AP+	50	2.35	5.37	1.19	$C_{10}H_{10}O_5$	Hydroxyferulate	Aromatic
117.0182	AP+	50	3.91	9.09	1.29	$C_4H_4O_4$	Fumarate	O. acid
179.0560	AP-	50	2.89	2.59	1.22	$C_6H_{12}O_6$	Glucose/Fructose	Sugar
307.1023	ES+	50	3.33	3.14	1.85	$C_{12}H_{18}O_9$	Galactopyranose	Sugar
149.0455	AP-	50	3.30	2.70	1.10	$C_5H_{10}O_5$	D-ribose	Sugar
						Profile h		
325.0929	AP-	AN	9.59	1.51	0.27	$C_{15}H_{18}O_8$	Coumarate glucose	Phenol
165.0545	AP+	50	3.27	0.86	0.42	$C_9H_8O_3$	4-Coumarate	P. acid

TABLE 2. (CONT'D) EXPRESSION PROFILES OF PUTATIVE METABOLITES IDENTIFIED DURING STRAWBERRY FRUIT DEVELOPMENT[a]

OM	IM	EX	R/G	T/G	W/G	EF	Putative Metabolite	CG
				Profile i				
281.2474	AP+	AN	20.80	1.93	0.97	$C_{18}H_{32}O_2$	Linoleic acid	F. Acid
255.2319	AP+	AN	24.30	1.34	0.85	$C_{16}H_{30}O_2$	Palmitoleic acid	F. Acid
283.2631	AP+	AN	32.56	1.50	0.99	$C_{18}H_{34}O_2$	Oleic acid	F. Acid
243.2319	AP+	AN	28.80	0.91	0.62	$C_{15}H_{30}O_2$	Pentadecanoic acid	F. Acid
257.2474	AP+	AN	40.20	1.67	0.80	$C_{16}H_{32}O_2$	Palmitic acid	F. Acid
229.2162	AP+	AN	29.80	1.38	1.00	$C_{14}H_{28}O_2$	3-methylbutyl nonanoate	Ester
285.2787	AP+	AN	18.03	1.87	1.24	$C_{18}H_{36}O_2$	Stearic acid	F. acid
291.1074	ES+	AN	10.04	1.00	1.00	$C_{12}H_{18}O_8$	Furaneol (R=D-glucopyranose)	Furan
377.1078	ES+	50	11.00	1.00	1.00	$C_{15}H_{20}O_{11}$	Furaneol (R=D-gluco-manosyl)	Furan
159.1380	AP+	AN	5.20	1.00	1.00	$C_9H_{18}O_2$	Nonanoic acid	F. Acid
173.1536	AP+	AN	5.49	1.00	1.00	$C_{10}H_{20}O_2$	Capric acid	F. Acid
313.3101	AP+	AN	7.85	1.00	1.00	$C_{20}H_{40}O_2$	Eicosanoic acid	F. Acid
279.2318	AP+	AN	8.52	1.97	1.10	$C_{18}H_{30}O_2$	Linolenic acid	F. Acid
149.0597	AP+	50	5.86	1.17	0.87	$C_9H_8O_2$	Cinnamate	P. Acid
201.1849	AP+	AN	5.69	1.00	1.00	$C_{12}H_{24}O_2$	1-octyl butanoate	Ester
215.2004	AP+	AN	6.67	1.00	1.00	$C_{13}H_{26}O_2$	1-methyhexyl hexanoate	Ester
227.2005	AP+	AN	6.93	0.91	0.90	$C_{14}H_{26}O_2$	Cis-hex-3-enyl octanoate	Ester
299.2944	AP+	AN	7.91	0.76	0.68	$C_{19}H_{38}O_2$	Methylstearate	Ester
209.0819	AP-	AN	5.8	1.00	1.00	$C_{11}H_{14}O_4$	Sinapyl alcohol	Phenolic
193.0494	AP+	50	5.31	1.94	0.99	$C_{10}H_8O_4$	Scopoletin	Benzopyranoid
153.1022	AP+	AN	6.79	1.00	1.00	$C_8H_{12}N_2O$	2-Isopropyl-3-methoxypyrazine	Alkaloid
263.2368	AP+	AN	6.65	1.57	0.59	$C_{18}H_{30}O$	Farnesyl acetone	Terpene
163.0389	AP+	AN	5.44	0.73	0.66	$C_9H_6O_3$	Hydroxycoumarine	Benzopyranoid
147.0441	AP+	AN	5.47	0.96	1.57	$C_9H_6O_2$	Coumarine	Benzopyranoid
269.0454	ES-	50	9.46	1.00	1.00	$C_{15}H_{11}O_5$	Pelargonidin	Flavonoid
181.1334	AP+	AN	5.33	1.00	1.00	$C_{10}H_{16}N_2O$	Smipine	Alkaloid
123.1168	AP+	AN	3.63	1.00	1.00	C_9H_{14}	Santene	Terpene
195.0651	AP+	50	3.91	1.89	1.00	$C_{10}H_{10}O_4$	Ferulate	P. Acid
225.0757	AP+	50	3.82	1.27	1.00	$C_{11}H_{12}O_5$	Sinapate	P. Acid
115.0502	AP+	AN	4.76	1.00	1.00	$C_4H_6N_2O_2$	3-cyanoalanine	A. acid
171.0288	AP+	AN	2.06	1.05	0.71	$C_7H_6O_5$	Gallic acid	P. Acid
143.0702	AP+	50	2.52	1.00	1.00	$C_7H_{10}O_3$	Furaneol (R=CH3)	Furan
129.0546	ES+	AN	4.50	1.00	1.00	$C_6H_8O_3$	Furaneol (R=H)	Furan
207.2107	AP+	AN	3.11	1.00	1.00	$C_{15}H_{26}$	4-Muurolene	Terpene
223.2057	AP+	AN	3.25	0.9	1.00	$C_{15}H_{26}O$	Nerolidol	Terpene
151.1481	AP+	AN	3.95	1.00	1.00	$C_{11}H_{18}$	Dimethyl -1,3,7- nonatriene	Terpene
137.1324	AP+	AN	4.63	1.00	0.78	$C_{10}H_{16}$	Pinene	Terpene
171.1379	AP+	AN	2.24	1.55	1.21	$C_{10}H_{18}O_2$	Dec-2-enoic acid	F. Acid
143.1006	AP+	AN	3.04	1.00	1.00	$C_8H_{14}O_2$	Oct-2-enoic acid	F. Acid
187.1692	AP+	AN	3.17	1.00	1.00	$C_{11}H_{22}O_2$	Undecanoic acid	F. Acid
165.0909	AP+	AN	2.96	1.80	0.91	$C_{10}H_{12}O_2$	2-phenylethyl acetate	Ester
151.0754	AP+	AN	3.27	1.00	1.00	$C_9H_{10}O_2$	2-methoxy-4-vinylphenol	Ester
167.0703	AP+	AN	3.32	1.00	1.00	$C_9H_{10}O_3$	Ethyl salicylate	Ester
101.0597	AP+	AN	3.30	1.00	1.00	$C_5H_8O_2$	2-methylbut-2-enoic acid	Acid
109.1011	AP+	AN	3.52	0.90	0.95	C_8H_{12}	Methyl-heptatriene	F. Acid
449.1079	ES+	50	2.14	1.00	1,00	$C_{21}H_{21}O_{11}$	Cyanidin glucoside	Flavonoid
111.0440	AP+	50	2.00	1.87	1.10	$C_6H_6O_2$	Catechol	Phenol
311.1127	ES+	AN	2.05	1.00	1.00	$C_{15}H_{18}O_7$	Cinnamate glucose	Phenol
209.0664	ES-	50	2.13	1.00	1.00	$C_7H_{14}O_7$	Sedoheptulose	Sugar
207.0661	AP-	50	2.38	1.00	1.00	$C_{11}H_{12}O_4$	Sinapaldehyde	Phenolic
179.0703	AP+	AN	2.29	0.94	1.22	$C_{10}H_{10}O_3$	Coniferaldehyde	Phenolic
154.0862	AP+	AN	2.54	1.00	1.00	$C_8H_{11}NO_2$	Dopamine	Alkaloid
143.0349	ES-	50	2.20	1.76	1.47	$C_6H_8O_4$	2-Methyleneglutarate	Acid

[a]Metabolites showing: early expression (group 1, profiles a, b, and c), intermediate expression (group 2, profiles d, e, and f), late expression (group 3, profiles g, h and i) during strawberry fruit development (see also Fig. 4A).

CG, chemical group; EF, empirical formula; OM, observed mass; IM, ionization mode; EX, extraction; P. Acid, phenolic acid; A. Acid, amino acid; A. Acid D., amino acid derivative; O. Acid, organic acid; F. Acid, fatty acid; AP+/-, APCI positive or negative; ES+/-, ESI positive or negative; 50- methanol/0.1% formic acid (50%/50%), AN- 100% acetonitrile. Examples of selected metabolic profiles that are significantly changed in the red (R), turning (T) and white (W) stages compared to the green (G) stage. Positive ratios from 2 to 5, 5 to 10 and >10 are shaded in yellow ( ), red ( ) and dark red ( ) respectively while negative ratios from 2 to 5, 5 to 10 folds and >10 are shaded in light green ( ), darker green ( ) and olive green ( ), respectively. Metabolites identified in the table were either a single hit in the natural product database search or were reported earlier to be produced in strawberry and other fruit or identified by us in strawberry.

Examples of profiles of selected putative metabolites observed for each group are shown in Figure 4A. The total number of mass changes observed in this study is summarised in Figure 4B. The least number of mass changes occurred from the turning to the white stage (504 up, 308 down). As was expected the greatest number of changes were observed when comparing the red stage to the green stage (1226 up, 1179 down). However, the greatest number of mass increases

between two successive developmental stages was between the turning and the red stage (948), whereas the greatest number of mass decreases was between the green and the white stage (1148). These mass fingerprints suggest that the green stage is most different in terms of metabolism, compared to the other 3 developmental stages analysed.

Although metabolism in early fruit development has hardly been investigated in strawberry or any other fruit, it is known that strawberry accumulates phenols such as tannins mainly in early stages (Cheng and Breen, 1991). We indeed observed relative high levels of masses possibly corresponding to these compounds, including the constituents of condensed (catechin) and hydrolysable tannins (glucogallin) in the green stage of development (Table 2). Other observed masses showing a similar profile (Table 2, group 1, profiles a and b) included several putative phenolic acids such as ellagic acid, hydroxybenzoic acid, vanillic acid and gentisic acid. Plants have the unique feature of accumulating organic acids in the cell vacuole. These simple acids can be divided into two groups: the tricarboxylic acid (TCA) cycle acids and other acids such as ascorbic acid (vitamin C). Among the TCA cycle acids, malic acid and citric acid most often accumulate in fruit, thus influencing their acidity (Seymour et al., 1993). While levels of the mass putatively corresponding to malic acid showed a dramatic decrease between the green stage and later stages (Table 2, profile a), other masses corresponding to TCA cycle acids showed a more gradual decrease during ripening (Table 2, profile c; citrate/isocitrate, aconitate and alpha-keto-glutarate). Putative metabolites showing intermediate expression profiles (Table 2, group 2) were mainly amino acids, which showed a clear reduction in levels from the green to the white stage. Some putative metabolites rise again from turning to the red stage (Table 2, profile e; e.g. glutamine and ascorbic acid) while others, after an increase from white to turning, decreased again upon further ripening (Table 2, profile d; e.g. phenylalanine and valine).

Fruit ripening is characterized by increases in levels or *de novo* production of a vast array of primary and secondary metabolites. These contribute to the marked alterations in fruit color, flavour/aroma, sweetness/acidity, texture and defense against pathogens. The most common metabolite profile observed in this study was a sharp increase between the turning and red stages of fruit development (Table 2, group 3, profile i). A large portion of the putative metabolites showing such an accumulation pattern could be associated with flavour and aroma and included an array of different ester types, alcohols, terpenoid and several derivatives of furaneol. At this stage, ripening strawberries also accumulate various types of phenolic and aromatic components including cinnamic acid and its derivatives (Table 2). Of the putative phenolic components, catechol and gallic acid might be degradation products of other phenols accumulating at earlier stages as described above. The increase in the various intermediates in monolignol biosynthesis during ripening (e.g. sinapyl alcohol and sinapaldehyde) provides additional evidence for the production of lignin, suggested to be a major constituent of the vasculature of the ripe strawberry fruit (Aharoni et al., Plant Physiology, *In press*). Color in strawberry is mainly provided by the anthocyanin pelargonidin-3-glucoside (Perkins-Veazie, 1995). From our mass profiling data it

seems that the accumulation of this flavonoid starts earlier in development than that of the flavour metabolites (between the white and turning stages; profile g) and coincides with the accumulation of hexose sugars (either glucose or fructose) and ribose.

Phenolic compounds fulfill a very broad range of physiological roles in plant tissues including fruits such as strawberry. Most plant phenols are derived from the phenylpropanoid and related pathways and possess a benzene ring that contains various attached substituents, such as carboxyl, hydroxyl and methoxyl groups, and often non-aromatic ring structures. We used the metabolism of phenols in the phenylpropanoid pathway as a model for examining the capability of our FTMS-based fingerprint method to track changes in a particular metabolic pathway. The metabolite data from this study was superimposed on previous microarray-based gene expression data acquired by analyzing comparable stages of fruit development (Aharoni et al., 2000). Figure 5 shows a schematic diagram of the phenylpropanoid and some diverging pathways. Six putative differentially expressed metabolites are represented in our FTMS data, all except phenylalanine previously reported as such in literature (cinnamate glucose, Latza et al., 1996; quinic acid, Moing et al., 2001; ellagic acid, Maas et al., 1991; condensed tannins, Cheng and Breen., 1991; pelargonidin-3-glucoside, Perkins-Veazie 1995). The only metabolite showing a double sigmoid accumulation was the aromatic amino acid phenylalanine (see Fig. 4), which is the major precursor for the phenylpropanoid and related pathways. At the green stage the relative high level of phenylalanine may support the accumulation of tannins (both hydrolysable and condensed) and some putative cinnamic acid derivatives (gentisic-, ellagic-, vanillic- and hydroxybenzoic acid). At the white stage its level is relatively low and at the turning stage it rises again. This second peak in production may be related to its role as precursor for both flavonoids (anthocyanins and flavonols), hydroxycinnamic acids, cinnamyl glucose, coumarins and monolignols, which characterize ripe fruit. Interestingly, the expression of the gene encoding the enzyme dihydroflavonol reductase (DFR) did not differ between green and red stage. Possibly, at early stages of strawberry development DFR is mainly involved in the biosynthesis of tannins, while at later stages it is involved in the biosynthesis of pelargonidin to provide the red color.

The parallels between gene and metabolite expression suggests that expression of genes encoding enzymes performing final committed steps in pathways often show direct relationships. The patterns of GT1 and pelargonidin glucoside as well as of GT2 and cinnamate glucose exemplify this relationship in strawberry. However, conclusions on such parallel patterns will be more difficult when genes and metabolites of intermediate biosynthetic steps of pathways are examined, especially when not all the branches of the pathways are known.

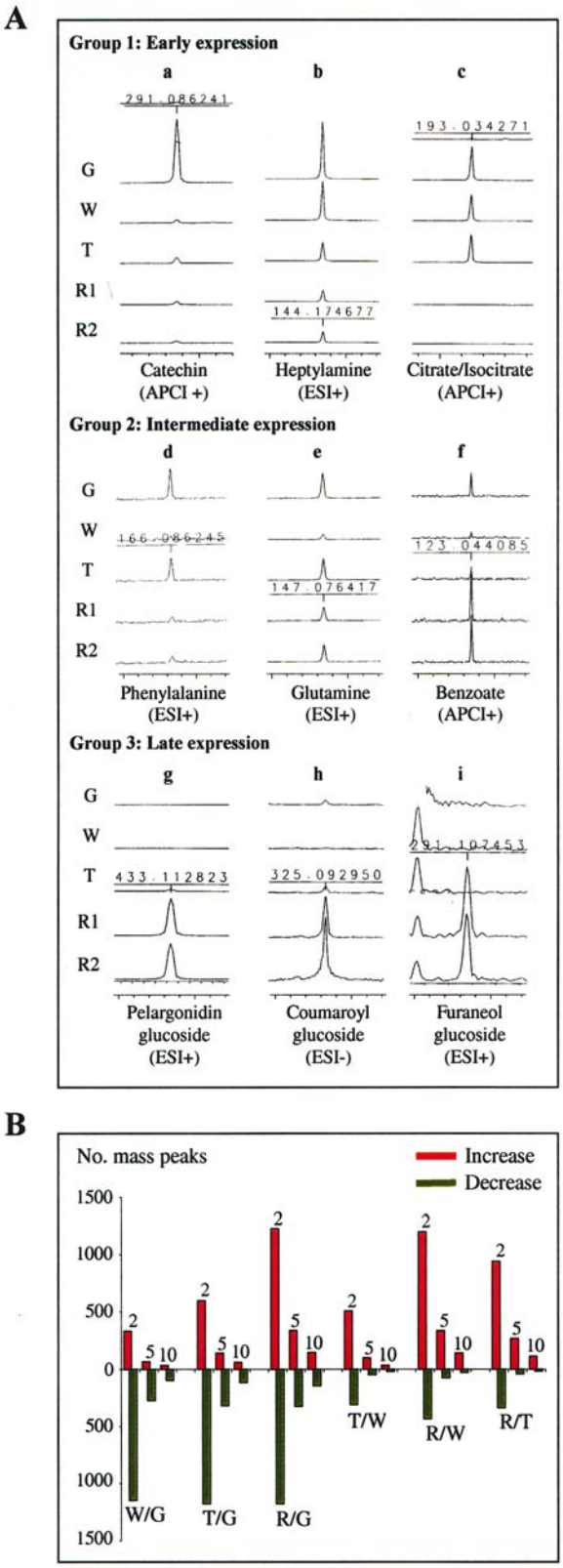


Figure 4 A range of metabolic profiles and overall metabolite changes in ripening strawberries. **A,** Examples of putative metabolites belonging to one of the 3 groups of profiles observed (see also Table 2). R1 and R2 are duplicate samples of the red stage. **B,** Numbers of masses changing in their expression levels: 2 = 2 to 5 fold, 5 = 5 to 10 fold and 10 = >10 fold. Mass increases and decreases correspond to the first ripening stage mentioned in the pair compared (e.g. white in W/G).

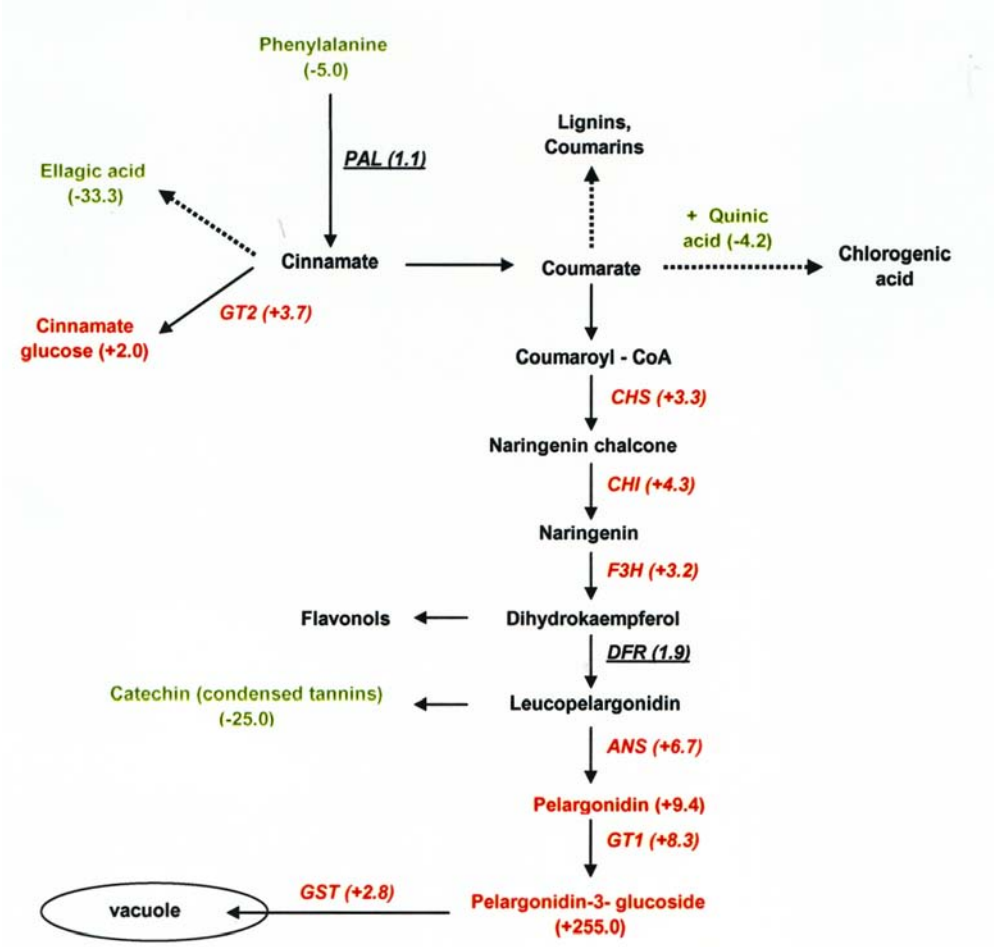


Figure 5 Metabolic and gene expression profiling comparisons in the phenylpropanoid and related pathways in ripening strawberries. Genes analysed for their expression by microarrays are indicated in italic capitals and their red versus green (R/G) expression ratios are given in parentheses (minimal ratio was determined by statistical analysis as 2.6; Aharoni et al., 2000). Putative metabolites or genes showing elevated expression in the red stage compared to green are depicted in red (+ value). Those showing an elevated expression in the green stage compared to red are depicted in green (- value). Those not detected are in black and those detected but with no significant change in accumulation are depicted in black and underlined. PAL- phenylalanine ammonia lyase; CHS- chalcone synthase; CHI- chalcone flavanone isomerase; F3H- flavanone-3-hydroxylase; DFR- dihydroflavonol 4-reductase; ANS- anthocyanidin synthase; GT- glucosyltransferase; GST- glutathione *S*-transferase. Dashed lines indicate multiple steps in the pathway. Except from phenylalanine all other differentially expressed metabolites detected were reported in literature as well (see main text). The enzymatic activity of the protein encoded by *GT2* was determined experimentally (S Lunkenbein, unpublished data) while other genes are putative and where defined by their homology to genes in the public databases.

The Application of FTMS for Metabolic Screening of Mutants

Screening for metabolic alterations in varies types of mutants, as for example transgenic plants overexpressing regulatory genes, may possibly be the most important application of non-targeted metabolic analysis using FTMS. In order to validate the capacity of the FTMS system to screen for metabolic alterations in varies types of mutants, we analyzed transgenic tobacco plants

overexpressing *FaMYB1*, a strawberry MYB transcription factor gene (Aharoni et al., 2001). Flowers of wild-type tobacco are colored red (Fig. 6A), while the *FaMYB1* transgenic lines show a clear reduction in petal pigmentation and appear white. The lines tested were a wild-type control (Wt), a first generation transgenic line showing the phenotype (T1-white), and two lines with red and white flowers (Bc-red and Bc-white), originating from a backcross progeny produced from crossing a wild type plant and the T1-white line (see Fig. 6A). Samples from flower petals of all four plant lines were extracted with 50/50 MeOH/0.1% formic acid and analyzed using FTMS with ESI and APCI in both positive and negative ionization modes.

TABLE 3. DIFFERENTIALLY EXPRESSED MASSES BETWEEN THE RED WILD TYPE AND THE WHITE FLOWER TOBACCO MUTANT OVER-EXPRESSING THE STRAWBERRY *FaMYB1* TRANSCRIPTION FACTOR

No.	IM	OM	T1-white	Bc-red	Bc-white
1	APCI-	164.0717	↑[a]	=	↑
2	APCI-	273.0768	↓	=	↓
3	APCI-	289.0717	↓	=	↓
4	APCI-	363.1658	↑	=	↑
5	APCI-	435.4205	↑	=	↑
6	APCI-	463.4522	↑	=	↑
7	APCI+	486.2601	↑	=	↑
8	ESI-	257.0954	↓	=	↓
9	ESI+	595.1657	↓	=	↓

[a]Increased (↑), decreased (↓) and unchanged (=) compared to the control wild type red flower tobacco plants.
IM, ionization mode; OM, observed mass; CG, chemical group.

Data analysis of the flower petal extracts revealed only nine mass peaks (Table 3), which appear to correlate with the white petal phenotype, i.e. showed either higher or lower expression in both the T1-white and Bc-white lines with no change in the Bc-red line compared to the Wt line. Despite the ultra-high mass resolution, we could not yet unambiguously identify the structure of 8 out of the nine metabolites although their putative empirical formula provided a clue to the possible chemical group they belong. Their unambiguous identification will require multiple stages of mass spectrometry (MS/MS) and/or NMR and known analytical standards. However, in the case of m/z 595.16572 (possible chemical formula, $C_{27}H_{31}O_{15}$+), a natural products search using this formula (regarding possibilities for both $M+H^+$ and M^+) revealed 17 possible compounds which all corresponded to be glycosides of either cyanidin, pelargonidin, peonidin or luteolinidin (all belonging to the group of flavonoids). Among the 5 cyanidin glycosides indicated was keracyanin (cyanidin-3-rhamnoglucoside) reported to accumulate in tobacco flowers (Duke, 1992). HPLC analysis with photodiode array detection of the red wild type and the white flower mutant confirmed that the levels of cyanidin-3-rhamnoglucoside are indeed markedly different between the two lines (Fig. 6B).

A
ai
(10^6)
Positive Ion ESI
Wt-control
595.165
Bc-white
Bc-red
595.165
T1-white
m/z
B
au
Wt-control
T1-white
Cyanidin-3-R
Wt-control
+ Cyanidin-3-R
rt

Figure 6 (previous page) The application of FTMS for metabolic screening of plant mutants. **A,** Extracts of tobacco mutant lines overexpressing the strawberry *FaMYB1* transcription factor showed severe alteration in their flower color compared to the wild type plants and were subjected for analysis using FTMS. Samples from four different plant lines were extracted with methanol/formic acid and ionised with the four ionization modes. Those lines included the wild type control (Wt), a first generation transgenic line showing the phenotype (T1-white) and two lines with red and white flowers (Bc-red and Bc-white), originating from a backcross progeny produced from crossing a wild type plant and the T1-white line. The mass peak shown (observed mass 595.16572 with $C_{27}H_{31}O_{15}$+ as the only possible chemical formula) was identified in the ESI positive ionization mode and was significantly decreased in the white mutant lines. A natural products search using the corresponding $[M]^+$ and $[M+H]^+$ formulae revealed cyanidin-3-rhamnoglucoside (keracyanin) as a possible candidate. ai, area of integration. **B,** HPLC analysis using the red wild type and the white flower mutant confirmed that cyanidin-3-rhamnoglucoside is a major metabolite altered between the two lines. au, absorbance units and rt, retention time.

DISCUSSION

A new and rapid procedure is presented for detection, relative quantification and putative identification of metabolites in biological samples. By analyzing four stages of strawberry fruit development extracted in two ways and ionized by four ionization modes, we could detect 5844 different mass peaks. By removal of a further overlap estimated as in the range of 10% between different ionization modes (D.B. Goodenowe, unpublished data), 5250 different mass peaks were eventually detected in this study. However, the total number of masses detected is estimated to be in the range between 5250 and 5844 because some masses would correspond to metabolites that have identical empirical formulas but different chemistry and can thus be detected in different modes. For example, an ester would be detected in ESI+ and/or APCI+ whereas a fatty acid with the same empirical formula would be detected in ESI- and/or APCI-. In the final summary these two different metabolites would be erroneously considered as one. In addition, within the same ionization mode the system does not discriminate between isomers, also resulting in a number of mass peaks that may be lower than the number of metabolites present.

The most likely empirical formula obtained for each mass was used to search a natural product database for putative metabolite identity. Changes in the levels of known fruit compounds, as well as putative compounds not previously associated with strawberry fruit development, revealed novel information on the metabolic transition from immature to ripe fruit. Combining all the metabolic activities identified during the early stages of strawberry fruit development revealed the process of accumulation of compounds that serve as raw material and building blocks for the production of ripening-associated metabolites. For example, the relative high levels of various organic, phenolic and amino acids in unripe fruit may serve as precursors for the formation of aromatic/phenolic components, while other amino acids such as glutamine may be used for nitrogen metabolism. Oxidative breakdown of accumulated unsaturated fatty acids may result in the biosynthesis of aroma compounds (i.e. aldehydes, alcohols and esters). Carbon skeletons produced from TCA cycle intermediates may be used to synthesise certain amino acids that, in

turn, can be converted into larger molecules. A most important output of these experiments was that the method described could simultaneously and rapidly detect masses putatively corresponding to various types of compounds known to accumulate during ripening of strawberry fruit. Ripening associated metabolites that could be detected included flavour compounds such as furans, ester derivatives (aldehydes, alcohols and esters) and terpenes, carbohydrates, anthocyanins and different other phenolics.

Although knowing the empirical formula of an unknown metabolite is a very powerful and specific clue as to its identity, it cannot provide unambiguous identification due to the fact that isomers have identical empirical formulas. In order to progress from the elemental composition of a metabolite to its geometric structure, separation by HPLC, multiple stages of mass spectrometry (MS/MS) and/or nuclear magnetic resonance (NMR) must be employed. In most of these situations known analytical standards of the isomers are required to identify unambiguously the structure of the metabolite. For example in the case of $C_6H_{12}O_6$, the results from a database search for a possible metabolite identity will include not only fructose and glucose as but also mannose, galactose and other less abundant sugars. Therefor the results are, at least in a part of the cases, a sum parameter for certain isomer classes, some of them may include a large number of individual members. Hence, the identification of the class of compounds to which an isomer may belong will be of a major benefit since this will allow to narrow down the search for metabolite identity either by searching the literature or by using a more specific analytical method for its identification. Indeed, the metabolite database search for $C_{27}H_{31}O_{15}+$, one of the mass peaks that was changed in the white tobacco mutant (Fig. 6), indicated that this metabolite belonged to the class of flavonoids. The result of an FTMS-mediated accurate mass-based empirical formula output must be therefore regarded as the equivalent at the metabolite level for a BLAST search output frequently used for the assignment of a putative identity to a gene or a protein sequence. Both in the accurate mass and in the BLAST outputs further experimental evidence must be provided to unambiguously determine the metabolite or gene/protein identity, respectively. In spite of these limitations, obtaining the empirical formula of a metabolite through the use of accurate mass determinations, as presented in this work, provides the most specific information regarding the identity of an unknown metabolite for which there is (by definition) no analytical standard available. Furthermore, empirical formula determinations as described here offer the only means of analyzing metabolites in an unbiased and non-targeted fashion.

The mass data output obtained by FTMS in the current study analysis was correlated with the data obtained from gene expression studies using DNA microarrays. Similar in both systems is that the peak/spot intensity can be analyzed and directly expressed as a ratio between two metabolite/mRNA populations. In microarray data analysis this methodology of using ratio's rather then absolute levels was shown to be powerful and important in overcoming a large portion of experimental variation (Aharoni and Vorst, 2002). Thus, in the cases where mass peaks detected by FTMS could be subsequently unambiguously identified (employing various methods described

above) and then associated with a metabolic pathway, integration of this two “profiling” sets will be feasible.

Two other issues, namely in-source ion suppression and adduct formation, are of concern for FTMS, as for all mass spectrometry based methods. Biological samples contain a mixture of components and one component may cause the signal intensity of another component to be suppressed and this has implications when trying to quantitate multi component samples (Sterner et al., 2000). Thus, samples used for comparison should preferably contain as much as possible similar matrices. A second limitation of the method is the formation of adducts during the ionization process. For instance, in positive ionization modes many metabolites may not be detected as only the common protonated adduct [M+H]+, but also as [M+Na]+, [M+K]+, [M+M+H]+, [M+solvent+H]+, [M+NH4]+, [M+HCOOH+H]+and others. Hence, analytical methods must be designed to minimize or eliminate such formations. If adduct formation is observed, the search for elemental composition of observed masses should include a search for adduct formation and special software must be developed in order to deconvolute the data. Non-hydrogen adduct formation was not observed in the current study. Adduct formation might also be an advantage for some applications (Gorlach and Richmond, 1999). For example, in global sample “fingerprinting”, the high vulnerability of ion abundance and adduct formation due to matrix effects could be exploited. On the other hand, some compounds do not form adducts at all but may be present in the injection solution as [M+] or [M-] ions and can thus be detected as such upon ionisation. For instance, in the methanol/formic acid extracts the anthocyanins pelargonidin-3-glucoside in red strawberry (Table 2) and cyanidin-3-rhamnoglucoside in red tobacco flowers (Fig. 6) were detected as their [M+] ions without detectable adduct formation.

We believe that the most useful application of the direct-injection FTMS method described will most probably be in the search for similarities and dissimilarities in large collections of biological samples, (e.g. plant mutant collections). Using the method for overall high-throughput and multiple comparisons of samples of different origins a direct infusion FTMS approach would be clearly advantageous over existing time-consuming metabolite analyses or screening methods. With a few minutes for data acquisition, the method is clearly faster than any chromatographic method, and has a very high information content. We demonstrated the utility of the method for such mass profiling by analyzing transgenic tobacco plants overexpressing a regulatory gene. Analysis of flower extracts in the four ionization modes resulted in the detection of nine masses that were significantly different between the transgenic and control plants. Among the deduced metabolites was cyanidin-3-rhamnoglucoside, which was markedly lowered in the transgenic line. The high-throughput, comprehensive, and complementary nature of this generic phenotyping technology should therefore greatly aid in the systematic analysis of gene function and plant physiological processes.

MATERIALS AND METHODS

Preparation of Extracts for FTMS Analysis

Three hundred mg FW fine-powdered frozen tissues of strawberry fruit (*Fragaria* x *ananassa*, cv. *Elsanta*) from green, white, turning and red stages of development, without achenes, and of tobacco flowers (*Nicotiana tabacum*, cv. SR1) were mixed with 3 ml of either methanol/0.1% formic acid (50%/50%) or 100% acetonitrile solutions. The extracts were sonicated (15 min, which at least appeared sufficient to extract phenolic compounds), filtered (0.2 μm PTFE filter) and stored at –80°C. Samples were diluted 1:19 prior to ESI and APCI analysis. 50/50 MeOH/0.1% ammonium hydroxide and 50/50 MeOH/0.1% formic acid were used as mobile phases for dilution of all negative and positive ion ionization analyses, respectively.

FTMS Instrumentation

All analyses were performed on a Bruker Daltonics APEX III Fourier Transform Mass Spectrometer (FTMS) equipped with a 7.0 Tesla actively shielded super conducting magnet and ESI and APCI sources. ESI, APCI, and ion transfer conditions were optimized using a standard mix of serine, tetra-alanine, reserpine, Hewlett-Packard tuning mix, and the adrenocorticotrophic hormone fragment 4-10. Instrument conditions were optimized for ion intensity and broadband accumulation over the mass range of 100-1000 amu. One megaword data files were acquired and a sinm data transformation was performed prior to Fourier transform and magnitude calculations.

FTMS Calibration

All samples were internally calibrated for mass accuracy over the approximate mass range of 100-1000 amu using a mixture of the above-mentioned standards. All mass deviances from the standard curves were <1.0 ppm over the mass range studied, although most were typically in the 0.1 to 0.2 ppm range.

FTMS Data Processing

The mass spectra from each analysis were integrated following calibration creating a peak list that contained the exact mass and absolute intensity of each peak. This raw peak list was then filtered to remove all ^{13}C isotopes. Mass analysis was then performed on each of the peaks in this filtered peak list to identify any peak that contained a chemically meaningful combination of Carbon, Hydrogen, Nitrogen, Oxygen, Phosphorus and Sulfur, with a single charge and with an error of

less than 1.0 ppm. Peaks that met these criteria along with the proposed empirical formulae and error of determination were stored in the final processed data file. These empirical formulas were screened for a possible metabolite by searching the Dictionary of Natural Products (159,000 compounds; Chapman and Hall/CRC, London, UK). In order to compare and summarize data across different ionization modes, detected mass peaks were converted to their corresponding neutral masses and used to sort the data accordingly.

HPLC Analysis

Filtered extracts were subjected to analysis by reversed phase HPLC with photodiode array detection (Aharoni et al. 2001). A gradient of acetonitril in 0.1% trifluoro acetic acid and a Symmetry C18-column (3.9x150 mm; Waters Chromatography, Etten-Leur, The Netherlands) at 40°C were used to separate the extracted compounds. Retention times and absorbance spectra, recorded at 240-600 nm, of eluting anthocyanins were compared with authentic reference compounds (Apin Chemicals, Abingdon Oxon, United Kingdom and Extrasynthese S.A., Genay Cedex, France).

ACKNOWLEDGEMENTS

We thank Robert Hall and Andy Pereira for critically reading the manuscript.

REFERENCES

AHARONI, **A., KEIZER, L.C.P., BOUWMEESTER, H.J., SUN, Z., ALVAREZ-HUERTA, M., VERHOEVEN, H.A., et al.** (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell **12,** 647-661.

AHARONI, A., DE VOS, C.H.R., WEIN, M., SUN, Z, KROON, A., MOL, J.N.M., et al. (2001) The Strawberry *FaMYB1* Transcription Factor Suppresses Anthocyanin and Flavonol Accumulation in Transgenic Tobacco. Plant J **28**, 319-332.

AHARONI, A., and VORST, O. (2002) DNA microarrays for functional plant genomics. Plant Mol Biol **48**, 99-118.

BRENNER, S., JOHNSON, M., BRIDGHAM, J., GOLDA, G., LLOYD, D.H., JOHNSON, D., et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol **18**, 630-634.

BROWN, P.O., and BOTSTEIN, D. (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21 Suppl. **S**, 33-37.

CHEE M., YANG R., HUBBELL E., BERNO A., HUANG X.C., STERN D., et al. (1996) Accessing genetic information with high-density DNA arrays. Science **274**, 610-614

CHENG, G.W., and BREEN, P.J. (1991) Activity of phenylalanine ammonia-lyase (PAL) and concentrations of anthocyanins and phenolics in developing strawberry fruit. J Am Soc Hort Sci **116**, 865-869.

DUKE, J.A. (1992). Handbook of phytochemical constituents of GRAS herbs and other economic plants. Boca Raton, FL. CRC Press.

DUTT, M.J., and LEE, K.H. (2000) Proteomic analysis. Curr Opin Biotech **11**, 176-179.

FIEHN, O., KLOSKA, S., and ALTMANN, T. (2001) Integrated studies on plant biology using multiparallel techniques. Curr Opin Biotech **12**, 82-86.

FIEHN, O., KOPKA, J., DOERMANN, P., ALTMANN, T., TRETHEWEY R.N., and WILLMITZER, L. (2000) Metabolite profiling for plant functional genomics. Nat Biotech **18**, 1157-1161.

GLASSBROOK, N., and RYALS, J. (2001) A systematic approach to biochemical profiling. Curr Opin Plant Biol **4**, 186-190.

GORLACH, E., and RICHMOND, R. (1999) Discovery of quasi-molecular ions in electrospray spectra by automated searching for simultaneous adduct mass differences. Anal Chem **71**, 5557-5562.

GRIERSON, D. (1998) Manipulation of fruit ripening by genetic modification. *In* K Lindsey, ed, Transgenic Plant Research. Harwood Academic Publishers, Amsterdam, The Netherlands, pp 109-124.

LATZA, S., GANSSER, D., and BERGER, R.G. (1996) Identification and accumulation of 1-O-trans-cinnamoyl-beta-D-glucopyranose in developing strawberry fruit (Fragaria ananassa Duch. Cv. Kent). J Agri Food Chem **44**, 1367-1370.

MAAS,. J.L., WANG, S.Y., and GALLETTA, G.J. (1991) Evaluation of Strawberry Cultivars for Ellagic Acid Content. HortScience **26**, 66-68.

MARTZEN, M.R., MCCRAITH, S.M., SPINELLI, S.L., TORRES, F.M., FIELDS, S., GRAYHACK, E.J., et al. (1999) A biochemical genomics approach for identifying genes by the activity of their products. Science **286**, 1153-1155.

MOING, A., RENAUD, C., GAUDILLERE, M., RAYMOND, P., ROUDEILLAC, P., and DENOYES ROTHAN, B. (2001) Biochemical changes during fruit development of four strawberry cultivars. J Am Soc Hort Sci **126**, 394-403.

PARINOV, S., and SUNDARESAN, V. (2000) Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotech **11**, 157-161.

PERKINS VEAZIE, P. (1995) Growth and ripening of strawberry fruit. Hort Rev **17**, 267-297.

REN, B., ROBERT, F., WYRICK, J.J., APARICIO, O., JENNINGS, E.G., SIMON, I., et al. (2001) Genome-wide location and function of DNA binding proteins. Science **290**, 2306-2309.

ROESSNER, U., WAGNER,. C., KOPKA, J., TRETHEWEY, R.N., and WILLMITZER, L. (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J **23**, 131-142.

ROESSNER, U., LUEDEMANN, A., BRUST, D., FIEHN, O., LINKE, T., WILLMITZER, L., et al. (2001a) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell **13**, 11-29.

ROESSNER, U., WILLMITZER, L., and FERNIE, A.R. (2001b) High-Resolution Metabolic Phenotyping of Genetically and Environmentally Diverse Potato Tuber Systems. Identification of Phenocopies. Plant Physiol **127**, 749-764.

SEYMOUR, G.B., TAYLOR, J.E., and TUCKER, G.A. (1993) Biochemistry of fruit ripening. Chapman & Hall, London, pp 454.

STERNER, J.L., JOHNSTON, M.V., NICOL, G.R., and RIDGE, D.P. (2000) Signal suppression in electrospray ionization fourier transform mass spectrometry of multi-component samples. J Mass Spec **35**, 385-391.

TRETHEWEY, R.N., KROTZKY, A.J., and WILLMITZER, L. (1999) Metabolic profiling: a Rosetta stone for genomics? Curr Opin Plant Biol **2**, 83-85.

UETZ, P., GIOT, L., CAGNEY, G., MANSFIELD T.A., JUDSON, R.S., KNIGHT J.R., et al. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403**, 623-627.

WEIGEL, D., AHN, J., BLAZQUEZ, M.A., BOREVITZ, J.O., CHRISTENSEN, S.K., FANKHAUSER, C., et al. (2000) Activation tagging in Arabidopsis. Plant Physiol **122**, 1003-1013.

WINZELER, E.A., SHOEMAKER, D.D., ASTROMOFF, A., LIANG, H., ANDERSON, K., ANDRE, B., et al. (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science **285**, 901-906

SUMMARY, CONCLUDING REMARKS AND FUTURE PROSPECTS

Summary

Fruit ripening is characterised by active metabolism and dramatic changes in texture and pigmentation which make the fruit a "biological bribe", attracting animals. A fully ripe and appealing fruit will actually ensure seeds dispersal and the production of a next generation. Early studies concluded that fruit ripening was a process of uncontrolled self-distraction and tissue senescence. Nowadays it is evident that this is not the case. On the contrary, ripening is a tightly regulated procedure involving the *De Novo* synthesis of mRNA, proteins, and metabolites, which are crucial for both regulating and implementing the process. A focal point of this study was to obtain a detailed "picture" of the genetic mechanisms governing strawberry fruit ripening. Of main interest was the biosynthesis of flavour and aroma compounds, a process hardly investigated at the molecular level in fruits. We utilised an array of molecular and biochemical tools to monitor the dramatic changes occurring in strawberry fruit during its maturation. Apart from observing the ripening process as a whole, we also performed more focused studies, isolating and characterising new genes taking part in the metabolism of important ripening metabolites.

The incorporation of DNA microarray technology to our research strategy provided us with a powerful way to monitor alterations in gene expression in Strawberry during fruit development and ripening. *Chapter 2* of the thesis reviews the principals, technical aspects and applications of microarrays. The main advantage of using microarray technology for the study of gene expression is that it allows the large-scale, quantitative and simultaneous analysis of mRNA levels in at least two samples under investigation. In a single microarray experiment the expression of hundreds and thousands of genes can be monitored. This study has highlighted a valuable aspect of this technology, that is the possibility to efficiently use it for the global investigation of more complex non-model plant species such as strawberry.

Initial experiments were performed using a microarray composed of 1700 strawberry cDNAs. These cDNAs were isolated in the course of a random sequencing of a strawberry ripe fruit cDNA library. Gene expression was monitored during different stages of fruit development and ripening (green, white, turning, red) in both achene and receptacle tissues. Overall 537 genes were detected which had at least once a significant change in expression during development or in between the

two tissue types. More than a hundred genes were identified as specifically ripening related from this study. They were to serve as a rich source for the selection of individual genes for downstream investigation. For the first time a global picture of key processes occurring simultaneously during ripening in both achene and receptacle tissues was obtained. Whilst gene expression in the achenes was mainly devoted to the acquisition of desiccation tolerance, expression in receptacle tissue was largely associated with the metabolism of ripening related compounds (pigments, cell wall components, fatty acids, volatile flavour etc.).

Two processes, namely the development of the vascular tissue and stress response emerged from our microarray study as playing an important role in strawberry fruit maturation. The importance of the vascular tissue to fruit development and ripening in strawberry and in other fruit species has to date been largely overlooked. The vascular tissue in the fruit connects the interior of the receptacle to both the achenes (located in the epidermal layer) and the pedicel. Apart from functioning as a transport system it may also play a significant role in the fruit texture. From our microarray data analysis we detected a remarkable parallelism in gene expression patterns between tracheary element (TE) differentiation in *Zinnia elegans* and strawberry ripening. The first three microarray experiments, showed that 31 individual genes out of the 112 ripening related genes identified (28%), were similar to genes identified as important for TE differentiation in *Zinnia elegans*. The results suggest that vascular development including lignification is an integral process in the wide strawberry development and ripening transcriptional program, which may in part be regulated by auxin.

A substantial number of the ripening related genes identified showed homology to stress related genes present in public sequence data bases. To investigate whether oxidative factors might be a possible cause of stress conditions in the fruit receptacle during the ripening process, an additional microarray experiment was performed examining which of the ripening related genes identified were induced in expression upon treatment of the fruit with a free radical generator. Hybridisation was conducted using a newly synthesised microarray, representing a limited set of 384 elements (termed “second generation microarray”). This new microarray was mainly composed of cDNA clones identified earlier as ripening regulated and receptacle associated. From this study we observed that 22 out of 45 cDNA clones previously identified as ripening regulated were induced by the stress treatment. The results demonstrate a link between oxidative stress and ripening.

Another important aspect of fruit ripening is the hormonal regulation of this complex biological process. Ethylene is a known mediator of ripening in various fruit such as tomato, banana and avocado. Banana and avocado normally defined as climacteric fruit, show a typical rise in ethylene levels prior to ripening, accompanied by a rise in respiration. Strawberry does not behave in this manner and is therefore defined as non-climacteric. The hormone auxin is known to play a major role in the regulation of strawberry fruit ripening. High auxin levels promote early growth of the receptacle. As auxin levels drop during the green to white stage, ripening is

triggered, mainly by the induction of gene expression. We used the second generation microarray to investigate how broad is the effect of auxin on ripening related gene expression in strawberry fruit. Results for our study showed that genes related to ripening processes such as pigmentation and volatile flavour production are clearly auxin dependent. However, a large set of ripening related genes are auxin independent such as genes related to stress responses and genes associated with cell wall metabolism. A detailed description of genes, gene expression, metabolic pathways and genetic programmes reflected by the microarray experiments described above are provided in *Chapter 3* and *Chapter 4*.

Accumulation and emission of volatile fruit flavours is one of the most spectacular characteristics of the ripe fruit. In fruit, including strawberry, flavour is a mix of hundreds of compounds belonging to several different chemical classes. The various groups include for example aldehydes, alcohols, esters, ketones, aromatic compounds, terpenes, furans, lactones, and sulphur compounds. These components are not unique volatiles of fruit. Some of them will dominate the scent of flowers and others are produced and emitted either constitutively or under certain environmental conditions by vegetative parts of plants. The contribution of each flavour compound to the overall fruit flavour is unequal, and some compounds although present in minor levels could have a dramatic influence on the flavour of a particular fruit. Our knowledge on the biosynthesis of such compounds is mostly derived from biochemical studies, investigating their levels in different plant species, predicting metabolic pathways and measuring enzyme activities. For a large set of volatile flavours of fruit the pathway to their biosynthesis is largely unknown. Our molecular genetic knowledge on how these compounds are formed in fruit is even more limited. In *Chapter 5* and *Chapter 6* the identification and characterisation of genes directly involved in the biosynthesis of volatile flavour compounds in strawberry is described. Volatile esters formed by are major components in the volatile profile of strawberry and most other fruit species. In strawberry alone more than 100 types of esters have been detected. The last committed step in ester biosynthesis, is a condensation reaction of an alcohol and an acyl moiety, carried out by an acyltransferase type enzyme.

Mining microarray gene expression data during strawberry development and searching sequence homologies in the public databases revealed a strong candidate (termed *SAAT*) for encoding the acyltransferase catalysing ester formation in the ripe strawberry fruit. *SAAT* showed high levels of transcripts during ripening, and its expression was fruit specific. Enzyme activity assays using the recombinant SAAT protein produced in *E.coli* cells, showed that the enzyme was capable of forming esters *in-vitro*. Interestingly, the SAAT recombinant enzyme could use an array of alcohols and acyl, forming a considerable series of esters. This broad enzyme activity might explain the biosynthesis of more than a 100 types of esters in the ripe strawberry fruit, by only a few ester-forming enzymes. During the course of this study *SAAT* homologs have been cloned and their corresponding recombinant enzymes have been produced from seven other fruit species (wild strawberry, lemon, mango, melon, tomato, apple, and banana). Acyl-transfer activity of some of

them, such as for example the wild strawberry enzyme, banana and lemon, was demonstrated. In addition, a further study examining petunia and tobacco plants constitutively expressing *SAAT*, showed that the availability of alcohol was limiting the production of esters in leaves and flowers of this plant species.

Another group of volatile components, also present in strawberry and other fruit species, are the 10 carbon monoterpenes and 15 carbon sesquiterpenes. Monoterpenes especially, are primarily known for their elevated levels in herbs (e.g. mint), which provide this plants with it's characteristic smell. Examining the volatile profile of wild and cultivated strawberry in our study revealed that their terpenoid composition was very different. Cultivated varieties produce mainly the terpene alcohols linalool and nerolidol whilst the wild species produce mainly olefinic monoterpenes (e.g. α- pinene, β- phellandrane, β- myrcene, and trans- ocimene). In *Chapter 6* the cloning and characterisation of genes encoding the enzymes catalysing the biosynthesis of terpenoids in both species is described. As a result of our study novel evidence on how volatile flavour compounds are gained and lost during domestication of crops was attained. Changes in the production of the compounds was a result of alteration to the coding regions of genes through both a gain and loss mechanisms. In the "gain" scenario, the 5' of the gene encoding a plastid targeting signal was truncated by a deletion and a stop-codon, causing a change in localization of the protein to the cytosolic compartment. In the cytosol, the enzyme (capable of producing the monoterpene linalool and the sesquiterpene nerolidol) could encounter its substrate and synthesise the products. However, the change in targeting was not sufficient and an additional change in expression, probably by alteration to the promoter region, were needed for high and fruit specific expression, and production of the desired terpene. Loss of the olefinic monoterpene production was suggested to occur as a result of the introduction of a truncated allele to the genome of the cultivated strawberry species. The presence of an aberrant mRNA in the cells, possibly caused the activation of RNA surveillance mechanisms, which resulted in the degradation of all similar mRNAs in the cells (even complete ones), and due to this no product was formed.

The concerted action of structural genes encoding enzymes involved in a certain metabolic pathway is most likely co-ordinated by the action of regulatory proteins, amongst them transcription factors. Transcription factors can activate or repress gene expression (in some cases both), either directly by binding to promoter regions upstream of target genes or by interacting with other factors which will affect downstream genes. In the course of our initial screen to identify ripening regulated genes in strawberry we detected a Myb transcription factor homolog, showing a ripening regulated expression pattern. More than a hundred and fifty different Myb factors have been identified in the genome of the model plant *Arabidopsis thaliana*, and it is conceivable that a similar number are present in the strawberry genome as well. To date Myb proteins have been largely associated with the regulation of metabolic pathways (mainly secondary metabolism, e.g. phenylpropanoid metabolism), and developmental processes (e.g. trichome development and lateral meristem initiation pathway). In order to gain further insight into the

function of the strawberry Myb (FaMYB1), we over expressed its corresponding cDNA in transgenic tobacco. Expression of *FaMYB1* resulted in a dramatic effect on flower petals, causing a white appearance compared to the red coloured petals of wild type plants. Pigmentation in tobacco as well as in strawberry and a vast number of other red coloured plant organs such as seeds, flowers and fruit is a result of the accumulation of anthocyanins, produced in the flavonoid biosynthetic pathway. Biochemical analysis of transgenic plants confirmed the decrease in anthocyanin pigment in transgenic flowers. The same analysis revealed a reduction in levels of another metabolite produced in the flavonoid biosynthetic pathway, namely the flavonol quercitin. Flavonols and other types of flavonoids have long been considered to play a role in UV protection, and often accumulate in epidermal cell layers. Examining the transgenic plants for alteration to gene expression and enzymatic activities in the flavonoid pathway and it branches, identified a reduction in gene expression and enzyme activities, in the lower part of the pathway. This part of the pathway is committed to flavonol and anthocyanin biosynthesis. Myb factors have previously been associated with the regulation of the flavonoid pathways, often acting as transcriptional activators. The results, presented in detail in *Chapter 7* of this thesis suggest that FaMYB1 acts as a repressor of flavonoid biosynthetic genes in late strawberry fruit ripening.

The last chapter of the thesis (*Chapter 8*), describes a new procedure for non-targeted and large scale metabolic screening. In recent years, as part of the "genomics revolution" and the major developments in functional genomics tools, researchers in plants as well as in other disciplines have attempted to develop methods for the large scale analysis of metabolites. In order to be compatible with data obtained from other levels of regulation, (e.g. microarray data on gene expression), the methods developed should be non-targeted to a specific metabolite or group of metabolites. This as opposed to nearly all metabolic analysis performed to date which have involved specific extraction, separation and detection methods for certain type of compounds. The complexity of metabolism, especially in plants, makes the development of such technologies nearly impossible and a great challenge. The method described in *Chapter 8* uses Fourier Transform Ion Cyclotron Mass Spectrometry (FTMS) to detect hundreds and thousands of mass peaks in a given plant extract. High mass accuracies determined by FTMS analysis for each peak in the spectrum allow the assignment of a possible empirical formulae for each mass and predict the identity of the metabolite detected (especially when masses below 300 are examined). Identification of the different metabolites is not unambiguous due to the presence of multiple compounds with identical empirical formula (e.g. isomers) and the potential for more then a single empirical formula matching a given molecular mass (especially in the case of high molecular weight). The utility of the method was tested by analysing metabolite changes between the different stages of strawberry fruit development and comparing the FaMYB1 expressing tobacco lines described above to wild type plants. The results provided the first preliminary data on the capability of the FTMS approach for non-targeted metabolic screening.

Concluding Remarks and Future Prospects

The scientific field of genomics although having emerged only in the last 8 years has already waged back its investments in terms of determining the function of hundreds of genes in a large number of organisms. However, sequence information by itself is not sufficient for determining gene function and understanding the biology behind it. The majority of experiments are based on observing variation (i.e. the differences between the normal and the unusual) which might either be induced or naturally occurring. Collection of genetic material with large variation and tools for mutant production are most often the basic requirement for a successful experiment. Apart from collecting or inducing variants we also need methods to detect differences and quantify them. Such approaches, integrated under the term functional genomics, enable us to produce, detect, and quantify variation on a larger scale than before and allow us to obtain a broader perspective of biological processes.

Functional genomics combines the analysis of variation at the several levels of genetic regulation, and uses bioinformatics to store, integrate and mine the results collected. This systematic approach has initially been applied to organisms such as the yeast, with only slightly more than 6000 genes in its genome and a very defined and extensively studied genetics. In plants, functional genomics approaches have mainly been applied to *Arabidopsis thaliana* and it is only in the last few years that similar methodologies have been used in other model organisms such as rice. Some of the functional genomic tools such as the production of a large amount of sequence data, gene expression and preliminary metabolite analyses have been proven in this thesis to be most powerful for the study of non-model and more complicated plant species such as strawberry. In strawberry our effort to obtain extensive biological insight and to fully exploit the power of for example microarray technology has been largely hampered by the lack of variable genetic material, namely mutant lines. Major insights into strawberry fruit biology were obtained in experiments utilising fruit with altered hormone levels (i.e. *Chapter 3*) and wild type strawberry fruit in the case of studying terpenoid related genes (i.e. *Chapter 6*). This apparent lack of genetic material could be overcome in the future by generating mutant populations as for example those produced by chemical treatment or radiation. Alternatively, transposon or T-DNA tagged populations could be generated and utilised for reverse and forward genetic screens.

The field of molecular biology has made a major contribution to our understanding of plant biology, especially for addressing function of genes. However, it is apparent that a successful research can not rely solely on molecular biology but should be combined with other fields, such as physiology and biochemistry. It is evident from this thesis work that the combination of molecular biology with biochemistry is very powerful and is an absolute request for the study of metabolic pathways of secondary metabolism, such as flavour biosynthesis. To provide a strong base for the molecular work, the biochemical investigation needs to meet three fundamental requirements. One is the determination of the various enzymatic and non-enzymatic steps in the

metabolic pathway of interest by means of for example feeding experiments and enzymatic assays. Such analysis should cover the entire pathway from primary substrates to final products, and also include possible branching of the pathways and modification of compounds, i.e. glycosylation, hydroxylation and methylation. A second requirement would be the proper experimental protocol for the extraction, separation, detection and identification of the compounds in the pathway. A third requirement, crucial for matching the biochemical and molecular data, is the determination of the level of the different components of the pathway, especially final products in different developmental stages, tissues and physiological conditions. It is only with such a basic and detailed biochemical data that research groups will be able to utilise efficiently advanced molecular tools to understand the molecular genetics, including regulation, of metabolic pathways.

It is often hard to envisage how one can study certain characteristic biological process active in non-model species by examining model species. However, such an approach of using model species for the study of more exotic and complex plants is becoming more common as our knowledge on the model species increases. The identification of orthologues genes, which share a common ancestor but have evolved a new function during evolution of species, makes such an approach feasible. For example, the study of fleshy fruit development and ripening might very well be performed in the model species *Arabidopsis thaliana*, which although produces a dry fruit, shares common genetic mechanisms for fruit set and maturation. Even in very special cases, such as the study of furaneol biosynthesis (the characteristic flavour component of strawberry), when it might be less appropriate to use model systems, it is most likely that orthologues of genes identified earlier in model species, could be of help. In this thesis work, the use of other plant species as models has made a major contribution for the characterisation of genes identified in strawberry (*Chapter 6* and *Chapter 7*), and should have been utilised even more extensively.

A related issue is whether the genetic material for the study of crops, such as fruit species should be the wild species or the present cultivated cultivars which are of commercial importance. It is often believed that in similarity to the use of model systems, studying the wild species will not correspond to the processes occurring in the cultivated varieties and will dissociate the study from any practical application. In the case of strawberry this might not be the case since wild strawberries have a clear advantage of being diploids (compared to the octaploid cultivated varieties) and a shorter lifecycle. Using the wild strawberry as a primary genetic material, rather then the domesticated species, might therefore be a shorter line of research towards an application.

In conclusion, *the time is ripe* to intensify our knowledge on development and ripening of fruit such as strawberry. A vast array of tools and approaches, including data from model species are available. Research groups, which will be open and knowledgeable to combine and efficiently utilise such technologies, approaches and data, will no doubt generate the next scientific breakthroughs.

SAMENVATTING (SUMMARY IN DUTCH)

Moleculaire technieken uit de functionele genomica, zoals het random sequencen van ESTs (Expressed Sequence Tags) en zelfs hele genomen, het bestuderen van de expressie van grote hoeveelheden genen met behulp van macro- en micro-arrays en het genereren van mutant populaties heeft de laatste jaren sterk bijgedragen aan het onderzoek aan modelplanten zoals met name *Arabidopsis thaliana* (de zandraket). Een groot gedeelte van deze moleculaire technieken kan echter ook worden gebruikt voor het onderzoek aan niet-model planten. Weliswaar is het werken aan dit soort planten veel moeilijker, maar het maakt het wel mogelijk unieke processen te bestuderen, die in modelplanten niet voorkomen, zoals bijvoorbeeld de vorming van smaakstoffen in vruchten.

Om inzicht te krijgen in de processen die optreden bij de vruchtrijping van aardbei en om belangrijke genen te isoleren die een rol spelen bij de vorming van de smaakstoffen in aardbei hebben wij in eerste instantie 1000 ESTs gesequenced uit een cDNA bank, gemaakt van rijpe aardbei (cv Elsanta). De informatie over de mogelijke identiteit van deze ESTs (verkregen door vergelijking met sequenties aanwezig in Internet-databases) werd gecombineerd met gegevens over expressie om zodoende interessante genen te identificeren. In eerste instantie werden die expressie studies uitgevoerd met behulp van Northern blots van 50 geselecteerde genen (selectie gebaseerd op interessante sequentie-homologie). Van deze 50 vertoonden er 15 een rijping-gerelateerd expressiepatroon, waaronder *FaMYB1*, een lid van de R2R3 MYB familie van transcriptie factoren. Verder onderzoek toonde aan dat *FaMYB1* een repressor is van de flavonoiden biosynthese laat in de vruchtrijping.

Om de genexpressie van grotere hoeveelheden genen te kunnen bestuderen, werd vervolgens een cDNA micro-array gemaakt met daarop 1700 cDNAs. Met deze micro-array werd de expressie van die 1700 genen tijdens de vruchtrijping bestudeerd en bepaald of de genen in het vruchtlichaam dan wel in de zaadjes van aardbei tot expressie kwamen. Een belangrijke ontdekking van deze studie was de identificatie van het *SAAT* gen dat codeert voor een alcohol-acyltransferase, dat verantwoordelijk is voor de vorming van esters in aardbei. Esters dragen in belangrijke mate bij aan de geur en smaak van aardbei, maar ook van veel andere vruchten. Op een tweede cDNA micro-array werden vervolgens 384 cDNAs gezet, die waren geselecteerd op basis van de resultaten met de eerste micro-array en specifiek waren voor de vruchtrijping. Hiermee werden de effecten bestudeerd van auxine en oxydatieve stress op genexpressie.

Al deze micro-array experimenten tezamen hebben geleid tot een uitgebreid en nieuw inzicht in de genexpressie programma's tijdens de vruchtrijping en tot de isolatie van nog diverse andere genen die betrokken zijn bij de vorming van smaakstoffen in aardbei en die momenteel worden gekarakteriseerd. In aanvulling op de grootschalige studie van genexpressie met behulp van de

micro-arrays, is ook gekeken naar de veranderingen in honderden metabolieten in aardbei waarbij gebruik is gemaakt van Fourier-Transformatie Ion Cyclotron Massa Spectrometrie (FTMS). Deze analysetechniek liet zien dat er tijdens de vruchtrijping sterke veranderingen optraden in veel van de stoffen aanwezig in aardbeienvruchten.

In conclusie: Deze studie heeft laten zien dat de integratie van een aantal van de nieuwe functionele genomica technieken van onschatbare waarde zijn voor zowel het vinden van nieuwe belangrijke genen als voor het doorgronden van de biologie ook van niet-model planten zoals aardbei.

NAWOORD (ACKNOWLEDGEMENTS)

Finally, after a long period, much longer then envisaged, I came to the end of my PhD here in Holland, and I would like to thank all the people that were involved during these years with both the scientific and personal aspects of my doctorate.

The decision to perform a Ph.D. in Holland was largely influenced by the recommendation I received on the CPRO institute and the head of the Cell Biology department at that time, Arjen. My main expectations were to increase my expertise and knowledge in the field of plant molecular biology and plant biochemistry, develop an extensive portfolio of high impact scientific publications and develop collaborations and links to research groups in The Netherlands, Europe and other parts of the world. Now, when I evaluate this period and the achievements (most of them described in this book), I am very satisfied and happy with the "brave" decision to leave temporarily but for a long period my family, friends and country and conduct my Ph.D. in CPRO-DLO / Plant Research International (PRI).

The data described in this thesis were the result of extensive and joint efforts of myself and a large group of other researchers involved in different aspects of the work. Colleagues from former Cell Biology and after Cell Cybernetics, CPRO/PRI, AB-DLO, students, guest researchers, collaborators from The Netherlands and abroad, all contributed in different ways to the output of this research project.

The major person behind my "Dutch" Ph.D. was Arjen, who above all the financial difficulties managed to arrange the grant and support the project. Arjen ! thank you for not giving up with bringing me here and for your scientific and personal example. From the beginning you defined this project as "pioneering, therefore we do not know how it will end", and like me I am sure you are satisfied with the final product. Together with Arjen, Harrie was a main "player" behind my arrival, setting up and executing the project plan. Harrie ! thank you for your support and back-up for a large and diverse set of experiments, from the first day of my arrival. We had some ups and downs during the time, especially when ASAPH impatiently turned into ASAP, but overall the project was a success and became a model for other new projects in the department. Jos, my promoter from Amsterdam, was among the few people who followed me along all the years from the beginning till the last months. Jos, I appreciate and thank you very much for your critical comments and advises on the work and on the publications. My warmest wishes for a full recovery and health.

Members of the former Cell Biology department were of great help along the years. Ric, Jan (Blaas), Arnaud, Reinoud, Ingrid, Robert (Hall), Robert (Sevenier), Hanny, Jeroen, Katinka, Tjitske, Twan, Harry (Jonker), Oscar, Michel, Andries, Elma, Jan (Schaart), Frans, Willem, and last but not least my neighbour from the north, Mazen ! I want to say BEDANKT to all of you for

explaining, teaching and helping. Sorry for not speaking fluently Dutch after all these years, except for a few “important” words such as “tandenstokertje” and “cocktail- prikkeltje”, which Jan taught me shortly after my arrival. Ric ! thank you for the assistance and help with the biochemistry, HPLC, etc.., I learned and enjoyed a lot. Most of the work we did together were actually side and unplanned projects, which eventually turned out to be important building blocks of the work on strawberry ripening and resulted in very good publications. To my office mates Joost (“hang loose” in Canada) and Adèle ! I enjoyed the time with you and I hope you did not suffer too much from my piles of papers and dirty kantine trays.

Among the “new” colleagues in Cell Cybernetics, I would like to thank Dirk as my cluster leader and for taking care for several months of my Ph.D., arranging all the things related to the Ph.D. and making sure that I was on the right track. Thank you, Dirk. Great thanks also to Harro and Francel, who worked together with me on several flavour related projects, mainly on the terpene part in the last 2 years. I appreciate very much your motivation, wish to work together and your important contribution to several so called “corridor projects”, with not really a project number. Those projects very often turn most successful, as in the case of the strawberry terpenoids. And to Marteen, for helpful discussions and for providing very often another perspective on things that seemed initially correct and obvious but were not always so, thanks. Raoul ! thank you for the confidence and appreciation of my research results, contribution to the last research project on metabolomics and taking care of continuous financial support for several extensions of my stay. To both you and Madelone, “toda raba” for Seder Pesach and Barmitzva, reminding me of Israel and jewish holidays which I always forget their date. I would also like to thank Ruud (for very active and critical discussions, not only about work but also on other things including politics), his current PhD student Tetti (Mrs. Coco) and Cinzia (Mrs. P450), for the nice time together.

Without strawberries none of this work could have been performed. I would like to thank the strawberry breeders, Bert, Bart and Jos, who were always most helpful in growing and providing plant material and strawberries in times when they were very difficult to obtain.

In 1996 I received financial support from the Oppenheim-Fishcelfonds, which allowed me to purchase a private PC, and I would like to thank the fond for the support.

During these years, several research guests and students spent time working with me on different aspects of the project. I would like to thank the Witzburg / Willi’s group, Martina, Stefan and of course Willi, for the excellent collaboration. Their complementary expertise in food chemistry had a major impact on the data output by the combination of biochemistry and molecular biology. Willi ! I enjoyed the collaboration with you and I hope to have more of it in the future. To the students who worked with me, all of them with great enthusiasm and the desire to learn and experiment. Sun, Edgar, Marianne, Mayte, Greg, Willjan, Patrick ! it was pleasure and fun to guide you all. I hope the time with me was of help for your present and future scientific career. One of the last research guests who worked with me during the doctorate was Ashok. I

enjoyed the few months we spent together, very enthusiastic to conduct all experiments and obtain all results we wanted (GFP, RFP, TAR, NOR, NOTA, BAJI, SAMBAR, PURI, PUNE, APERNA and all the rest). Ashok ! It was great to have you here both as a friend and a colleague in the lab.

One of the major building blocks of the project was the use of DNA microarrays. The person who introduced me to the technique in a relatively early stage of its use in plants was Mark Schena and I would like to thank him for it. Analysing array data is a crucial part of the whole microarray procedure and Paul is the person that deserves great credit for developing the program in-house, as a side project, with great motivation in the weekends and evenings, whenever he could slot in some spare time. Thanks Paul ! your involvement in the statistical analysis was crucial in the early publication on the use of the method in plants. Related to the statistical analysis, Chris was always happy to listen and help in solving problems, or trying to understand how to handle the FTMS data. Thanks Chris!

One of the things I did not realise early in the Ph.D., is that converting data to a publication is not a matter of 2 or 3 weeks, but a much longer story. The person who was most helpful in this phase was Ann. Loaded with enormous ambition and a strong wish to convert every piece of information generated to either a patent or a high impact publication, she greatly helped me in the last 2 years of the Ph.D. I appreciate much and thank you for every minute you spent with me on getting the microarray and other papers published (6 versions of the first paper!), the long discussions and corrections, even after you left PRI. I learned a lot on experimental design, the importance of applied vs. basic research, including patents, and finalising a well written scientific communication. Keep in touch and, as you always say, “keep your n…. on”!

Since the beginning of this year I have started a post-doc in Andy’s group in Genomics. Andy already contributed much earlier with very helpful discussion and suggestions about my work and publications. Andy, thank you for this, for your investment in corrections and for getting me acquainted with the Arabidopsis world of science. I hope we can generate successful results together from my period in your group. I would also like to thank Willem, my current Business Unit’s head, for his role as my promoter in the final steps of my doctorate.

During the years, a large amount of foreign guests from different countries have been visiting the institute for periods of few weeks to several years, and we spent nice time together. Lucia was the first one I met, actually “warmly recommended” to meet by Arjen, who claimed that “she will take care of everything you need to know”, and he was right. Lucia ! thanks a lot for taking care of me in the beginning of my period here, trips, parties, and administration (the caverna). Italians coming to Wageningen very often outbreed … I would like to thank Martin (for sharing the Rubichen with me), Silvia (ciao pepp) and Marcel (from the village of Stockum), Diana and Jeroen, for the fun time together. Grazie mille also to Cristina and Simona, two other members of gruppo Milano, and to Mariateresa (mammateresa and Mrs. cozze). Apart from Italians, I spent joyful time with the Bulgarian crowd (pajarne commando!). Nikolay, Rumyana, Violeta, Elena,

Arnoud, Pavlina, Sergei, Tzanco ! thank you for all the good time, shopskas, burek, rakia and so on. I would like to thank also Nayelli and Stefan, another couple “made in PRI”, for the nice time together during the years. To Rajesh, Sapna (r.a.j.e.s.h. a.n.d. s.a.p.n.a.) and Seetharam! thanks for keeping my capsaicin levels high and the nice time we spent together (especially eating Indian food). I hope to see you again, maybe in India.

To my family in Israel who followed physically from far away, but actually very close, the progress of my Ph.D. Mother, Father, Ike and Snaeet, Maayan and Felix ! I know it took me longer time than expected to answer the question “ma im hadoktorat?”, and long and often visits were not my strong side. Hopefully this will change in the coming future. Along all the years, I knew that there was always you, interested, supportive, and always with a good advice. Mother and Father ! you especially followed every small step along the way, disappointments, successes, state of paper writing and their rejections and acceptances. I thank you for your support and encouragement in the initial step of performing the Ph.D. in Holland, and along the way up to date.

And last, but not least, my gratitude to Raffa, my most successful and promising result obtained in the Ph.D. Raffa !! we passed a long way together during this period, in a foreign country, with pressure and tough demands from ourselves. You were extremely supportive, helpful, compromising, backing-up and understanding, and I appreciate and thank you very much for this.

I thank all of you for your contribution to the success of this work.

Asaph

Wageningen, August 2002

PUBLICATIONS

Journals

Asaph Aharoni, Ashok P. Giri, Francel W.A. Verstappen, Robert Sevenier, Harrie A. Verhoeven, Gerard H. Rouwendaal, Maarten A. Jongsma, Wilfried Schwab & Harro J. Bouwmeester (2002) Gain and Loss of Flavour and Aroma Compounds During Crop Domestication. *Submitted* (*Chapter 6* of this thesis).

Asaph Aharoni, C.H. Ric de Vos, Harrie A. Verhoeven, Chris A. Maliepaard, Gary Kruppa, Raoul Bino, and Dayan Goodenowe (2002) Nontargeted Metabolome Analysis by Use of Fourier Transform Ion Cyclotron Mass Spectrometry (FTMS). OMICS: A Journal of Integrative Biology 6 (3) *In Press* (*Chapter 8* of this thesis).

Asaph Aharoni and Ann P. O'Connell (2002) Microarray Gene Expression Analysis During Strawberry Achenes and Receptecle Maturation. Journal of Experimental Botany, *In Press* (*Chapter 4* of this thesis).

Asaph Aharoni, Leopold C. P. Keizer, Hetty C. van den Broeck, Gregory Bois, Patrick Smit, Ric C. H. De Vos and Ann P. O'Connell (2002) Novel Insight into Vascular, Stress and Auxin Dependent and Independent Gene Expression Programs in Strawberry, a Non-Climacteric Fruit. Plant Physiology 129: 1019-1031(*Chapter 3* of this thesis).

Asaph Aharoni and Oscar Vorst (2001) DNA Microarrays for Functional Plant Genomics. Plant Molecular Biology 48 (1-2): 99-118, (*Chapter 2* of this thesis).

Asaph Aharoni, C. H. Ric De Vos, Martina Wein, Zhongkui Sun, Raffaella Greco, Arthur Kroon, Joseph N.M. Mol and Ann P. O'Connell (2001) The Strawberry FaMYB1 Transcription Factor Suppresses Anthocyanin and Flavonol Accumulation in Transgenic Tobacco. The Plant Journal 28(3): 319-332 (*Chapter 7* of this thesis).

Asaph Aharoni, Leopold C. P. Keizer, Harro J. Bouwmeester, Zhongkui Sun, Mayte Alvarez Huerta, Harrie A. Verhoeven, Jan Blaas, Adèle M. M. L. van Houwelingen, Ric C. H. De Vos, Hilko van der Voet, Ritsert C. Jansen, Monique Guis, Jos Mol, Ronald W. Davis, Mark Schena, Arjen J. van Tunen, and Ann P. O'Connell (2000) Identification of the SAAT Gene Involved in Strawberry Flavor Biogenesis by Use of DNA Microarrays. The Plant Cell 12: 647-662 (*Chapter 5* of this thesis).

B. Lemieux, A. Aharoni and M. Schena (1998) Overview of DNA chip technology. Molecular Breeding 4: 277-289.

Schwab W, Aharoni A, Raab T, Perez AG, Sanz C. (2001) Cytosolic aldolase is a ripening related enzyme in strawberry fruits (Fragaria x ananassa). Phytochemistry 56(5): 407-415.

Nicole L.W. van Hal, Oscar Vorst, Adele M.M.L. van Houwelingen, Esther J. KoK, Ad Peijnenburg, Asaph Aharoni, Arjen J. van Tunen and Jaap Keijer (2000) The application of DNA microarrays in gene expression analysis. Journal of Biotechnology 78: 271-280.

Book Chapters

Asaph Aharoni and Ann P. O'Connell (2002) Identification of strawberry flavour related genes by the use of DNA microarrays. In: S. M. Jain, D. S. Brar and B. S. Ahloowalia (Editors), Molecular techniques in crop improvement, Kluwer Acad. Publish., Netherlands.

Asaph Aharoni and Ann P. O'Connell (2002) Use of DNA microarrays in the identification of genes involved in strawberry flavour formation. In: J. F. Jackson, H. F. Linskens and R. B. Inman (Editors), Molecular methods of plant analysis (Vol. 21), Analysis of taste and aroma, Springer-Verlag, Berlin, Heidelberg.

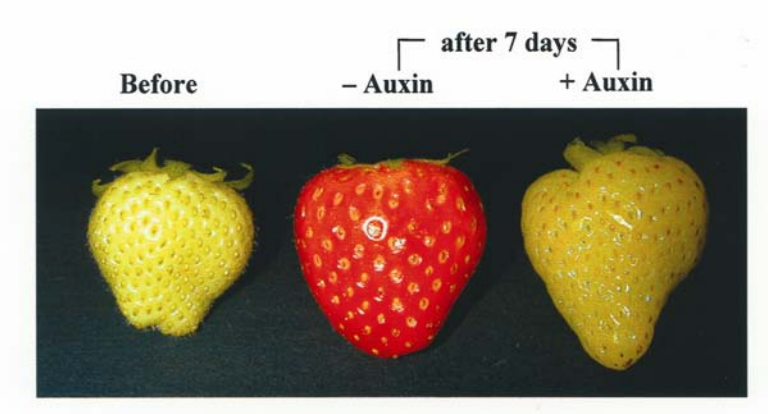
Patent Applications

Aharoni et al., (2001)
Terpene synthase/cyclase and olefin synthase and uses thereof (Submitted- January 2001; PCT-NL02-00089)

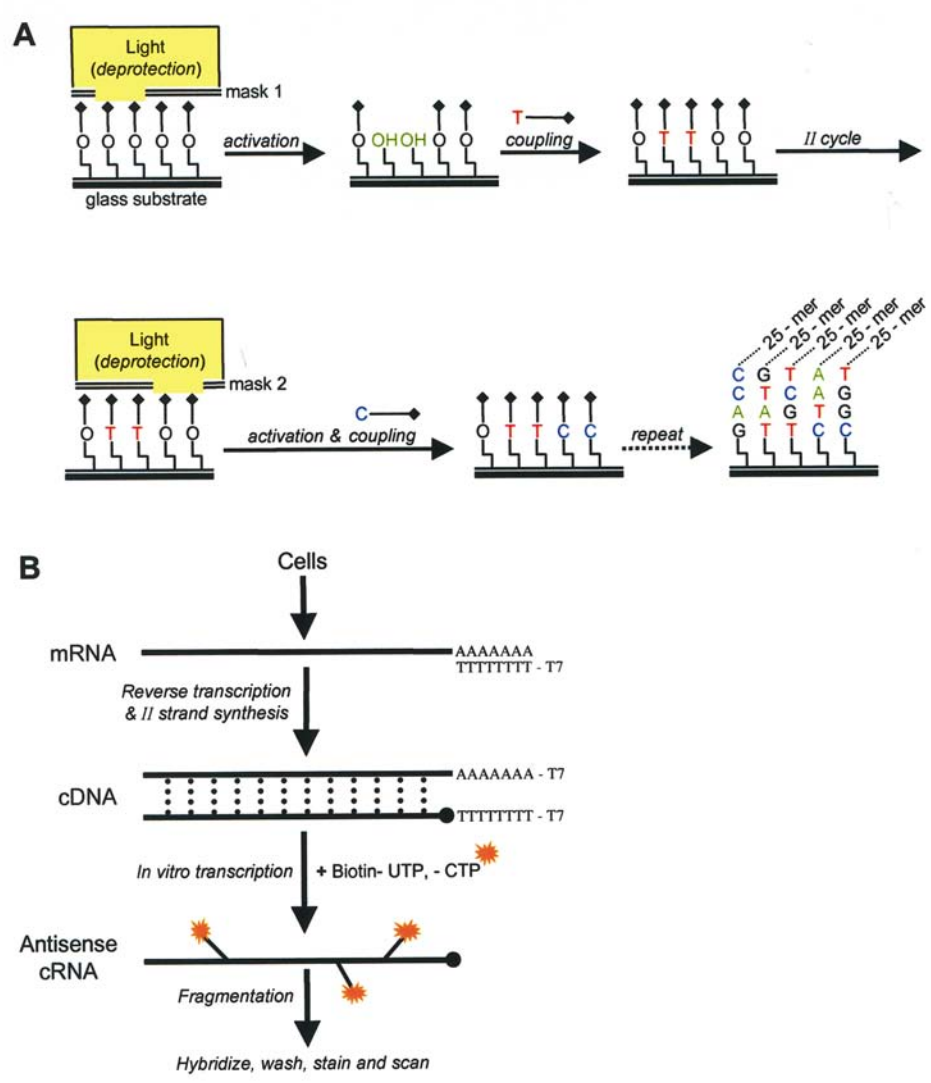
Aharoni et al., (2000)
Fruit flavour related genes and use thereof (publication date, June 2000; WO219991202)

CURRICULUM VITAE

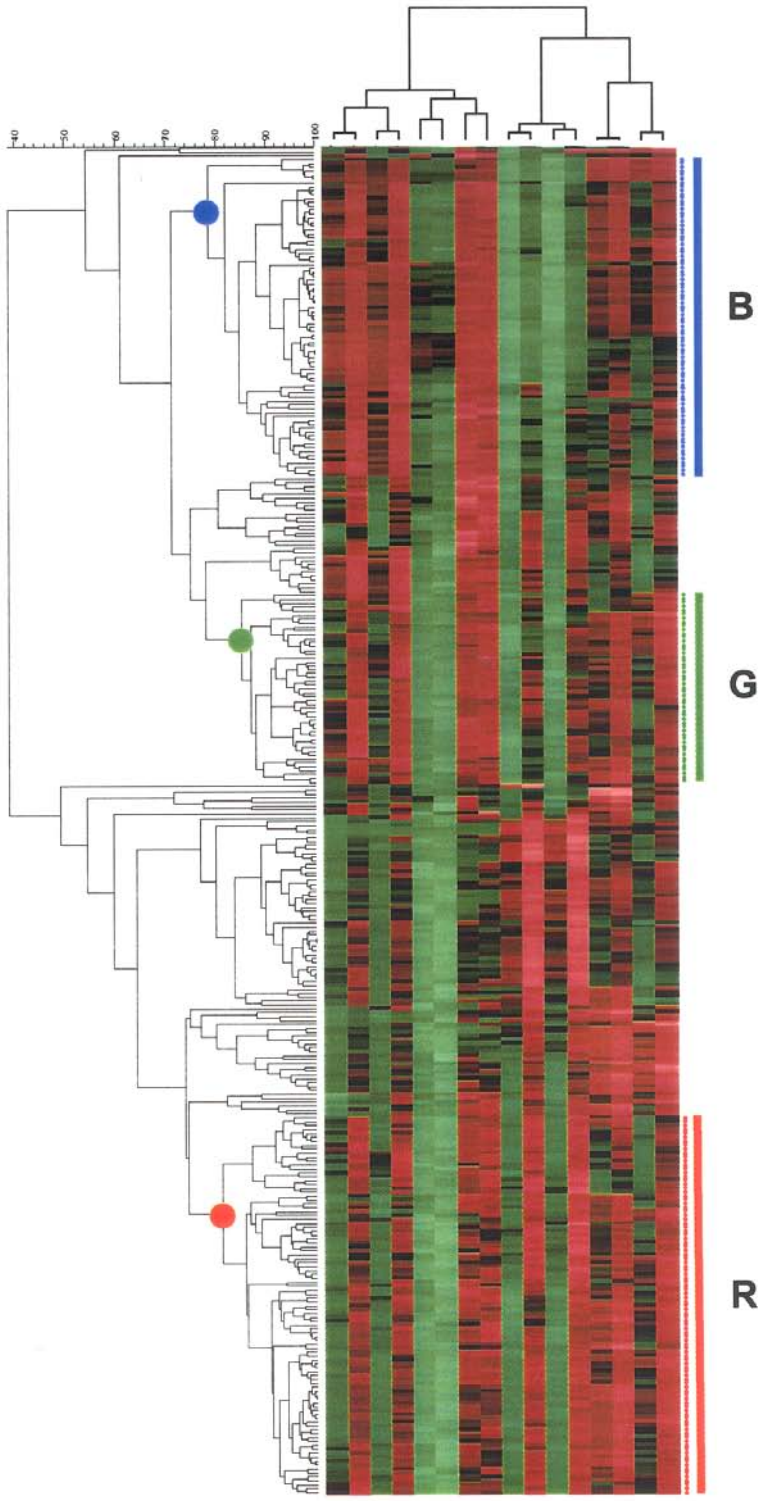
Asaph Aharoni, the author of this thesis, was born in Israel on the 25th of November 1966. He grew up in the city of Petach-Tikva, located in the centre of Israel, near Tel-Aviv. In 1985, after finishing high school, he started a 4 years army service. Once his military service terminated he carried out a one and a half year trip, travelling all around South America. After returning to Israel he started studying for a B.Sc. degree in Agriculture Sciences in the Hebrew University of Jerusalem, Faculty of Agriculture, Rehovot. From 1994 till 1996 he pursued an M.Sc. program in Agriculture Sciences in the same university at the Department of Horticulture. He was awarded CUM LAUDE for his thesis work entitled "Development of an Efficient Regeneration and Transformation Methods for Carnation and Gypsophila Plants". In May 1996 he moved to The Netherlands and started a PhD program in CPRO-DLO (currently Plant Research International), at the department of Cell Biology (currently Cell Cybernetics). The results of his research on the molecular genetic processes underlying strawberry fruit development and ripening are described in this thesis. Since January 2002 he joined as a post-doc the group of Dr. Andy Pereira in the Business Unit Genomics at Plant research International. His current research uses the model plant Arabidopsis to dissect the genetic regulation of metabolic pathways in plants.



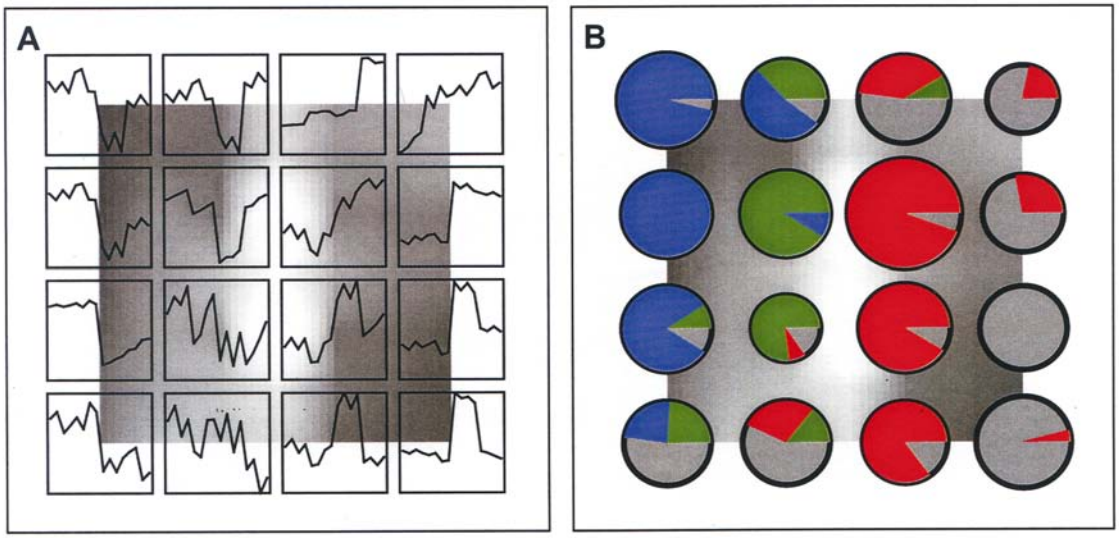
CHAPTER 1 - Figure 2



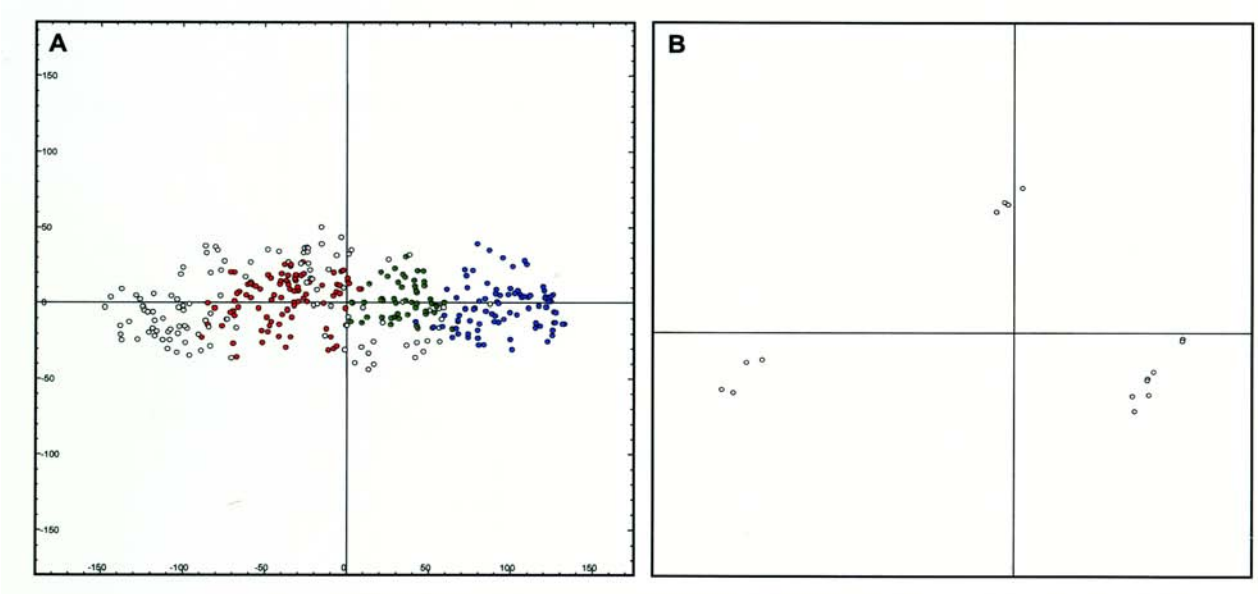
CHAPTER 2 - Figure 3



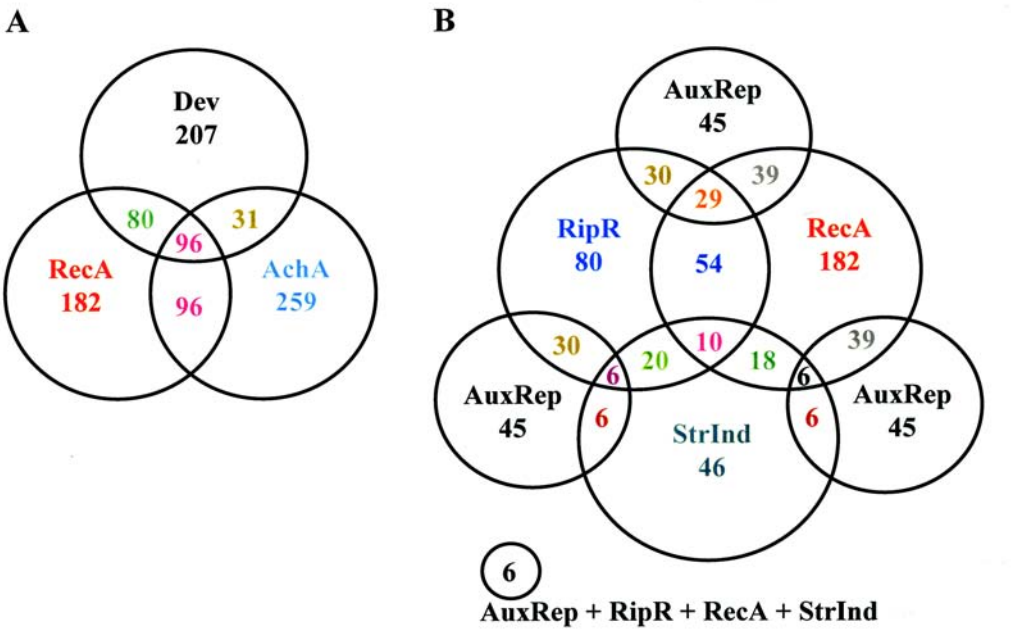
CHAPTER 2 - Figure 5



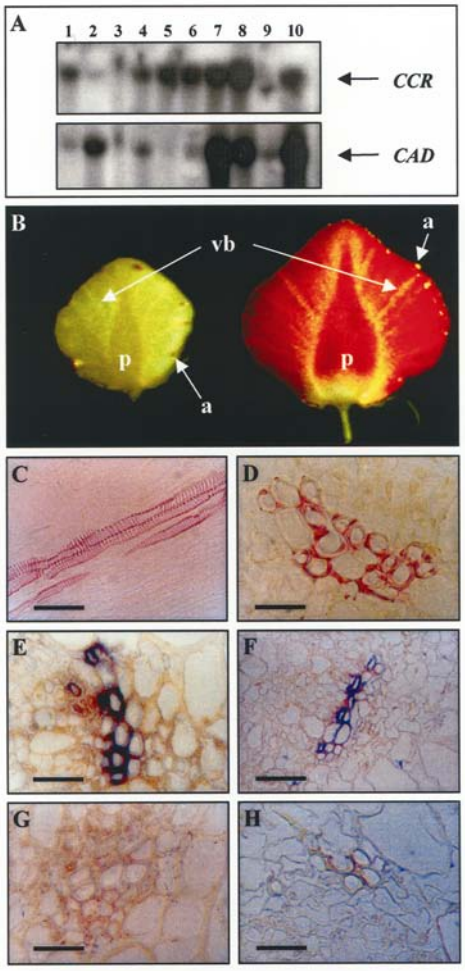
CHAPTER 2 - Figure 6



CHAPTER 2 - Figure 7



CHAPTER 3 - Figure 2



CHAPTER 3 - Figure 3

CHAPTER 3 - Table III. Identification of auxin dependent genes and processes in strawberry

PRI Clone	Auxin Response (fold)[a]	R/A[b]	Homolog Definition (GenBank accession number) (E-value)[c]	Putative Function
Auxin Repressed and Ripening Regulated Genes				
A104	5.9	8.4	Glutathione *S*-transferase (Y07721) (6e-67)	Anthocyanin sequestration
A135	6.7[d]	22.2	Chalcone synthase 2 (D26594) (6e-69)	Flavonoid pathway
C122	6.5	4.3	Profilin (AF129427) (4e-49)	Actin binding
C179	4.9[d]	6.5	Flavanone 3-hydroxylase (X69664) (9e-75)	Flavonoid pathway
C23	2.3	9.7 ↑	60S Ribosomal protein L13E (Z22620) (6e-60)	Unknown
D117	1.9[d]	3.7	Acyl carrier protein (AJ001446) (1e-29)	Fatty acid synthesis
D23	1.9[d]	10.2	NADPH-dependent oxidoreductase (AF108438) (2e-43)	Phytoalexin synthesis
E149	1.9	5.0 ↑	Quinone reductase-like protein (AL163972) (3e-36)	Reduction of quinones
E27	6.5[d]	16.3 ↑	Alcohol acyltransferase (AF193789) (3e-50)	Ester formation
E30	1.8[d]	8.5	Pectate lyase (U63550) (3e-54)	Cell wall degradation
E80	2.2[d]	8.7	Endo-1,4-beta-glucanase (AF074923) (3e-55)	Cell wall degradation
F1	2.0	n.d.	LTP (Q43681) (2e-22)	Lipid transfer
F102	3.6[d]	8.4	O-methyltransferase (AF220491) (1e-18)	Metabolites methylation
F157	2.6	20.3	Chalcone-flavonone isomerase (AB010692) (6e-38)	Flavonoid pathway
F193	2.0	3.3 ↑	Cinnamyl alcohol dehydrogenase (U63534) (2e-53)	Lignin biosynthesis
G13	2.7	8.8	Dioxygenase (AC007504) (4e-19)	Unknown
G175	2.1	3.5	Isoflavone reductase related (AF071477) (9e-46)	Phytoalexin synthesis
G84	7.2	3.7	Beta-tubulin (D63138) (3e-80)	Cytoskeleton organisation
H142	8.7	21.1	Dioxygenase (U97530) (1e-08)	Unknown
H159	4.8[d]	3.6	Ripening-induced protein (AJ001445) (9e-28)	Unknown
H51	2.0[d]	6.2	Pyruvate decarboxylase (AF193791) (2e-73)	Aldehyde formation
H61	3.3[d]	16.0	Flavanone 3-hydroxylase (X69664) (7e-73)	Flavonoid pathway
H81	2.4	5.5	Unknown (AL138646) (2e-30)	Unknown
JB136	2.9	4.6 ↑	Farnesyl pyrophosphate synthase (U15777) (2e-54)	Terpene metabolism
JB173	2.5[d]	12.6	Chalcone reductase-like (AC007259) (0.002)	Phytoalexin synthesis
JB19	2.0[d]	4.6	Malonyl-CoA decarboxylase (AF069441) (2e-47)	Fatty acid synthesis
JB202	3.0[d]	4.0	Cysteine proteinase (Z99954) (3e-68)	Protein breakdown
JB77	3.6	19.1	Anthocyanidin synthase (X7160) (4e-31)	Flavonoid pathway
Auxin Induced and Early-Mid Development Regulated Genes				
C176	2.2	n.d.	Copper homeostasis factor (AL163763) (8e-23)	Protection against oxidative stress
E1	1.7	3.5	Metallothionein-like protein (AJ001444) (2e-15)	Binding various heavy metals
F29	1.9	n.d.	ABC transporter-like (AL138656) (3e-32)	Unknown
H117	1.8	n.d.	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (X83499) (2e-56)	Methionine synthesis
H163	1.8	n.d.	Pectin esterase-like (AJ237985) (1e-21)	Cell wall metabolism
JB67	2.1	n.d.	SAM synthetase (AF271220) (7e-52)	*S*-adenosyl methionine synthesis

Pigmentation · Flavour · Stress Response
Fatty Acid Metabolism · Cell Structure · Unknown
Cell Wall Metabolism · Other

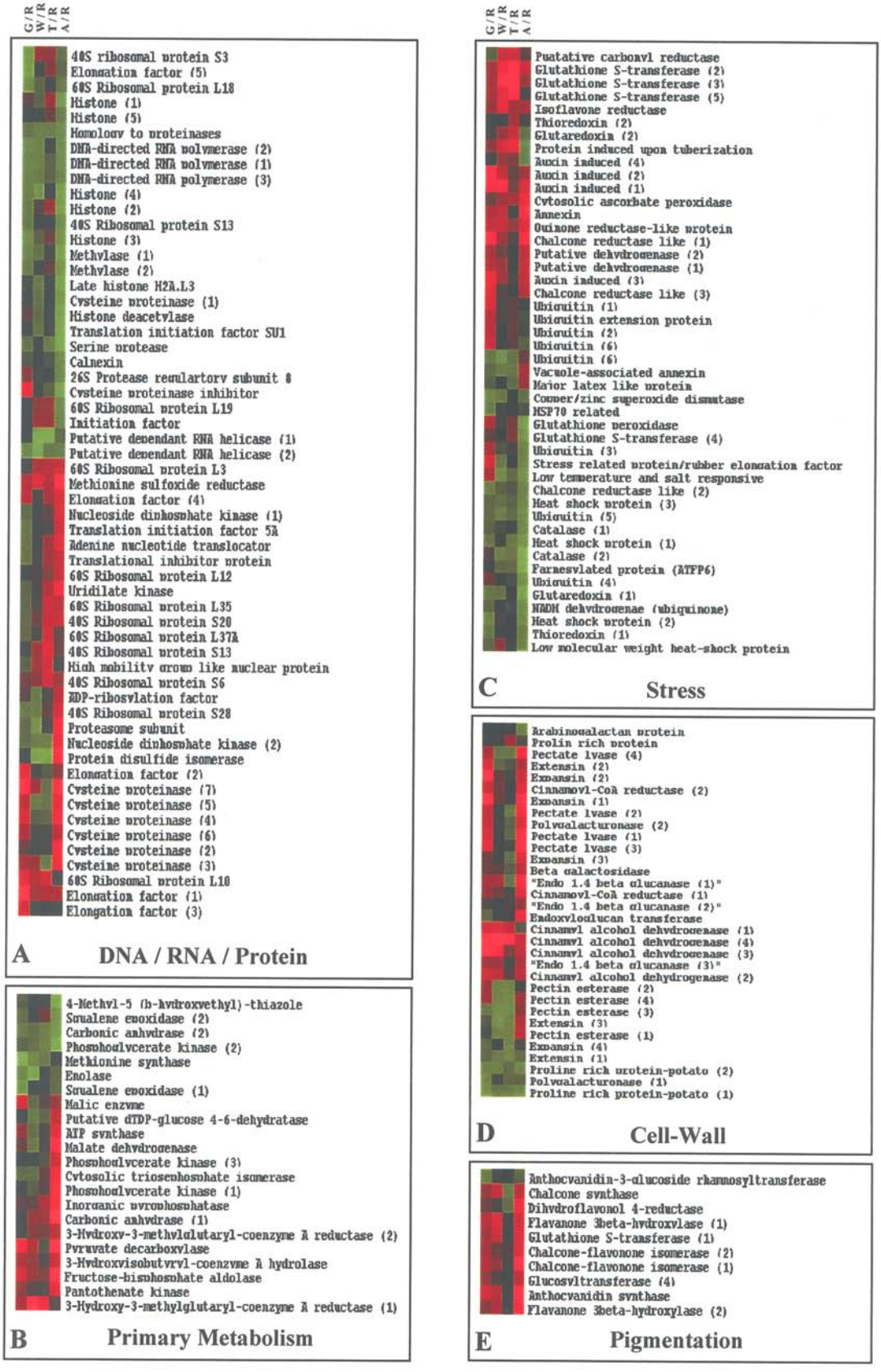
[a]Expression ratio detected in microarray experiment VI examining response to auxin. In the first set of genes it represents fold repression due to auxin treatment, while in the second set it represents fold induction by the auxin treatment.

[b]Expression ratio in the fourth microarray experiment (IV) comparing expression in receptacle tissue (R) vs. achene tissue (A). Genes identified in experiment V as stress induced are marked by ↑. n.d., No difference in expression between achene and receptacle.

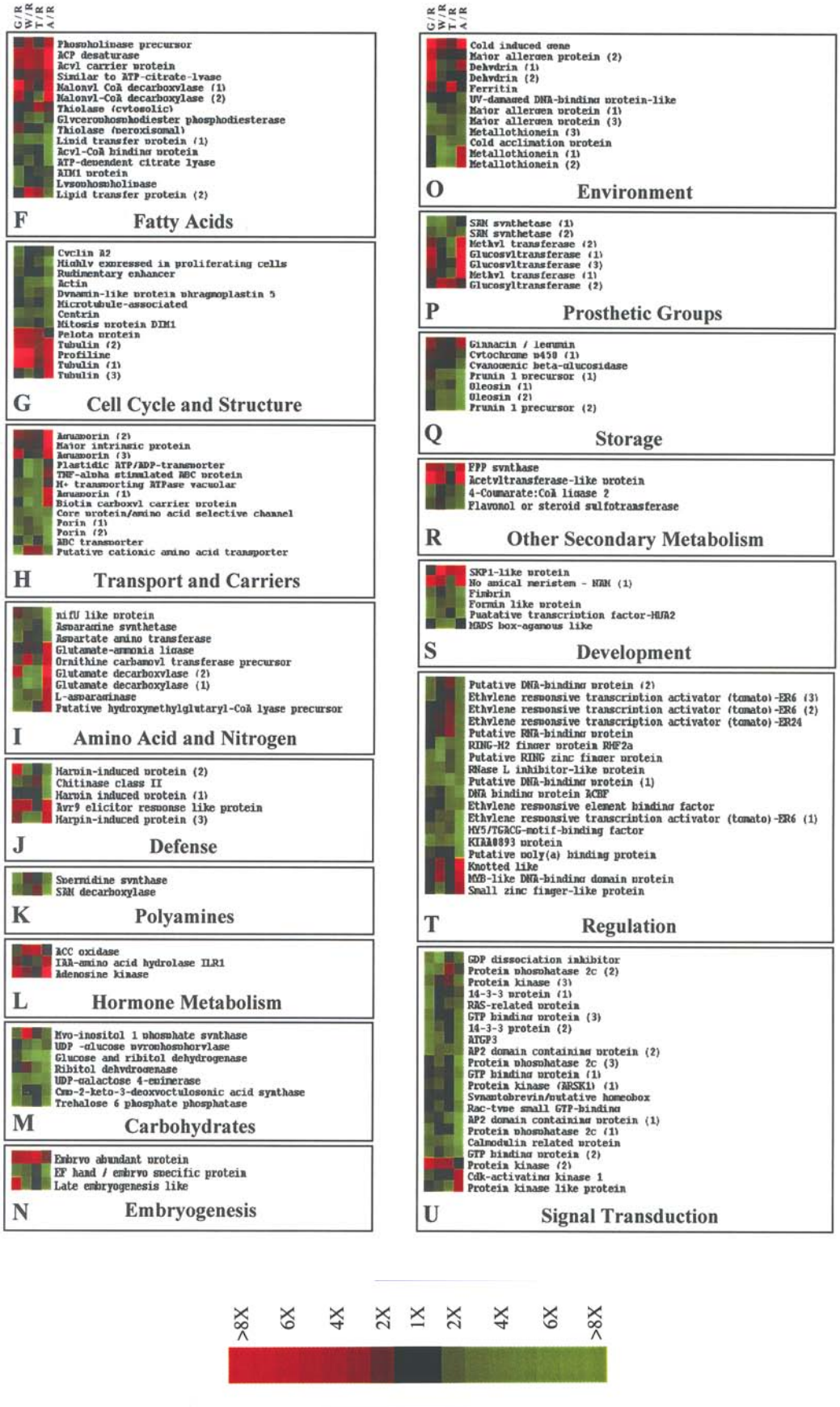
[c]Definition, accession (nucleotide sequence) and E-value (Altschul et al., 1990) of the first BLAST X homolog.

[d]Described earlier as ripening regulated and auxin repressed (e.g. Manning, 1998).

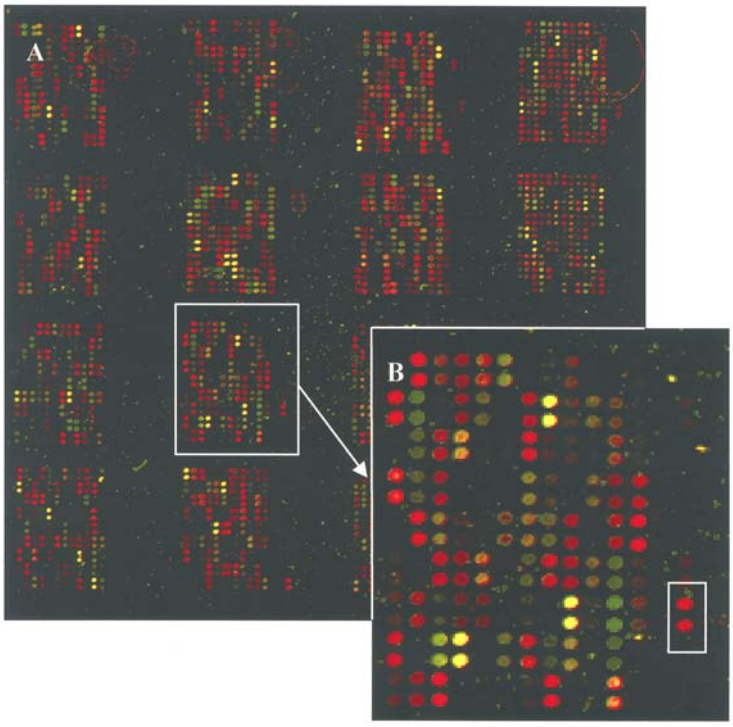
According to their putative function, genes were associated to one of the processes depicted below the table.



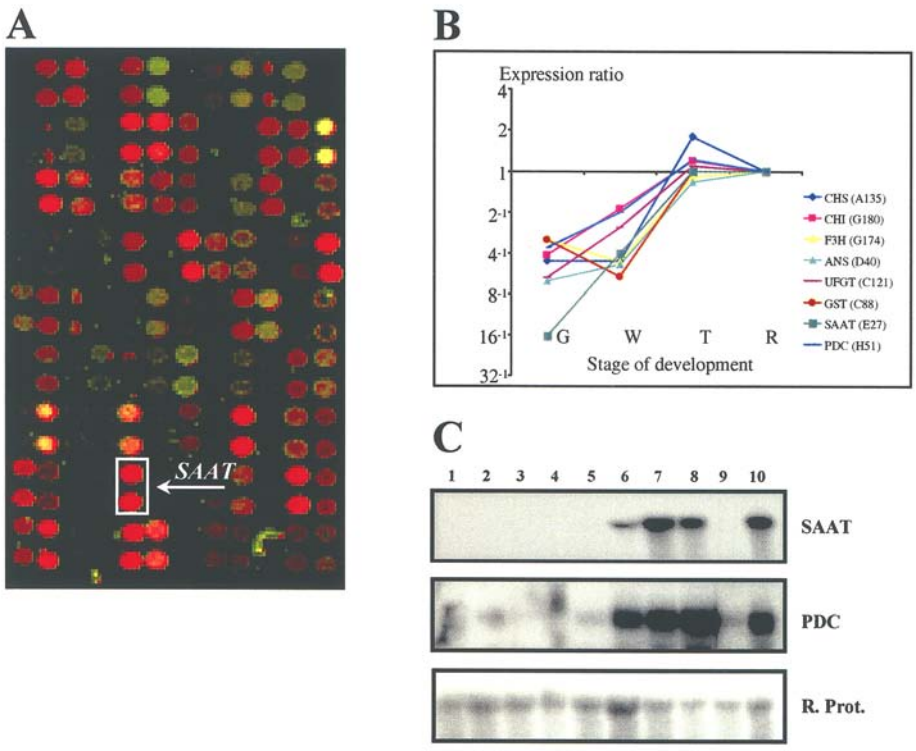
CHAPTER 4 - Figure 3



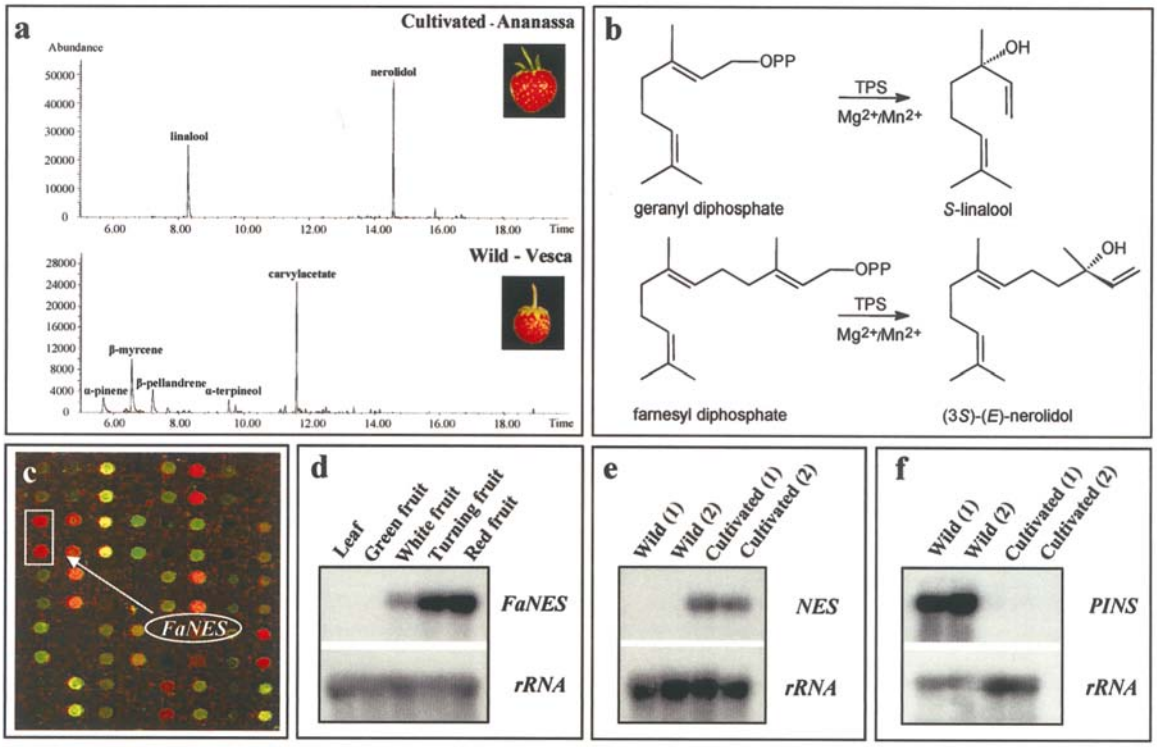
CHAPTER 4 - Figure 3 (continued)



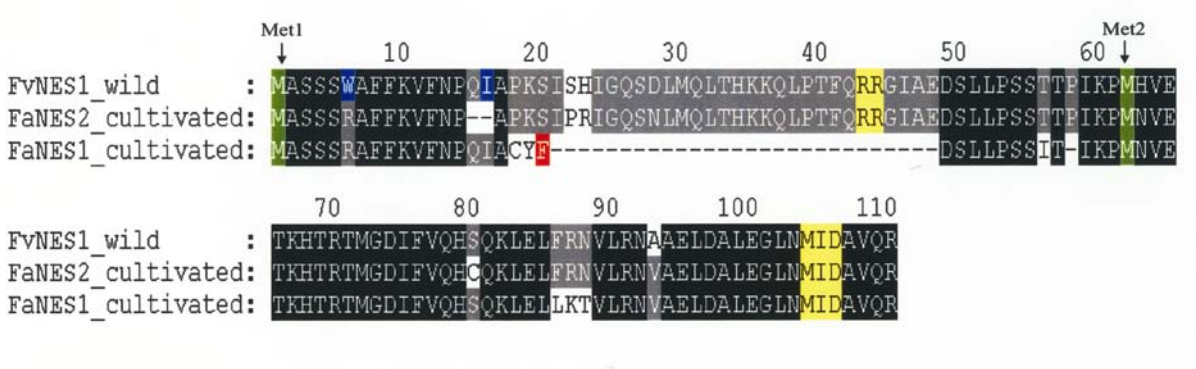
CHAPTER 5 - Figure 2



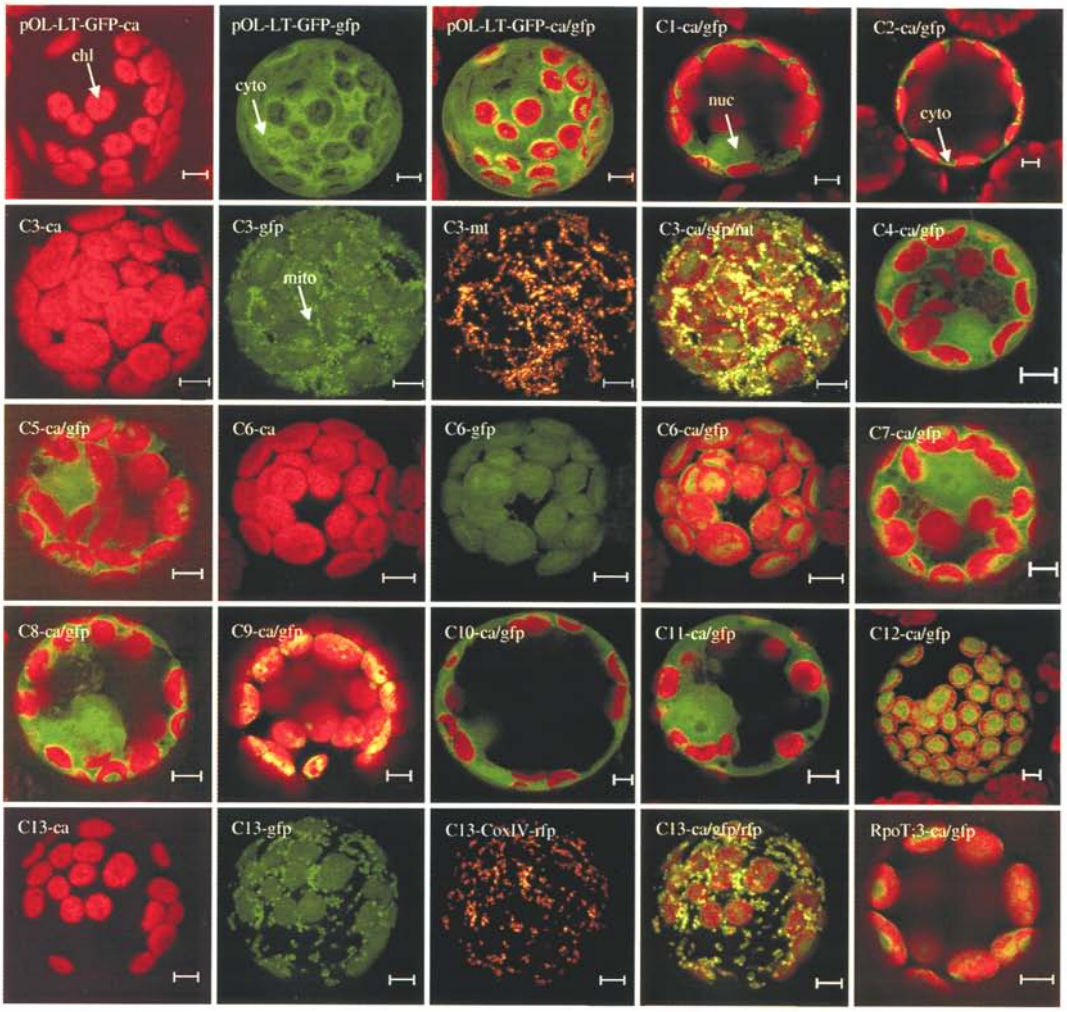
CHAPTER 5 - Figure 4

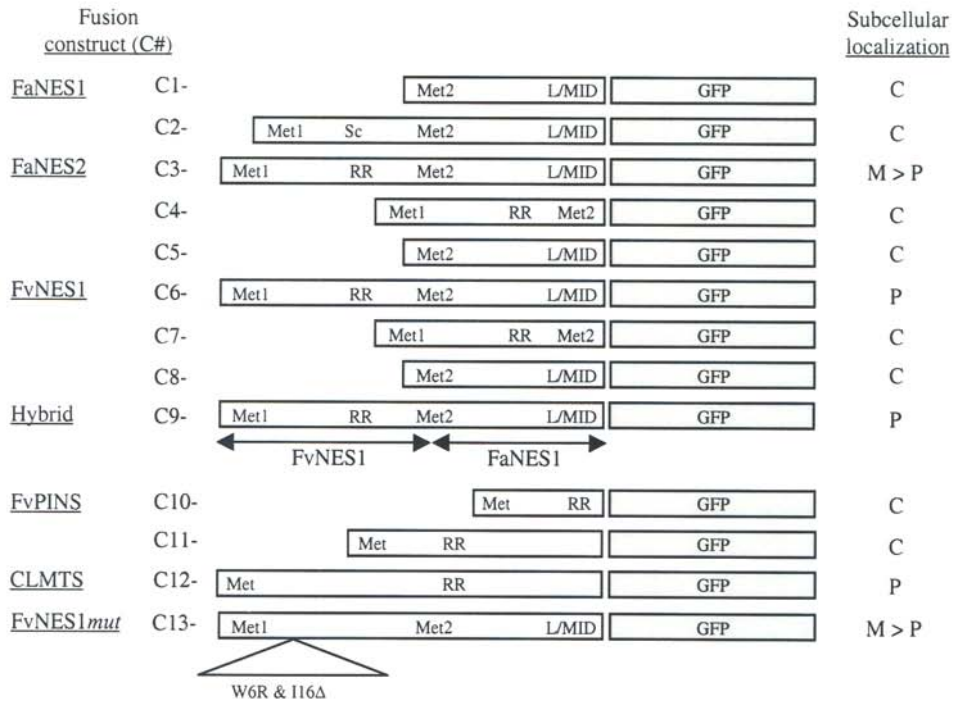


CHAPTER 6 - Figure 1

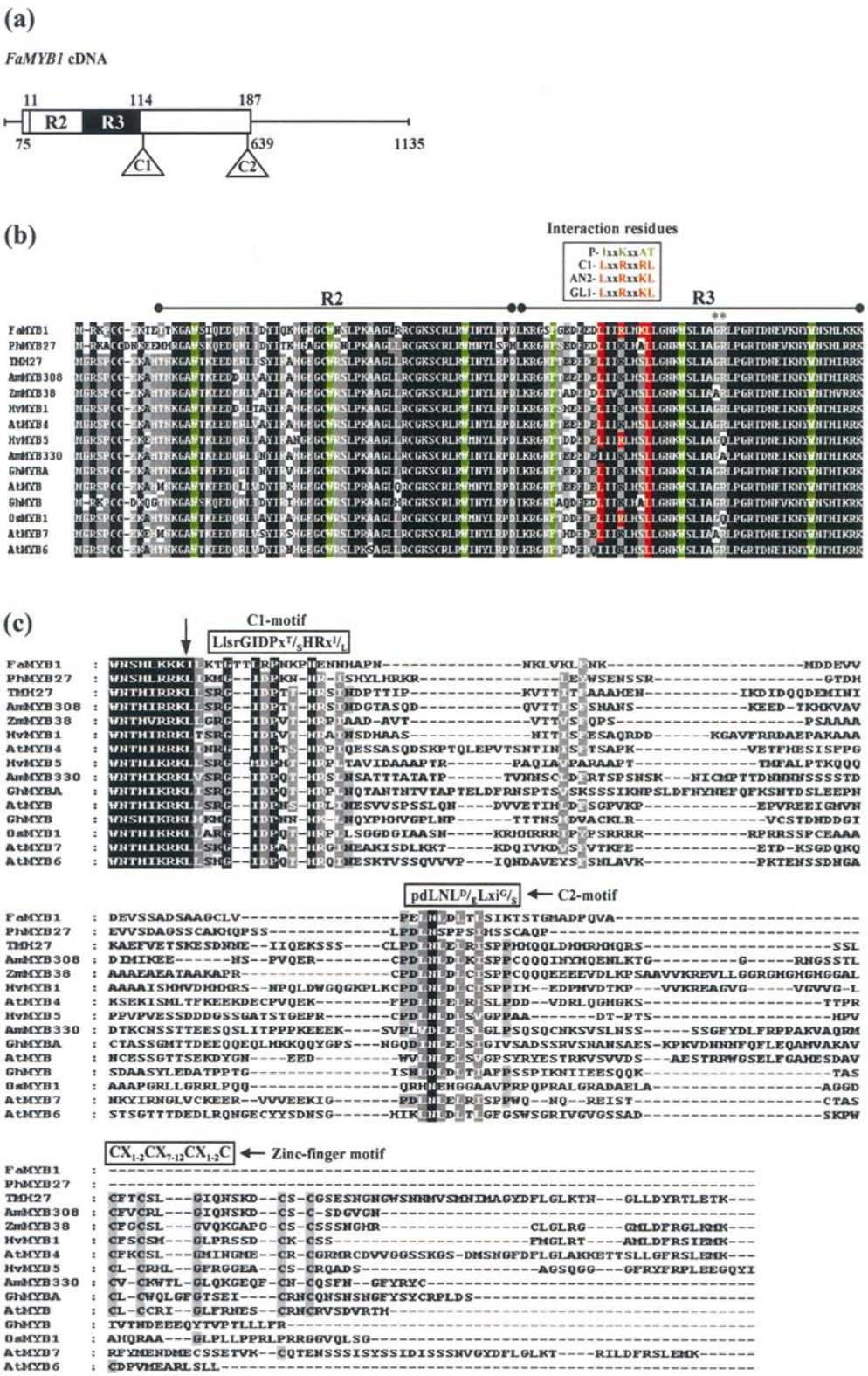


CHAPTER 6 - Figure 2

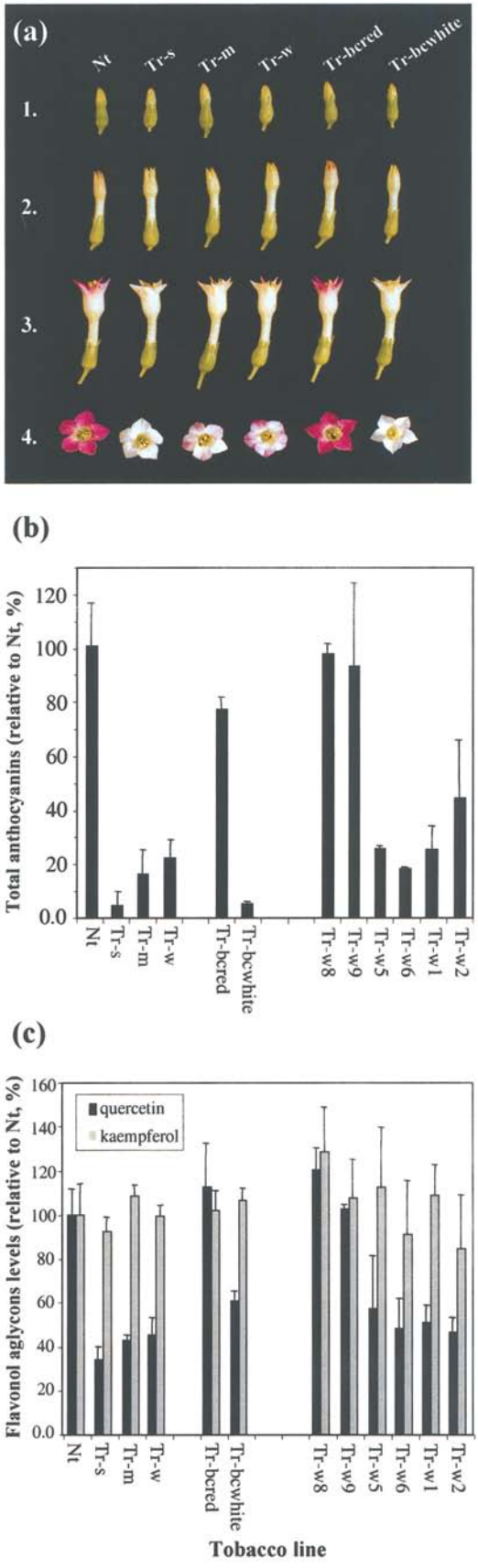




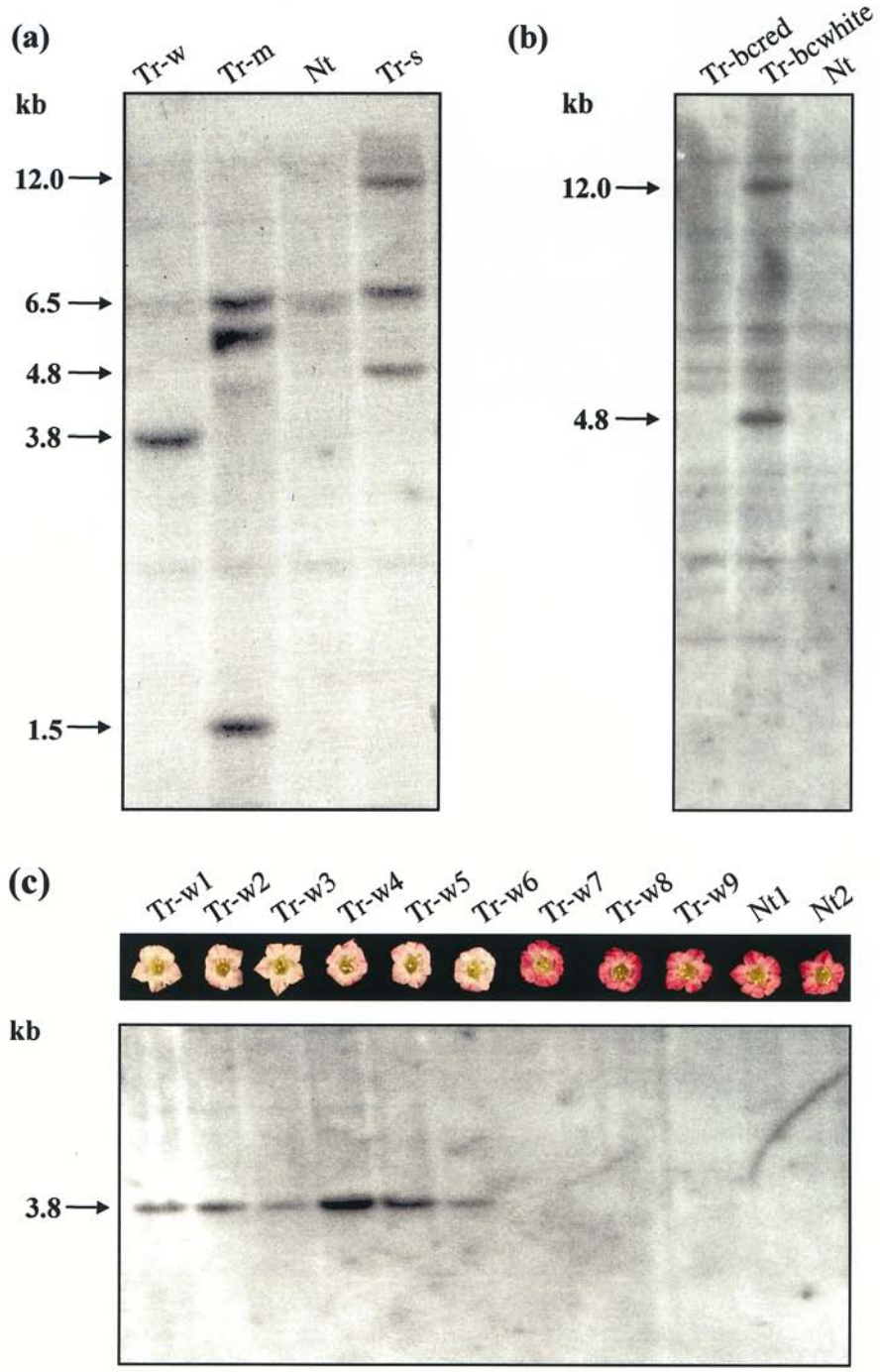
CHAPTER 6 - Figure 6



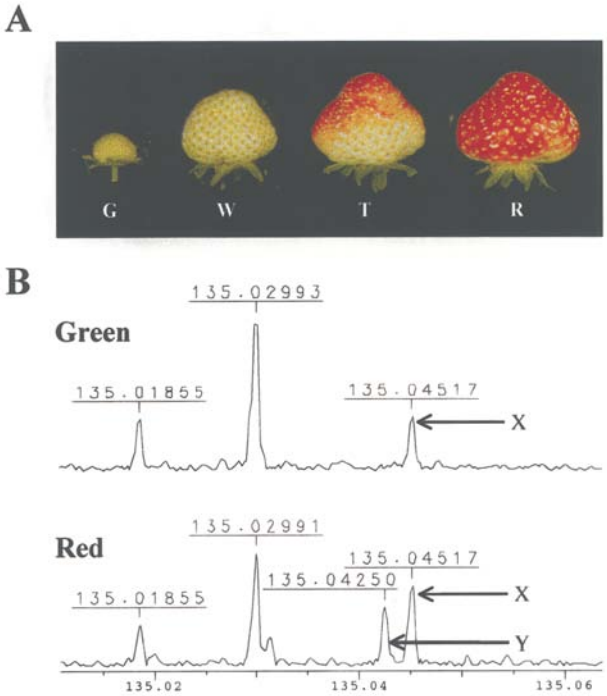
CHAPTER 7 - Figure 1



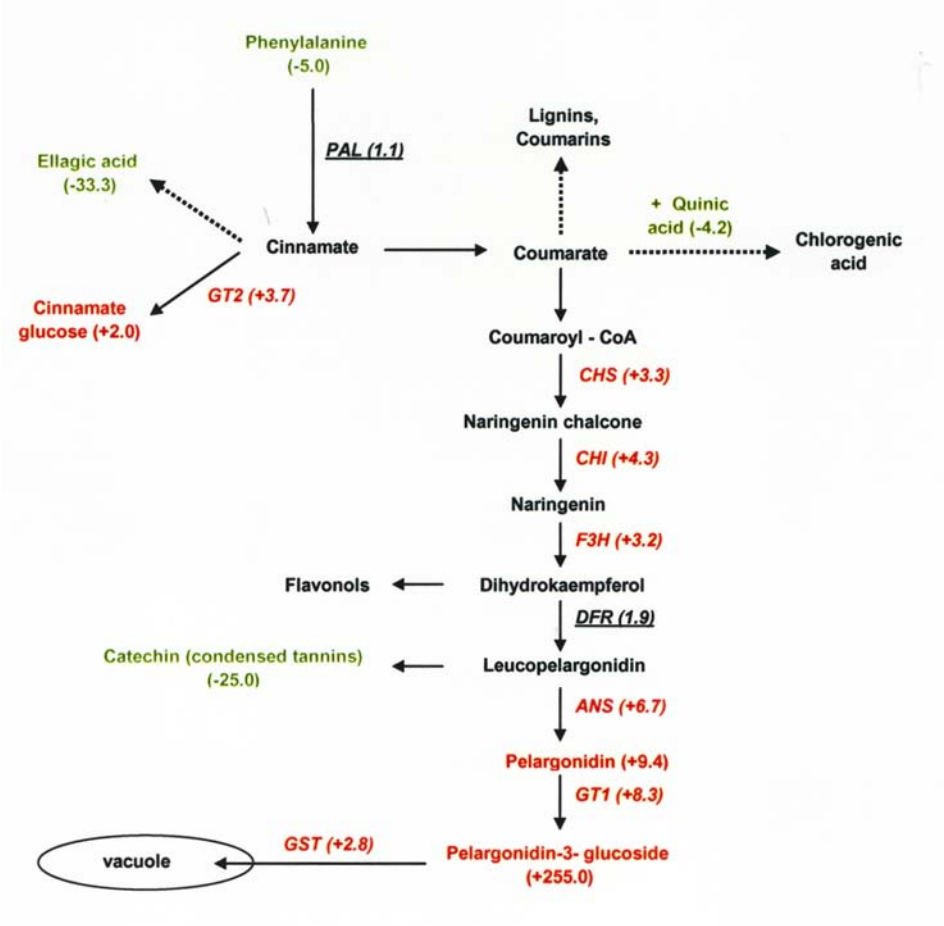
CHAPTER 7 - Figure 3



CHAPTER 7 - Figure 4



CHAPTER 8 - Figure 1



CHAPTER 8 - Figure 5

TABLE 2. EXPRESSION PROFILES OF PUTATIVE METABOLITES IDENTIFIED DURING STRAWBERRY FRUIT DEVELOPMENT[a]

OM	IM	EX	R/G	T/G	W/G	EF	Putative Metabolite	CG
Group 1: Early expression						**Profile a**		
300.9990	AP-	AN	0.03	0.03	0.03	$C_{14}H_6O_8$	Ellagic acid	P. Acid
137.0243	AP-	AN	0.09	0.08	0.09	$C_7H_6O_3$	Hydroxybenzoic acid	P. Acid
175.1188	ES+	50	0.06	0.06	0.10	$C_6H_{14}N_4O_2$	Arginine	A. acid
167.0349	AP-	AN	0.08	0.05	0.06	$C_8H_8O_4$	Vanillic acid	P. Acid
133.0142	ES-	AN	0.04	0.04	0.04	$C_4H_6O_5$	Malic acid	O. acid
2890717	ES-	50	0.04	0.07	0.12	$C_{15}H_{14}O_6$	Catechin (epi)/leucopelargonidin	Phenolic
179.0349	AP-	50	0.16	0.14	0.11	$C_9H_8O_4$	Caffeate	P. Acid
175.0764	AP-	AN	0.13	0.13	0.12	$C_{11}H_{12}O_2$	Ethyl cinnamate	Ester
464.0882	ES-	50	0.13	0.13	0.13	$C_{21}H_{20}O_{12}$	Quercetin glucoside	Flavonoid
161.0608	AP-	AN	0.19	0.1	0.14	$C_{10}H_{10}O_2$	Methyl cinnamate	Ester
289.0707	AP+	50	0.11	0.15	0.28	$C_{15}H_{12}O_6$	Dihydrokaempferol	Flavonoid
305.0664	AP-	AN	0.37	0.37	0.37	$C_{15}H_{14}O_7$	Leucocyanidin	Flavonoid
193.0640	ES+	AN	0.22	0.22	0.22	$C_6H_{12}N_2O_3S$	*N*-Alanylcysteine	A. Acid d.
155.0338	AP+	50	0.31	0.26	0.25	$C_7H_6O_4$	Gentisic/Protocatechuic acid	P. Acid
231.0297	AP-	50	0.25	0.25	0.25	$C_{12}H_8O_5$	Malusfuran	Furan
101.0961	AP+	AN	0.25	0.25	0.25	$C_6H_{12}O$	(E)2-hexenol / hexanal	Alcohol
331.0671	AP-	50	0.23	0.23	0.23	$C_{13}H_{16}O_{10}$	Glucogallin	Phenolic
220.1180	ES+	AN	0.27	0.27	0.27	$C_9H_{17}NO_5$	Pantothenic acid	Vitamin
141.0182	ES+	AN	0.49	0.48	0.46	$C_6H_4O_4$	Dihydroxybenzoquinone	Phenolic
305.0654	AP+	50	0.42	0.42	0.42	$C_{15}H_{12}O_7$	Dihydroquercetin	Flavonoid
						Profile b		
144.1746	ES+	50	0.2	0.33	0.68	$C_9H_{21}N$	Heptylamine	Amine
193.0706	AP+	50	0.24	0.34	0.64	$C_7H_{12}O_6$	Quinic acid	O. Acid
						Profile c		
193.0342	AP+	AN	0.01	0.78	0.81	$C_6H_8O_7$	Citrate / Isocitrate	O. acid
175.0236	AP+	50	0.10	0.74	1.63	$C_6H_6O_6$	Aconitate / Dehydroascorbic acid	O. acid
147.0288	AP+	AN	0.18	0.99	0.64	$C_5H_6O_5$	Alpha-ketoglutarate	O. acid
210.0607	AP+	AN	0.19	0.76	0.67	$C_6H_{11}NO_7$	xylo-5-Hexulosonic acid	S. Acid
138.0550	ES+	50	0.43	0.64	1.26	$C_7H_7NO_2$	Anthranilate	Acid
210.1277	AP+	AN	0.30	0.85	0.88	$C_{15}H_{13}N$	*N*-Propyl carbazole	Alkaloid
Group 2: Intermediate expression						**Profile d**		
166.0863	ES+	50	0.20	0.75	0.15	$C_9H_{11}NO_2$	Phenylalanine	A. acid
132.1019	ES+	AN	0.41	0.71	0.29	$C_6H_{13}NO_2$	Leucine	A. acid
148.0604	ES+	50	0.36	0.56	0.36	$C_5H_9NO_4$	Glutamate	A. acid
118.0862	ES+	50	0.30	0.67	0.38	$C_5H_{11}NO_2$	Valine	A. acid
						Profile e		
123.0451	AP-	AN	0.9	0.9	0.11	$C_7H_8O_2$	3-methylcatechol	Phenol
147.0764	ES+	50	0.61	0.87	0.23	$C_5H_{10}N_2O_3$	Glutamine	A. acid
134.0448	AP+	50	0.64	0.72	0.44	$C_4H_7NO_4$	Aspartate	A. acid
109.0284	AP+	AN	1.18	0.62	0.33	$C_6H_4O_2$	1,4-Benzoquinone	Phenolic
175.0248	ES-	50	0.90	0.53	0.45	$C_6H_8O_6$	Ascorbic Acid	O. acid
						Profile f		
123.0440	AP+	AN	2.08	0.43	0.30	$C_7H_6O_2$	Benzoic acid	P. acid
271.2631	AP+	AN	2.01	0.37	0.33	$C_{17}H_{34}O_2$	Methyl hexadecanoate	Ester
Group 3: Late expression						**Profile g**		
433.1127	ES+	50	255.00	12.90	1.00	$C_{21}H_{21}O_{10}$	Pelargonidin-3-glucoside	Flavonoid
211.0601	AP+	50	2.35	5.37	1.19	$C_{10}H_{10}O_5$	Hydroxyferulate	Aromatic
117.0182	AP+	50	3.91	9.09	1.29	$C_4H_4O_4$	Fumarate	O. acid
179.0560	AP-	50	2.89	2.59	1.22	$C_6H_{12}O_6$	Glucose/Fructose	Sugar
307.1023	ES+	50	3.33	3.14	1.85	$C_{12}H_{18}O_9$	Galactopyranose	Sugar
149.0455	AP-	50	3.30	2.70	1.10	$C_5H_{10}O_5$	D-ribose	Sugar
						Profile h		
325.0929	AP-	AN	9.59	1.51	0.27	$C_{15}H_{18}O_8$	Coumarate glucose	Phenol
165.0545	AP+	50	3.27	0.86	0.42	$C_9H_8O_3$	4-Coumarate	P. acid

CHAPTER 8 - Table 2

TABLE 2. (CONT'D) EXPRESSION PROFILES OF PUTATIVE METABOLITES IDENTIFIED DURING STRAWBERRY FRUIT DEVELOPMENT[a]

OM	IM	EX	R/G	T/G	W/G	EF	Putative Metabolite	CG
				Profile i				
281.2474	AP+	AN	20.80	1.93	0.97	$C_{18}H_{32}O_2$	Linoleic acid	F. Acid
255.2319	AP+	AN	24.30	1.34	0.85	$C_{16}H_{30}O_2$	Palmitoleic acid	F. Acid
283.2631	AP+	AN	32.56	1.50	0.99	$C_{18}H_{34}O_2$	Oleic acid	F. Acid
243.2319	AP+	AN	28.80	0.91	0.62	$C_{15}H_{30}O_2$	Pentadecanoic acid	F. Acid
257.2474	AP+	AN	40.20	1.67	0.80	$C_{16}H_{32}O_2$	Palmitic acid	F. Acid
229.2162	AP+	AN	29.80	1.38	1.00	$C_{14}H_{28}O_2$	3-methylbutyl nonanoate	Ester
285.2787	AP+	AN	18.03	1.87	1.24	$C_{18}H_{36}O_2$	Stearic acid	F. acid
291.1074	ES+	AN	10.04	1.00	1.00	$C_{12}H_{18}O_8$	Furaneol (R=D-glucopyranose)	Furan
377.1078	ES+	50	11.00	1.00	1.00	$C_{15}H_{20}O_{11}$	Furaneol (R=D-gluco-manosyl)	Furan
159.1380	AP+	AN	5.20	1.00	1.00	$C_9H_{18}O_2$	Nonanoic acid	F. Acid
173.1536	AP+	AN	5.49	1.00	1.00	$C_{10}H_{20}O_2$	Capric acid	F. Acid
313.3101	AP+	AN	7.85	1.00	1.00	$C_{20}H_{40}O_2$	Eicosanoic acid	F. Acid
279.2318	AP+	AN	8.52	1.97	1.10	$C_{18}H_{30}O_2$	Linolenic acid	F. Acid
149.0597	AP+	50	5.86	1.17	0.87	$C_9H_8O_2$	Cinnamate	P. Acid
201.1849	AP+	AN	5.69	1.00	1.00	$C_{12}H_{24}O_2$	1-octyl butanoate	Ester
215.2004	AP+	AN	6.67	1.00	1.00	$C_{13}H_{26}O_2$	1-methyhexyl hexanoate	Ester
227.2005	AP+	AN	6.93	0.91	0.90	$C_{14}H_{26}O_2$	Cis-hex-3-enyl octanoate	Ester
299.2944	AP+	AN	7.91	0.76	0.68	$C_{19}H_{38}O_2$	Methylstearate	Ester
209.0819	AP-	AN	5.8	1.00	1.00	$C_{11}H_{14}O_4$	Sinapyl alcohol	Phenolic
193.0494	AP+	50	5.31	1.94	0.99	$C_{10}H_8O_4$	Scopoletin	Benzopyranoid
153.1022	AP+	AN	6.79	1.00	1.00	$C_8H_{12}N_2O$	2-Isopropyl-3-methoxypyrazine	Alkaloid
263.2368	AP+	AN	6.65	1.57	0.59	$C_{18}H_{30}O$	Farnesyl acetone	Terpene
163.0389	AP+	AN	5.44	0.73	0.66	$C_9H_6O_3$	Hydroxycoumarine	Benzopyranoid
147.0441	AP+	AN	5.47	0.96	1.57	$C_9H_6O_2$	Coumarine	Benzopyranoid
269.0454	ES-	50	9.46	1.00	1.00	$C_{15}H_{11}O_5$	Pelargonidin	Flavonoid
181.1334	AP+	AN	5.33	1.00	1.00	$C_{10}H_{16}N_2O$	Smipine	Alkaloid
123.1168	AP+	AN	3.63	1.00	1.00	C_9H_{14}	Santene	Terpene
195.0651	AP+	50	3.91	1.89	1.00	$C_{10}H_{10}O_4$	Ferulate	P. Acid
225.0757	AP+	50	3.82	1.27	1.00	$C_{11}H_{12}O_5$	Sinapate	P. Acid
115.0502	AP+	AN	4.76	1.00	1.00	$C_4H_6N_2O_2$	3-cyanoalanine	A. acid
171.0288	AP+	AN	2.06	1.05	0.71	$C_7H_6O_5$	Gallic acid	P. Acid
143.0702	AP+	50	2.52	1.00	1.00	$C_7H_{10}O_3$	Furaneol (R=CH3)	Furan
129.0546	ES+	AN	4.50	1.00	1.00	$C_6H_8O_3$	Furaneol (R=H)	Furan
207.2107	AP+	AN	3.11	1.00	1.00	$C_{15}H_{26}$	4-Muurolene	Terpene
223.2057	AP+	AN	3.25	0.9	1.00	$C_{15}H_{26}O$	Nerolidol	Terpene
151.1481	AP+	AN	3.95	1.00	1.00	$C_{11}H_{18}$	Dimethyl -1,3,7- nonatriene	Terpene
137.1324	AP+	AN	4.63	1.00	0.78	$C_{10}H_{16}$	Pinene	Terpene
171.1379	AP+	AN	2.24	1.55	1.21	$C_{10}H_{18}O_2$	Dec-2-enoic acid	F. Acid
143.1006	AP+	AN	3.04	1.00	1.00	$C_8H_{14}O_2$	Oct-2-enoic acid	F. Acid
187.1692	AP+	AN	3.17	1.00	1.00	$C_{11}H_{22}O_2$	Undecanoic acid	F. Acid
165.0909	AP+	AN	2.96	1.80	0.91	$C_{10}H_{12}O_2$	2-phenylethyl acetate	Ester
151.0754	AP+	AN	3.27	1.00	1.00	$C_9H_{10}O_2$	2-methoxy-4-vinylphenol	Ester
167.0703	AP+	AN	3.32	1.00	1.00	$C_9H_{10}O_3$	Ethyl salicylate	Ester
101.0597	AP+	AN	3.30	1.00	1.00	$C_5H_8O_2$	2-methylbut-2-enoic acid	Acid
109.1011	AP+	AN	3.52	0.90	0.95	C_8H_{12}	Methyl-heptatriene	F. Acid
449.1079	ES+	50	2.14	1.00	1,00	$C_{21}H_{21}O_{11}$	Cyanidin glucoside	Flavonoid
111.0440	AP+	50	2.00	1.87	1.10	$C_6H_6O_2$	Catechol	Phenol
311.1127	ES+	AN	2.05	1.00	1.00	$C_{15}H_{18}O_7$	Cinnamate glucose	Phenol
209.0664	ES-	50	2.13	1.00	1.00	$C_7H_{14}O_7$	Sedoheptulose	Sugar
207.0661	AP-	50	2.38	1.00	1.00	$C_{11}H_{12}O_4$	Sinapaldehyde	Phenolic
179.0703	AP+	AN	2.29	0.94	1.22	$C_{10}H_{10}O_3$	Coniferaldehyde	Phenolic
154.0862	AP+	AN	2.54	1.00	1.00	$C_8H_{11}NO_2$	Dopamine	Alkaloid
143.0349	ES-	50	2.20	1.76	1.47	$C_6H_8O_4$	2-Methyleneglutarate	Acid

[a]Metabolites showing: early expression (group 1, profiles a, b, and c), intermediate expression (group 2, profiles d, e, and f), late expression (group 3, profiles g, h and i) during strawberry fruit development (see also Fig. 4A).

CG, chemical group; EF, empirical formula; OM, observed mass; IM, ionization mode; EX, extraction; P. Acid, phenolic acid; A. Acid, amino acid; A. Acid D., amino acid derivative; O. Acid, organic acid; F. Acid, fatty acid; AP+/-, APCI positive or negative; ES+/-, ESI positive or negative; 50- methanol/0.1% formic acid (50%/50%), AN- 100% acetonitrile. Examples of selected metabolic profiles that are significantly changed in the red (R), turning (T) and white (W) stages compared to the green (G) stage. Positive ratios from 2 to 5, 5 to 10 and >10 are shaded in yellow ( ), red ( ) and dark red ( ) respectively while negative ratios from 2 to 5, 5 to 10 folds and >10 are shaded in light green ( ), darker green ( ) and olive green ( ), respectively. Metabolites identified in the table were either a single hit in the natural product database search or were reported earlier to be produced in strawberry and other fruit or identified by us in strawberry.

CHAPTER 8 - Table 2 (continued)