Developing pineapple fruit has a small transcriptome dominated by metallothionein

Richard Moyle1, David J. Fairbairn1, Jonni Ripi1, Mark Crowe2 and Jose R. Botella1,*

1 Plant Genetic Engineering Laboratory, Department of Botany, University of Queensland, QLD 4072, Australia
2 Institute of Molecular Biosciences, University of Queensland, Australia

Received 5 April 2004; Accepted 20 August 2004

Abstract

In a first step toward understanding the molecular basis of pineapple fruit development, a sequencing project was initiated to survey a range of expressed sequences from green unripe and yellow ripe fruit tissue. A highly abundant metallothionein transcript was identified during library construction, and was estimated to account for up to 50% of all EST library clones. Library clones with metallothionein subtracted were sequenced, and 408 unripe green and 1140 ripe yellow edited EST clone sequences were retrieved. Clone redundancy was high, with the combined 1548 clone sequences clustering into just 634 contigs comprising 191 consensus sequences and 443 singletons. Half of the EST clone sequences clustered within 13.5% and 9.3% of contigs from green unripe and yellow ripe libraries, respectively, indicating that a small subset of genes dominate the majority of the transcriptome. Furthermore, sequence cluster analysis, northern analysis, and functional classification revealed major differences between genes expressed in the unripe green and ripe yellow fruit tissues. Abundant genes identified from the green fruit include a fruit bromelain and a bromelain inhibitor. Abundant genes identified in the yellow fruit library include a MADS box gene, and several genes normally associated with protein synthesis, including homologues of ribosomal L10 and the translation factors SUI1 and eIF5A. Both the green unripe and yellow ripe libraries contained high proportions of clones associated with oxidative stress responses and the detoxification of free radicals.

Keywords: Ananas comosus, EST, metallothionein, non-climacteric fruit, pineapple, fruit ripening.

Introduction

Pineapple (Ananas comosus L. Merril) fruit cropping and processing are important horticultural industries in many countries with tropical climates. In terms of worldwide production, pineapple is currently the third most important tropical fruit after bananas and mangoes. Pineapple is cultivated mainly for fresh or canned fruit and juice, but is also the only source of bromelain, a complex proteolytic enzyme used in the pharmaceutical market and as a meat-tenderizing agent.

Pineapples are diploid (2n=50) perennial monocotyledonous plants with a terminal inflorescence consisting of 50–200 individual hermaphrodite flowers that cluster together. As individual fruits develop from the flowers, they join together forming a cone shaped, compound, juicy, fleshy multiple fruit of approximately 30 cm or more in height. Pineapple maturity is evaluated by the fruit eye flatness, the extent of skin yellowing and by aroma. It takes 3–5 months for pineapple fruit to reach maturity and pineapple fruit quality is highest if the fruit matures on the plant. Pineapples harvested prematurely do not continue to ripen or sweeten, as there are no starch reserves in the fruit to be converted to sugar.

In general, fleshy fruits are divided in two large groups, climacteric and non-climacteric, based upon the presence or absence of an autocatalytic ethylene burst during ripening. Pineapple is an example of a non-climacteric fruit. It does not have an autocatalytic ethylene burst during ripening, and exogenous application of ethylene does not rapidly accelerate fruit ripening. In climacteric fruits, ethylene plays a crucial role in the control of the ripening process, by regulating the transcription of a large number of genes. The ethylene biosynthetic pathway is well established (Yang and Hoffman, 1984) and the expression patterns of the main ethylene biosynthetic genes, 1-aminocyclopropane-1-carboxylic
acid (ACC) synthase (ACS) and ACC oxidase (ACO), have been well characterized (Giovannoni, 2001). Conversely, almost nothing is known of ripening processes in non-climacteric fruit, although ripening-related genes from non-climacteric fruits have been reported with grape, *Citrus unshiu*, and strawberry among those most studied. Davies and Robinson (2000) undertook the isolation of 17 ripening-associated cDNAs by differential screening of a shiraz grape library. An EST sequencing project on grape buds has also been reported, but the library clones were isolated to investigate gene expression leading up to bud release from dormancy (Pacey-Miller et al., 2003). A small-scale sequencing project from *Citrus unshiu* resulted in the isolation of 297 EST sequences (Moriguchi et al., 1998). In addition, an EST sequencing project has been reported from strawberry, resulting in the isolation of 1100 clone sequences (Aharoni et al., 2002; Aharoni and O’Connell, 2002). Microarray analysis of the strawberry clones identified 112 genes that were ripening-regulated (Aharoni and O’Connell, 2002). A number of previously identified genes involved in cell wall softening and pigmentation were differentially expressed during strawberry fruit ripening. Oxidative stress-related genes were also observed as up-regulated in ripening strawberry fruits, and it has been suggested that strawberry fruits may contain a transcriptional programme responsive to oxidative stress induced during ripening (Aharoni and O’Connell, 2002). Despite these efforts to isolate ripening-related clone sequences, the relatively small number of genes available for study has hampered the identification of genes controlling the ripening processes in non-climacteric fruits.

The development and automation of high-throughput sequencing technology provides an opportunity to study genomes at both the DNA and RNA level. Genome-sequencing projects can lead to the identification of the complete catalogue of genes within an organism, independent of gene expression levels. Alternatively, large-scale end-sequencing of cDNA library clones generates expressed sequence tags (ESTs) that can be used to obtain a snap-shot of gene expression in a particular tissue and stage of development at a fraction of the cost of genome sequencing (Adams et al., 1993). EST sequencing has proved a popular and cost-effective method of isolating vast collections of coding sequences from a plethora of species. As at February 2004, the GenBank dbEST (database of expressed sequence tags) contained 20 039 613 EST sequences from 611 organisms (Boguski et al., 1993). Substantial numbers of EST sequences have been generated from a number of plants, with plant species comprising 12 of the 30 most abundant EST collections. In addition, comparing EST clones generated from different tissues at different stages of development provides a method for comparing relative gene expression and identifying subsets of genes and corresponding functions that probably play vital roles within specific tissue types and stages of development.

Despite the economic importance of pineapple as a crop, surprisingly little research has been undertaken to understand the molecular basis of pineapple fruit development. A search of the GenBank nucleotide sequence database (Benson et al., 2003) yielded only 51 sequence entries for *Ananas comosus* (as of February 2004), the majority being isolated from root and vegetative tissues. Almost half of these sequence entries resulted from a study reporting 24 genes isolated from pineapple roots by differential display (Neuteboom et al., 2002). This study reports the first EST sequencing project from pineapple, resulting in the isolation of 1548 fruit EST clone sequences. As far as is known, this is the first report of EST sequencing on any pineapple tissue or indeed on any bromeliad. The isolation of a range of pineapple gene sequences from green unripe and yellow ripe fruit tissue is a first step toward building a valuable resource that will help in understanding the molecular basis of fruit development and will aid pineapple improvement programmes.

**Materials and methods**

**Plant material**

Commercial field-grown pineapples (*A. comosus* L. cv. Smooth Cayenne) were harvested early in the morning and graded according to ripening appearance. Selected pineapple fruits were separated into top, middle, and bottom thirds, with the skin and pith tissues removed. The remaining fruit flesh was snap frozen in liquid nitrogen, pulverized, and stored at −80°C.

**cDNA library construction and sequencing**

PolyA RNA was isolated from the middle third of pineapple 4 (Fig. 1), a mature size but unripe green pineapple, and pineapple 9 (Fig. 1), a mature ripe yellow pineapple, using the oligo dT dynabead kit (Dynal). PolyA RNA was transcribed into cDNA and subsequently amplified using the SMART cDNA synthesis kit (Clontech Laboratories, Palo Alto, CA). SMART cDNA synthesis utilizes a mechanism to enhance the proportion of cDNA containing full-length coding sequences, allows for PCR amplification of the synthesized cDNA, and introduces *Sfi* restriction enzyme sites for directional cloning. The resulting cDNA was size fractionated by electrophoresis through 0.8% agarose gels. cDNA fragments running higher than approximately 700 bp were recovered and purified using the GeneClean II kit (Bio101). The resulting cDNA pools were digested with *Sfi* restriction enzyme and directionally cloned into pBS-*Sfi*. pBS-*Sfi* was engineered by inserting a linker containing *Sfi* directional cloning sites between the *EcoRI* and *XhoI* restriction enzyme sites of pBLUESCRIPT (Stratagene). Transformed colonies were picked at

![Fig. 1. Commercial field-grown pineapples harvested and sorted 1–9 according to visible ripening characteristics.](image-url)
random and placed, using a grid pattern, on duplicate 25 cm² LB ampicillin plates, one of which was subsequently screened at high stringency (0.1× SSC/0.1% SDS at 65 °C) with a radiolabelled metallothionein clone Ac180 using a colony lift procedure (Sambrook et al., 1989). Colonies not hybridizing to the metallothionein Ac180 probe were picked from the duplicate plate for growth in 96 format deep well plates. A robotic platform (Biomek 9600) was used to replicate the 96 well clone plates into multiple glycerol stock bank collections. Cultures were grown from glycerol stock plates and 5’ end sequencing was conducted on purified plasmid preparations.

**Bioinformatics**

Raw sequences were manually edited for sequence quality, trimmed of plasmid contaminant and polyA tail in the sequence viewer program Chromas v2.13 (Technelysium). Clones yielding less than 150 bp of insert sequence were eliminated from the data set. All EST sequences were submitted into the GenBank dbEST. Edited sequences were compiled into contigs or singletons using SeqMan sequence assembly software (DNASTAR Inc. Madison, USA), and key parameters of minimum 90% match over at least 45 bp overlap. Comparison analysis of the contig consensus sequences was conducted with the advanced basic alignment search tool, BLASTX server (Altschul et al., 1990) and the National Center for Biotechnological Information (www.ncbi.nih.gov) non-redundant protein database and resulting information parsed from nearest neighbour hits. Contig consensus sequences returning hits better than the 10E⁻²⁰ E-value cutoff were BLASTX searched against MATDB (mips.gsf.de/proj/thal/db/index.html) to retrieve putative functional classifications. Functional classes and subclasses retrieved were subjected to manual curation.

**Northern analysis**

Total RNA was extracted from meristem, young leaves (shorter than 3 cm in length), mature-sized green leaves, the middle third of pineapples 1–5 and 7–9 (Fig. 1) using a pineapple RNA extraction protocol (Cazzonelli et al., 1998). Total RNA was isolated from root tissue using TRIzol (Invitrogen). Five microgram aliquots of total RNA from pineapple 1–5, 7–9, meristem, young leaf, old leaf, and roots were size fractionated through two 1% agarose gels, and transferred to HyBond N+ membranes (Amersham) by capillary transfer (Sambrook et al., 1989). The duplicate membranes were cross-linked by UV and prehybridized in Church and Gilbert buffer (Church and Gilbert, 1984) at 50 °C. Radiolabelled probes were synthesized using the Strip EZ DNA kit (Ambion) and hybridized to membranes at 55–65 °C. Membranes were progressively washed to an high stringency of 0.1% SSC/0.1% SDS at 60 °C prior to exposure in phosphoimager cassettes. Membranes were stripped and reprobed according to the Strip EZ DNA kit instructions (Ambion).

**Real-time PCR**

For real-time quantitative PCR experiments (RT-qPCR), 1 µg of total RNA was reverse transcribed using Superscript III (Invitrogen) following the manufacturer’s instructions. The resulting cDNA was subsequently taken up in a volume of 250 µl and the fruit and meristem samples further diluted 100-fold. Pineapple metallothionein Ac180 and β-actin target genes were amplified using gene-specific primers designed from the coding sequence, and in the case of Ac180 over an RNA splice junction, using the Primer Express 1.5 software (Applied Biosystems, Foster City, CA). The primer pairs for Ac180 were 5’-TTC TTGGCTCCTTA TTGATGCT GTCAT-3’ and 5’-AGCCGT TTCCCTTCCTCAACA-3’ and for β-actin 5’-ATGG AAGCTGCG GTTATCCA-3’ and 5’-CCACC ACTGAGCAGT GTGTT-3’. Both primers pairs amplified 101 bp amplicons with similar melting temperatures of 81 °C. RT-qPCR was carried out using the ABI PRISM 7700 sequence detector and SYBR Green Master mix (Applied Biosystems) using primers at a final concentration of 0.1 µM each and 2.5 µl (the equivalent of 10 or 0.1 ng total RNA) of cDNA as template. PCR-cycling conditions comprised an initial polymerase activation step at 95 °C for 10 min, followed by 45 cycles at 95 °C for 15 s and 60 °C for 1 min. Real-time DNA amplification was monitored and analysed using the Sequence Detector 1.7 program (Applied Biosystems). Threshold cycle data were exported to EXCEL and the quantification of Ac180 gene expression relative to the β-actin reference gene was calculated using the Pfaffl equation (Pfaffl, 2001). This equation takes into account differing primer efficiencies. Cycle threshold values were corrected for genomic DNA contamination in the RNA sample. The young leaf sample was used as a calibrator to allow comparisons across samples to be made. RT-PCR experiments were repeated on three separate occasions in duplicate.

**Results and discussion**

**cDNA library construction**

PCR amplification of both unripe green and ripe yellow fruit cDNA and subsequent fractionation through an agarose gel revealed an intense band within the green and yellow fruit cDNA pools. Subsequent sequence analysis on a test plate of 96 randomly picked library colonies resulted in the identification of 40 clones as a member of the metallothionein gene family. Subsequent colonies were pre-screened with radiolabelled Ac180 to reduce redundancy due to the abundant metallothionein clone. One problem traditionally associated with EST library construction is the overabundance of short and truncated EST fragments due to incomplete reverse-transcription and the ligation bias toward small inserts. This can result in a disproportionately high number of false undiscovered ‘novel’ sequences due to insufficient coding sequence to establish identity accurately by homology to existing coding sequences. To minimize the overabundance of very short inserts and enhance the proportion of full-length and large EST fragments in the libraries, the amplified cDNA were size-fractionated and those running at less than approximately 700 bp were discarded. Although the resulting cDNA pools still contain fragments of less the 700 bp, it is possible that this size-fractionation strategy may subsequently result in the under-representation of very short genes in the cDNA libraries. Restriction mapping of 24 randomly picked clones revealed an estimated average EST insert size of approximately 1.2 kb for each library. A total of 480 unripe green pineapple and 1536 ripe yellow pineapple clones were subsequently picked for sequencing.

**Sequence analysis**

Raw sequence quality was consistently high, with over 80% of clones returning an average Phred20 score of more than 700 bp. After editing raw sequences for quality and discarding insert sequences of less than 150 bp, a total of 408 green unripe and 1140 yellow ripe pineapple EST sequences were retrieved at an average read length of 785 bp.
All edited EST sequences have been submitted to the GenBank dbEST.

The large read lengths obtained compare favourably with other recently reported EST sequence projects. For example, Bhalerao et al. (2003) report average read lengths of 349 bp and 355 bp from autumn leaf and young leaf libraries, while Ronning et al. (2003) report an overall average read length of 525 bp from various potato EST libraries. Maximum read lengths are desirable, as short truncated EST clones containing 3' untranslated sequence with little or no coding sequence cannot be accurately identified by homology to protein coding sequences in public databases. The length of the sequences generated also impacts on EST cluster analysis, as the longer the EST sequences are, the more likely that multiple EST clones encoded from a single gene will overlap during contig assembly.

Cluster analysis reveals high redundancy within the pineapple fruit EST libraries

All edited EST sequences were clustered using SeqMan sequence assembly software (DNASTAR). The 1548 sequences clustered into 634 contigs. The unripe green and ripe yellow sequences clustered with 62% and 54% redundancy, respectively. Further analysis of all contigs containing two or more green and/or yellow fruit clones revealed that 1105 ESTs (71% of all sequences) cluster within the 191 contigs, whereas just 443 ESTs (29%) represent singletons. To investigate the high level of redundancy, contigs were listed by clone abundance—from contigs containing the largest number of EST clones to those containing a single EST clone from either unripe green or ripe yellow libraries. The resulting data were converted to percentages to enable a direct comparison between the green unripe and yellow ripe cumulative abundance of EST clones (Fig. 2). The results reveal 50% of EST clone sequences cluster within just 13.5% and 9.3% of contigs from green unripe and yellow ripe libraries, respectively. Thus a small subset of genes appears to dominate the majority of the transcriptome in both unripe green and ripe yellow fruit. However, as library construction involved an amplification step, the possibility that amplification bias may have contributed to the high redundancies cannot be discounted.

Bioinformatics

Contig consensus sequences were exported in FASTA format and batch BLASTX searches performed against the GenBank non-redundant database. A stringent BlastX E-value cut-off of 10E-20 was chosen to ensure that the annotations were based only on genes with a high degree of similarity to this study's cDNA clones (Hu et al., 2003; Whitfield et al., 2003). A semi-automated process of parsing BLAST hits and manually curating the putative annotations resulted in a spreadsheet of information including cloneID, contig number, number of clones in each contig, nearest BLASTX neighbour, accession number of match, length of match, percentage similarity, and putative annotation (supplementary data at JXB online).

Approximately 17% of all sequences did not have significant homology to coding sequences in the GenBank non-redundant database and therefore could not be annotated by similarity. This 'undiscovered' sequence subset would probably contain some novel coding sequences that have not previously been discovered. However, it is unrealistic to suggest that all of the undiscovered EST subset encode for novel proteins with no significant homology to those in public sequence databases. Instead, many clones falling under the undiscovered sequence subset are likely to contain insufficient coding sequence to accurately assign an annotation based on homology. It is also possible that some of these unidentified sequences may not encode proteins, but instead may function as RNA molecules. Furthermore, some of the undiscovered sequence ESTs may contain short stretches of protein coding sequence homologous to proteins in the non-redundant database but, due to a small open reading frame or short length of overlapping coding sequence, do not produce an expected value better then this cutoff limit of 10E-20.

The majority (83%) of EST clones were putatively annotated by similarity to coding sequences in the GenBank non-redundant database with over 55% of all EST clones sharing 80% or higher similarity with existing sequences. Approximately 17% of clones had similarity of between 60–80%, and less than 5% shared lower than 60% similarity to known coding sequences. Contig consensus sequences returning a valid BLASTX hit from GenBank were individually searched against the MIPS MATDB database (Schoof et al., 2002). Manual curation of MIPS MATDB functional classification resulted in the assignment of major functional classes and subclasses.

![Cumulative abundance of EST Clone Sequences](image-url)
to 1184 EST clones generated from the green unripe and yellow ripe pineapple libraries.

**Metallothionein clone Ac180 is very highly expressed in pineapple fruits**

Fractionating an aliquot of the unripe green and ripe yellow pineapple cDNA through an agarose gel revealed the presence of a very highly abundant transcript of approximately 500–600 bp in length. Despite gel purification to enrich for cDNA greater than 700 bp, sequence analysis of an initial test plate of 96 clones picked resulted in the identification of 40 clones as a member of the metallothionein gene family, subsequently named Ac180. Excluding the polyA tail, the metallothionein EST clones were approximately 500 bp in length, matching the estimated size of the abundant transcript in the amplified fruit cDNA pools. When subsequent library colonies were screened using radiolabelled Ac180 as a probe, up to 50% of the fruit library colonies screened hybridized strongly. Reports of highly abundant clones isolated in plant EST sequencing projects include 14% of EST sequences from a poplar young leaf library encoding the small subunit of Rubisco (Bhalerao et al., 2003) and 20% of EST sequences from a poplar cambium transition to dormancy library encoding a major storage protein (Schrader et al., 2004). Furthermore, a small-scale EST sequencing project from a *Citrus unshiu* fruit library identified 20% of clone sequences as encoding for metallothionein (Moriguchi et al., 1998). However, as far as is known, there have been no reports of a single gene product dominating a transcriptome to the extent that the metallothionein clone Ac180 appears to dominate the unripe green and ripe yellow pineapple fruit tissue. However, as the libraries were constructed from PCR amplified cDNA, the possibility that the abundance of the metallothionein clone Ac180 is exaggerated due to amplification bias cannot be discounted.

The metallothionein Ac180 expression was further analysed by quantitative real-time PCR using first strand cDNA synthesized from the middle third of the nine pineapple harvests (Fig. 1). β-actin (Ac42) was used as a normalization control (Fig. 3). The expression of the metallothionein Ac180 gene increased during fruit development and was highest in mature fruit. Expression was highest in fruit tissue, but high levels relative to β-actin were also observed in root and leaf tissue, particularly in the older leaf tissue. Metallothionein Ac180 expression was calculated to be more than 50-fold higher than β-actin (Ac42) in the ripening fruit samples 7–9, confirming metallothionein transcripts are highly abundant in ripe pineapple fruit.

A second metallothionein gene, Ac181, was also discovered in the yellow fruit library with 17 copies isolated. No copies of Ac181 were isolated from the green fruit library. The two metallothionein sequences share approximately 80% identity at the amino acid and nucleotide level.

Although Ac180 appears to be considerably more abundant than Ac181, Northern analysis revealed both genes exhibit a similar expression pattern across the range of ripening pineapple fruits sampled, and are similarly up-regulated in older leaves (Fig. 4).

Metallothioneins are small cysteine-rich proteins required for heavy metal tolerance in animals and fungi. In plants, metallothionein genes are up-regulated in response to heavy metal stress (van Hoof et al., 2001), natural and induced leaf senescence (Buchanan-Wollaston, 1994; Chen et al., 2003), ethylene-induced abscission (Coupe et al., 1995), drought and salt stress (Oztur et al., 2002), light stress (Dunaeva and Adamska, 2001), nematode infection (Potenza et al., 2001), pathogen-induced necrosis (Butt et al., 1998), abscisic acid (Chatthai et al., 1997; Reynolds and Crawford, 1996), wounding (Choi et al., 1996), sucrose starvation and heat shock (Hsieh et al., 1996), and cold stress (Reid and Ross, 1997). Metallothionein genes up-regulated during fruit development have been isolated from climacteric fruits such as banana (Clendennen and May, 1997), apple (Reid and Ross, 1997), and kiwifruit (Ledger and Gardner, 1994) and non-climacteric fruit such as grape (Davies and Robinson, 2000), *Citrus unshiu* (Moriguchi et al., 1998), and strawberry (Nam et al., 1999). Two kiwifruit metallothionein-like clones were differentially expressed during fruit development. pKIWI4 was up-regulated early while pKIWI3 was up-regulated late in fruit development. Similarly, two banana metallothionein-like clones were found to be differentially expressed during fruit ripening, pBAN3-23 expression decreased during fruit ripening while pBAN3-6 expression peaked during a later stage of fruit ripening. For *Citrus unshiu*, 62 metallothionein clones were isolated from the 297 clones analysed from a mature fruit library. A comparison of the expression
Functional classifications were assigned to contig consensus sequences that had an E-value better than 10E-20, using a manually curated system based on MIPS MATDB functional classification. Results obtained from contig consensus sequences were subsequently assigned to the individual EST sequences clustering within each contig. The highly abundant metallothionein gene Ac180 was excluded from the functional classification analysis.

Functional class distribution indicated that 10% of unripe green and 6% of ripe yellow clones encode proteins falling within the cell rescue, defence, and virulence major class (Fig. 5). This represents an unusually high proportion of clones, although similar percentages of cell rescue, defence, and virulence clones have been found in senescing leaves (Bhalerao et al., 2003). The majority of the clones clustered within the subclasses representing detoxification and stress responses, particularly oxidative stress and the detoxification of reactive oxygen species and free radicals. Indices of oxidative processes have been measured in tomato fruits (Jimenez et al., 2002). Hydrogen peroxide content, lipid peroxidation, and protein oxidation were all found to increase at the breaker stage (fully developed green fruit just prior to turning red). Six EST copies of a (S)-2-hydroxy-acid gene (Ac3388) were identified from the green pineapple fruit. (S)-2-hydroxy-acid oxidase (glycolate oxidase) catalyses a reaction resulting in the production of H2O2. It is a known source of peroxide produced during photorespiration and is enhanced during abiotic stress responses (Mittler, 2002). As (S)-2-hydroxy-acid oxidase appears to be abundant in the green unripe fruit, it is speculated that it is a significant source of peroxide, contributing to a state of oxidative stress within the fruit. Peroxiredoxin and glutathione transferase are two genes involved in the detoxification of H2O2. Peroxiredoxin (Ac21) and glutathione transferase (Ac135) were among the most abundant gene sequences from the green fruit library, with six EST copies isolated each (Table 1). Three copies of peroxiredoxin (Ac21) were also isolated from the yellow fruit library. Clones encoding additional glutathione transferase genes were isolated from the green fruit library (three copies of Ac134 and two copies of Ac54), and the yellow fruit library (three copies of Ac55, three copies of Ac2038, and two copies of Ac2248). Glutathione transferases are up-regulated in many plants in response to a range of stress conditions (Conklin and Last, 1995; Marrs and Walbot, 1997; Mittova et al., 2003; Moons, 2003; Ruzsa et al., 1999; Skipsey et al., 1997; Thom et al., 2002) and have been reported as up-regulated during fruit ripening in strawberry (Aharoni et al., 2002; Aharoni and O’Connell, 2002). In plants, the functions of glutathione transferase include detoxification of xenobiotics and the stabilization of flavanoids (Dixon et al., 2002). In addition, stress-inducible glutathione transferases have been shown to protect plants from oxidative injury by functioning as glutathione peroxidases (Cummins et al., 1999; Roxas et al., 1997). It is probable that some glutathione transferase genes are up-regulated by the general oxidative stress caused during a range of stress conditions (Dixon et al., 2002), and that oxidative stress also causes up-regulation of glutathione transferases in fruits. Similarly, plant peroxiredoxins are up-regulated during stress. They function as antioxidant proteins, detoxifying various peroxide substrates (Dietz, 2003;
Hofmann et al., 2002). In addition, two ascorbate peroxidase genes, Ac50 (two copies) and Ac3096 (one copy), one monodehydroascorbate reductase gene (Ac3333), and a glutaredoxin gene (Ac37) were isolated from the yellow fruit library. Ascorbate peroxidase and monodehydroascorbate reductase also function to scavenge $H_2O_2$ (Mittler, 2002) while glutaredoxin is described as an electron donor to peroxiredoxin (Rouhier et al., 2003). Three superoxide dismutase genes, Ac119 (one copy), Ac3200 (one copy) and Ac3219 (three copies) were isolated from the yellow fruit library. One copy of Ac3219 was also isolated from the green fruit library. Superoxide dismutase functions as a scavenger of $O_2^-$ (Alscher et al., 2002). The relative abundance of antioxidant proteins and enzymes of the ROS scavenging pathways in pineapple fruits supports the view that ripening is an oxidative phenomenon requiring a balance between the production of ROS and their removal by antioxidant systems.

**The green unripe and yellow ripe pineapple fruit express different subsets of abundant genes**

Unripe green and ripe yellow fruit clones were sorted by contig abundance for direct comparison. The fruit libraries shared few abundant genes in common, with most abundant
gene species being preferentially isolated from either the unripe green or ripe yellow fruit (Tables 1, 2). Abundant clones preferentially isolated from green fruit included a bromelain inhibitor (Ac124) and a fruit bromelain (Ac122). The bromelain inhibitor was the most abundant clone type from the green fruit library with 52 clones isolated. Northern analysis confirmed these two genes are down-regulated during fruit ripening (Fig. 4). Why bromelain and a bromelain inhibitor are simultaneously expressed in developing fruits is unclear.

Cinnamyl alcohol dehydrogenase (CAD) (Ac176) clones were isolated from both the green and yellow fruit libraries, but were more abundant in the green than the yellow fruit library. A CAD gene was reported to be up-regulated during fruit ripening in strawberry (Aharoni et al., 2002; Blanco-Portales et al., 2002). CAD up-regulation in strawberry fruits has been suggested to play a role in the interconversion of aldehydes and alcohols implicated in flavour as well as the lignification of vascular elements (Aharoni et al., 2002).

Eight copies of a putative PHD transcription factor (Ac3400) were isolated from the green fruit library, while none were identified among the yellow fruit library clones. It is possible that this putative transcription factor directs changes in gene expression during the course of fruit ripening. Conversely, a MADS box transcription factor homologue (Ac146) was the most abundant clone type isolated from the subtracted yellow fruit library, but was not isolated from the green fruit library. MADS-box genes encode for a large family of transcription factor proteins involved in regulating various aspects of plant development, particularly flowering and organ or meristem identity. Analysis of the Arabidopsis genome revealed 107 genes encode MADS box proteins, although the function of most remain unknown (Parenicova et al., 2003). Northern analysis of developing pineapple fruit confirmed that Ac146 expression increases during ripening and is most abundant in yellow ripe fruit tissue (Fig. 4). No expression was detected in meristem, leaf or root tissue. It is somewhat unusual for a transcription factor to be among the most highly expressed genes, and, therefore, it is hypothesized that the differential abundance of this MADS box gene probably plays a significant role in regulating differential gene expression during fruit ripening.

Interestingly, a protein translation factor SUI1, a eukaryotic translation initiation factor 5A isoform (eIF 5A), a 60S ribosomal protein L10, and a putative nascent polypeptide-associated complex alpha chain were all abundant among the yellow fruit library clones yet absent in green fruit library clones. Why these translation-related genes are strongly up-regulated in the ripe yellow fruit is not immediately obvious as one would expect protein synthesis genes to be relatively evenly expressed throughout the plant. However, recent research in non-plant systems has determined that L10, SUI1, and eRF5A perform alternative functions in addition to protein synthesis. The pineapple 60S ribosomal L10 gene is a homologue of the human QM ribosomal protein (Dowdy et al., 1991) and the Jun interaction factor Jif-1 from chicken (Monteclaro and Vogt, 1993). Both QM ribosomal protein and Jif-1 function as tumour suppressor genes (Dowdy et al., 1991; Monteclaro and Vogt, 1993), and Jif-1 has been demonstrated specifically to bind to c-Jun, causing the inhibition of a certain AP-1 transcription factor function. SUI1 translation initiation factors have been implicated in the nonsense-mediated decay pathway (NMD) in addition to their role in recognition of the AUG codon during translation initiation (Cui et al., 1999). The NMD pathway is an mRNA surveillance pathway, responsible for recognizing aberrant mRNA transcripts with premature stop codons and accelerating their...
Hypusine modified eIF5A is the active form of the protein and is thought to act as a nucleo-cytoplasmic transporter of mRNA from the nucleus to the cytoplasm, in addition to translation initiation factor activity (Bevec and Hauber, 1997; Rosorius et al., 1999). The eIF5A mediated translocation of mRNA may occur for only certain RNA species, as hypusine-modified eIF5A was found to bind RNA in a sequence-specific manner (Xu and Chen, 2001). Yeast cells, in which DHS has been inactivated are incapable of dividing, prompted the proposal that hypusine-containing eIF-5A facilitates translation of a subset of mRNAs required for cell division. In Arabidopsis, eIF5A expression peaks in rosette leaves at days 14 and 35 after emergence, coincident with the onset of bolting and the later stages of leaf senescence (Wang et al., 2003). In tomato, eIF-5A mRNA show a parallel increase in abundance in senescing flowers, senescing fruit, and osmotically stressed tomato leaves exhibiting programmed cell death (Wang et al., 2001). Northern analysis of the pineapple eIF5A shows an up-regulation of the eIF5A during fruit ripening and in older leaves, consistent with the expression data reported for Arabidopsis and tomato eIF5A (Wang et al., 2001, 2003).

Two highly abundant undiscovered sequences were present in the ripe yellow fruit library, but absent in green fruit library. The undiscovered sequence Ac81 represented by 44 ripe yellow fruit clones appeared to have a very short open reading frame encoding 64 amino acids. Although annotated as an undiscovered sequence, there are two hypothetical genes from rice that share approximately 64% similarity. The short open reading frame of only 64 amino acids resulted in an expected value falling below the $10E^{-20}$ threshold, leading to the annotation as an undiscovered sequence. The undiscovered sequence Ac157 represented by 14 ripe yellow fruit clones is just 283 bp in length and contains no significant open reading frame, the longest starting at the fifth ATG and being just 16 amino acids in length. As 13 of the 14 ESTs for this gene begin at the same base and there were nine variations in the polyA addition base, it is believed that the full-length clone sequence for this transcript has been isolated rather than partial sequence clones containing only a 3' untranslated sequence. Together with the lack of a reasonable open-reading frame, it is hypothesized that Ac157 may function as an RNA molecule rather than encode for a protein.

Abundant genes found in common and in relatively proportionate numbers between the unripe green and ripe yellow libraries included an S-like ribonuclease, mitochondrial ADP/ATP carrier protein, and a putative transcription factor IIB. However, northern analysis revealed that the S-like ribonuclease and mitochondrial ADP/ATP carrier protein genes are up-regulated during ripening, with highest expression in the ripe yellow fruit. S-like ribonucleases have been isolated from many plant species and characterized as being up-regulated during processes including senescence, phosphate starvation, wounding, and drought stress (Liang et al., 2002; Ma and Oliveira, 2000; Salekdeh et al., 2002; Taylor et al., 1993). It has been suggested that S-like ribonucleases are up-regulated during senescence and in response to phosphate starvation to play a role in the remobilization of phosphate.

Comparing clone distribution by functional class reveals major differences between the unripe green and ripe yellow fruit libraries

A number of differences between the unripe green and ripe yellow fruit libraries were observed after clustering clones by assigned functional classification (Fig. 5). The largest difference was in the protein synthesis category, where 10% of ripe yellow fruit clones and 2% of unripe green fruit...
clones clustered. The difference is largely attributable to three abundant gene products, a 60S ribosomal protein L10, an abundant SUI1 translation factor, and isoforms of a eukaryotic translation initiation factor 5A.

Nine per cent of unripe green fruit and 4% of ripe yellow fruit clones fell within the cellular transport functional class. A single abundant gene encoding a mitochondrial ADP/ATP carrier protein was largely responsible for the difference, being twice as abundant in green fruit than in yellow fruit. Conversely, the transport facilitation class represented 2% of ripe yellow fruit clones and 1% of unripe green fruit clones. The difference is largely due to the eight clones isolated from ripe yellow fruit encoding various subunits of H+ transporting vacuolar ATPase, compared with just one clone isolated from unripe green fruit.

The transcription class contained 9% of clones from ripe yellow fruit and 5% of unripe green fruit clones. The percentage difference can be accounted for by the presence of the abundant MADS box gene AC146 from the ripe yellow library that was not found in the green fruit library. Interestingly, a closer inspection revealed no transcription factor genes were isolated in common from both the unripe green and ripe yellow libraries. The yellow fruit library contained three different MADS box genes, a scarecrow-like protein clone, a bHLH transcription factor, and two different zinc finger proteins. The green unripe library contained seven copies of a putative PHD finger transcription factor protein, an APETALA2 homologue, and a G-box binding factor.

The cellular communication and signal transduction functional class contained approximately 3–4% of clones from both the unripe green and ripe yellow libraries. However, closer inspection revealed major differences in the composition of the subclasses. The green unripe library contained many clones that fell within the plant hormonal regulation subclass whereas the yellow unripe fruit predominantly contained clones involved in Ca2+-mediated and G-protein-mediated signal transduction.

Approximately 3% of clones from each library fell within the energy functional class. However, the composition of energy-related clones differs between each library. The ripe yellow fruit library produced 18 clones related to electron transport and membrane-associated energy conservation that were not present in the unripe green fruit library clones. These clones encode for photosystem II protein K (seven clones), cytochrome b5 (seven clones), and cytochrome b3 reductase (four clones). The green unripe fruit produced six clones encoding a putative (S)-2-hydroxy-acid oxidase assigned to the respiration subclass that were absent from the yellow fruit library clones.

Differences were also observed between various metabolism subclasses. The ripe yellow library contained many clones associated with cysteine biosynthesis that were absent from the unripe green library clones. In addition, 12 clones of the subclass glycine degradation encoding for glycine decarboxylase were isolated from the ripe yellow fruit library, whereas just one clone was isolated from the green fruit library. The nitrogen and sulphur utilization and metabolism subclasses contained 11 clones from the yellow fruit library and only one from the green fruit library.

Approximately 37–40% of the unripe green and ripe yellow fruit sequences analysed shared homology to unclassified MATDB sequences, and as such were assigned as of unknown or not clear-cut function.

Summary

A pineapple EST sequencing project was initiated to study gene expression during fruit ripening and to develop a valuable molecular resource for pineapple biotechnology. Significant differences have been identified in the most abundant genes expressed in the green unripe tissue compared with the yellow ripe fruit tissue. The differential expression patterns of the most abundant EST species isolated were validated by northern analysis of fruit tissue spanning a range of developmental and ripening stages, as well as apical meristem, leaf, and root tissues (Fig. 4). Northern analysis confirmed the fruit bromelain and bromelain inhibitor genes found to be abundant in the green fruit library are down-regulated during fruit ripening. Two metallothionein genes (Ac180 and Ac181), a MADS box gene (Ac146), a protein translation factor SUI1 (Ac75), and a 60S ribosomal protein L10 (Ac8) were all abundant in the yellow fruit library relative to the green fruit library. All were shown to be strongly up-regulated during fruit ripening by northern analysis. With the exception of MADS box protein Ac146, the abundant clones isolated from the yellow fruit library exhibit a similar expression pattern of up-regulation during fruit ripening and up-regulation in older leaves. These results suggest that a suite of genes up-regulated during ripening in fruits is similarly up-regulated during maturation or senescence in leaves, perhaps by common regulatory processes.

The metallothionein clone Ac180 was extremely abundant in the green unripe and yellow ripened libraries. Over 40% of all library colonies hybridized strongly to a radiolabelled Ac180 probe. The metallothionein Ac180 expression level was further analysed by quantitative real-time PCR. Expression of metallothionein Ac180 was calculated at over 50-fold higher than the β-actin control in ripening fruit tissues. Why metallothionein is so strongly expressed in ripening pineapple fruit tissue remains unclear.

The present study provides a valuable resource for the future study of fruit ripening in pineapple, and in non-climacteric fruits in general. A number of genes likely to play vital roles in fruit ripening and senescence have been identified, providing many candidate genes for future analysis. In addition, the isolation of the EST clone set will allow microarrays to be constructed for the large-scale examination of gene expression over a range of fruit development stages. It is hoped to investigate the similarities and differences in genes expressed in this non-climacteric fruit
model compared with climacteric fruit systems. It is also possible that key gene sequences identified could prove useful for future molecular breeding programmes, as markers to aid selection of desirable traits in commercial pineapple lines. The identification of pineapple fruit-specific genes will also facilitate the isolation of promoter sequences conferring tight gene control during fruit ripening.

**Supplementary data**

For supplementary data, please refer to JXB online.

**Acknowledgements**

This work was supported by the Australian Research Council (ARC) and Golden Circle Limited.

**References**


