DOCTOR OF PHILOSOPHY

Global investigations into the metabolic regulation of gene expression during germination and seedling establishment in Arabidopsis thaliana

Allen, Elizabeth

Award date:
2007

Awarding institution:
Bangor University

Global investigations into the metabolic regulation of gene expression during germination and seedling establishment in *Arabidopsis thaliana*

Elizabeth Jane Allen

A thesis submitted to Bangor University in candidature for the degree of Doctor of Philosophy

School of Biological Sciences
Bangor University
Gywnedd, LL57 2UW

October 2007

Summary

Metabolite and gene expression profiles were utilised in order to investigate potential metabolic signals with the ability to affect gene expression in germinating seeds and young seedlings of *Arabidopsis thaliana*. A previously suggested role of acetate in germination was investigated using *acn1*, a previously identified acetate non utilisation mutant. A general repression in metabolism was observed in the mutant. A number of genes involved in seed storage and development were repressed in *acn1* indicating that disruption to acetate metabolism affects aspects of seedling development and metabolism. The *acn1* mutant was used to investigate interactions between acetate and carbohydrate signalling. The expression of sucrose responsive genes was altered in *acn1* suggesting that organic acid signalling mechanisms cross-talk with sugar, stress and developmental signalling pathways. Non-targeted metabolite and gene expression profiles were produced for imbibed seeds and young seedlings. A metabolic switch preceded a re-organisation in gene expression for the metabolites measured in this experiment. Correlation of metabolite and gene expression profiles identified potential regulatory relationships between genes and metabolites during development.

Contents

Figure contents

Chapter 6 Integration of transcript and metabolite data

Table contents

Acknowledgements

I would like to thank the many people who have helped me over the past few years. I am indebted to my supervisors Dr. Mark Hooks and Prof. Deri Tomos for their unfailing encouragement and advice throughout my PhD. Their enthusiastic contribution to scientific discussions has enhanced my enjoyment of my PhD immensely. I would like to thank Dr. Karen Greville for her help in the production of Affymetrix microarrays. My thanks go to Dr. Martine Dieuaide at INRA-Bordeaux for metabolite quantification. I would like to express my gratitude to Dr. Joel Milner at Glasgow University for assistance with quantitative real-time PCR analysis. I am also grateful to Dr. Tim Ebbels at the Universty College of London for his help with data analysis. The members of lab F4 and F21 who have provided a friendly and supportive atmosphere are too numerous to mention but special thanks go to Dr. Karen Roberts, Elaine Murphy, Louise Thurlow, Dr. Aidan Farrell and Dr. Barry Grail.

I am extremely grateful for the continuous support provided by my family. The invaluable contributions of time, finances and opportune distractions from my parents, sister, tadcu and numerous other relations have been greatly appreciated. A very special mention goes to Tom Fryer for, amongst many noteworthy feats, mastering the minutiae of academic reference management in a remarkably short space of time.

My heartfelt thanks go to the founding members of the Women in Science Society. Rhona Bennett, Susan Brittain, Megan Davies, Alice Gray, Guo Wei, Alessa Jaendling, Dr. Cath Kay, Sarah Nicholl and Amanda Powell have at various times and in a variety of inventive ways provided support in times of emotional and scientific unrest. I am also very appreciative of the tolerance displayed by my friends not in the academic community. Sarah Davies, Joanna Carey and Wayne Cooper are among those who have patiently stood by, awaiting my reappearance from the laboratory. Finally I would like to mention the very good people at the Hendricks gin distillery without whom this thesis would almost certainly not have materialised.

Happy is he who gets to know the reasons for things.
Virgil (70-19 BC)

Abbreviations used in the text

2D-SOM	2-dimensional self organising map
A_{260}	absorbance at 260 nm
A_{280}	absorbance at 280 nm
ABA	abscisic acid
Ac	acetate
acn (ACN)	acetate non-utilisation
ADP	adenosine-5'-diphosphate
ATP	adenosine-5'-triphosphate
bHLH	basic helix-loop-helix
BR	brassinosteroid
BSA	bovine serum albumin
cAMP	cyclic-AMP
Cvi	Cape Verde Islands
Col	Columbia
dH_2O	deionised water
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ESI-Q-TOF	ElectroSpray Quadrupole Time Of Flight
FAc	fluoroacetate
g	gravitational force (times gravity)
GA	gibberellic acid
icl (ICL)	isocitrate lyase
IUPAC	International Union of Pure and Applied Chemistry
LB	Luria-Bertani medium
LEA	late embryogenesis abundant
Ler	Landsburg erecta
M	molar (g/l)
MALDI-TOF	Matrix Assisted Laser Desorption/ Ionisation Time Of Flight
mRNA	messenger ribonucleic acid

MS	mass spectrometry
NMR	nuclear magnetic resonance
RNA	ribonucleic acid
PCA	principal component analysis
PCR	polymerase chain reaction
rpm	revolutions per minute
Q-RT-PCR	quantitative real time polymerase chain reaction
RT	reverse transcription
RT-PCR	reverse transcription polymerase chain reaction
SAGE	serial analysis of gene expression
Tris	tris (hydroxymethyl) amino-methane
UV-A	ultraviolet A
UV-B	ultraviolet B
v/v	volume to volume ratio
w/v	weight (mass) to volume ratio

Chapter 1

Introduction

1.1 Seed development and germination

The developmental period encompassing seed development, germination and the establishment of photosynthetically competent seedlings is of great biological and economical importance. The production of a mature seed able to sense and respond appropriately to environmental conditions is essential for the dispersal and continued survival of a plant species. A tightly regulated process occurs to produce a mature seed containing sufficient storage reserves to allow germination and the establishment of a self-sufficient plant to occur. The onset of germination and maintenance of dormant conditions has widespread economic implications in cultivation of crop species and survival of undesirable weed species. Approximately 70% of food supplies are derived from seeds, while a large proportion of the remainder comes from animals fed on seed stock (Bewley and Black, 1994). The high proportion of stored oils in oilseed crops may be exploited through metabolic engineering to produce neutraceutical polyunsaturated fatty acids, biodiesel and alternatives to petrochemical supplies (Thelen and Ohlrogge, 2002). The availability of suitable model plant systems such as the dicotyledonous plant *Arabidopsis thaliana* which is closely related to major oilseed crops provides a means for further investigations into these processes. This thesis aims to investigate potential regulatory mechanisms during the later stages of the transition from heterotrophic to predominantly photoautotrophic metabolism in Arabidopsis seedlings.

1.1.1 Embryogenesis and seed maturation

Plant embryogenesis and maturation is a complex developmental process cumulating in the production of a mature seed, primed for germination. The fertilisation of the egg-cell by the sperm cells commences the diploid phase of the plant life cycle. Fertilisation is a two stage event; one sperm cell fertilises the egg cell to produce the zygote while a second sperm cell unites with two polar nuclei within the embryo sac to initiate differentiation of endosperm tissue (Goldberg et al., 1994; Howell, 1998). Unlike early animal development the majority of plant development occurs post-embryogenesis, consequently plant embryos are comparatively simple in structure. Plant embryos are generally composed of two main systems, the axis and the cotyledons. Although specifics vary according to the species, axis tissue is typically composed of meristem tissue, epicotyl, hypocotyl and the radicle (Howell, 1998).

The process of embryogenesis may be sub-divided into three main phases, each of which is associated with distinct physiological and regulatory events: (1) postfertilisation-proembryo, (2) globular-heart transition and (3) organ expansion and maturation (Lindsey and Topping, 1993; West and Harada, 1993). One of the first, and most important, events in embryogenesis is the establishment of polarity within the embryo which occurs prior to the first cell division in the proembryo stage. The alignment of the embryonic axis with the micropyle axis of the ovule indicates that maternally-derived signalling might regulate this process (Laux and Jüurgens, 1997). Asymmetric cellular cleavage continues in the zygote, resulting in the production of a bilaterally symmetrical heart stage embryo where tissue differentiation into cotyledons, root and shoot apices and provascular tissue has begun (Meinke, 1991; West and Harada, 1993). Following the production of embryonic structures, a period of growth occurs during which time the embryo expands to fill the seed sac (Goldberg et al., 1994). Physiological constraints such as the maternal tissue surrounding the embryo and the density of tissue within the embryo itself have impeded the analysis of embryonic gene expression events. However, the expression of a large number of embryonic genes has been described including glycosylated extracellular proteins thought to be involved in transcriptional regulation, signal

transduction and cutinisation of the embryo surface (Sterk et al., 1991; van Engelen et al., 1993; Thomas, 1993; Girke et al., 2000).

After embryo expansion cell division in the embryo ceases (Raz et al., 2001). The final stages of embryogenesis are involved in the production of a mature seed fully equipped to commence the process of germination. Prior to the establishment of photosynthetic apparatus in seedlings, the seed undergoes an autotrophic growth phase (Bewley, 1997). Consequently, sufficient storage products must be accumulated during embryogenesis to support growth during this period. The two main storage compounds produced in Arabidopsis embryos are the 12S (cruciferin) and 2S (arabin) which are identified in the hypocotyls and cotyledon tissues of mature embryos (Heath et al., 1986). Metabolite profiles produced at varying stages of Arabidopsis seed development identified a decrease in amino acids, sugars and organic acids during seed maturation, suggesting the incorporation of these metabolites into storage compounds (Fait et al., 2006).

Following the production of a mature seed with adequate storage reserves, a program of water-loss is initiated, allowing the seed to enter a state of desiccation. Control mechanisms are also established to prevent premature germination of the seed on the mother plant and to establish a dormant state for the seed so that it may persist in the environment until environmental factors are favourable for germination to occur (Lindsey and Topping, 1993; West and Harada, 1993). Some plant species have been shown to gain photosynthetic capacity during the process of seed maturation. This results in an increase in oxygen content in the seed, restoring the energy status of the tissue (Geigenberger, 2003). The transition to the desiccated seed is associated with an accumulation of distinct sugars, fatty acids and free amino acids suggesting the establishment of a metabolic state which supports the rapid metabolic recovery upon the onset of seed germination (Fait et al., 2006).

The main genetic events associated with the seed maturation phase include the induction of genes encoding seed storage proteins and the late embryogenesis abundant (LEA) proteins. A correlation between drought stress and the abundance of LEA transcripts has suggested a role in dehydration tolerance for LEA proteins especially during seed desiccation (reviewed in Wise and Tunnacliffe, 2004; Tunnacliffe and Wise, 2007). Gene transcription activity

during embryogenesis does not appear to produce mRNA exclusively for the use of the developing seeds; a large number of extant mRNA species have been identified in mature dry seeds which are otherwise metabolically inactive (Dure and Waters, 1965; Ishibashi et al., 1990; Almoguera and Jordano, 1992; Nakabayashi et al., 2005). Transcriptomic analysis of stored mRNA in Arabidopsis seeds identified a number of regulatory mechanisms encoded by the gene transcripts including epigenetic chromatin structure, chromosomal groupings of co-regulated genes and *cis*-acting promoter elements, suggesting a regulatory role for stored mRNA species (Nakabayashi et al., 2005).

1.1.2 Seed dormancy and germination

A program of dormancy is established during the late stages of embryogenesis in order to prevent precocious seed germination, thereby optimising the chances of survival for seedlings. However, due to the termination of the dormancy period in a variety of species-specific mechanisms no clear definition of dormancy has been accepted by the scientific community (Bewley and Black, 1994; Vleeshouwers et al., 1995; Lange, 1996; Bewley, 1997; Koornneef et al., 2002). The definition of seed dormancy as the failure of an intact viable seed to complete germination under favourable conditions proposed by Bewley (1997) is used in this work. The major factors contributing to the successful breakage of dormancy and subsequent seed germination are the constraints of the tissue surrounding the embryo and the growth potential of the embryo itself. Secondary regulatory factors such as after-ripening (the period of seed storage after harvest), light and temperature are also able to influence the timing and rate of germination (Bewley and Black, 1994; Bewley, 1997; Koornneef et al., 2002; Kucera et al, 2005). Dormancy itself may be considered as a characteristic of the seed which defines germination conditions as much as any of the previously identified regulatory factors (Vleeshowers et al., 1995; Thompson, 2000; Fenner and Thompson, 2005). The depth of dormancy imposed upon a seed varies dramatically according to ecotype. For example, dry seeds of Arabidopsis Columbia (Col) and Landsberg erecta (Ler) ecotypes lose dormancy after a

period of a few weeks at 25°C while seeds of the ecotype C24 do not lose dormancy for a period of several months. Strong dormancy is observed in the Cape Verde Islands (Cvi) and Kashmir-2 (Kas2) ecotypes, requiring a period of nearly a year to become non-dormant (Clerkx et al., 2003).

The process of seed germination commences with the uptake of water by the mature dry seed and terminates with the elongation of the embryonic axis, i.e. the protrusion of the radicle through the tissue surrounding the seed embryo, (Bewley and Black, 1994; Bewley, 1997). Water uptake by the quiescent seed occurs in a triphasic manner. An initial rapid intake of water at the onset of germination is followed by a plateau phase until germination has been completed. The advent of post-germinative growth is associated with a further uptake of water (Bewley, 1997). The initial influx of water into the dry seed results in damage to membrane structures and leakage of solutes prior to re-hydration of membranes to a liquid crystalline physical state (Crowe and Crowe, 1992; Bewley, 1997).

The quiescent mature seed contains very little moisture (in the range of 5 to 15%) and demonstrates very low metabolic activity. Upon re-hydration of the quiescent seed, a burst in metabolic activity is observed probably supported by the consumption of the metabolites accumulated during the transition to seed desiccation (Bewley and Black, 1994; Bewley, 1997; Fait et al., 2006). It is assumed that the enzymes and structures required for post-imbibition metabolism are stored in the dry seed, the period of quiescence does not affect their function and that the initial uptake of water is sufficient to allow the resumption of metabolic activity within the seed (Bewley, 1997). Extant mitochondria produced during seed development are able to produce sufficient ATP to support metabolism during seed imbibition (Attucci et al., 1990). As development progresses mitochondria are repaired or new organelles synthesised, dependent on the type of storage reserves present in the seed. Seeds with starch as the main storage compound repair existing mitochondria while seeds with reserves composed primarily of oil produce new mitochondria (Morohashi and Bewley, 1980; Morohashi 1986).

The quiescent seed contains a population of stored mRNA which survive the period of desiccation and are transcribed at the onset of imbibition (Raghavan,

2000). Translational apparatus is also stored in the dry seed in the form of ribosomes, the numbers of which decrease as polysomal protein synthesis complexes are formed. *De novo* synthesis of ribosomes is observed within hours of the onset of imbibition for the continued production of proteins (Dommes and Van der Walle, 1990). The extant mRNAs are formed during the later stages of embryogenesis to synthesise proteins required during germination. A large proportion of these mRNA species has been shown not to be specifically associated with germination and are generally required for the production of components of normal cellular metabolism (Bewley and Marcus, 1990; Fu et al., 2005).

The burst of metabolic activity observed immediately post-imbibition is associated with the recovery of respiratory activity by the germinating seed and may be divided into four phases commencing with the rapid uptake of oxygen in the first few minutes following imbibition. Oxygen uptake is subsequently stabilised or increases slowly during the second phase. The penetration of the seed coat by the radicle is correlated with a second sharp increase in oxygen uptake in the tissue of cotyledons and the radicle (phase three). The final phase involves a decrease in respiration in cotyledon tissue which is induced by the onset of senescence (Botha et al., 1992). Gaseous diffusion is impeded in seeds by the presence of the testa and the density of material within the seed, resulting in the production of excess pyruvate and a period of anoxia during the second phase of oxygen consumption (Crawford, 1977; Bewley, 1997).

1.1.3 Radicle extension and the completion of germination

The penetration of the tissues surrounding the embryo by the radicle is generally accepted as the visible sign that the germination process has been completed (Bewley, 1997). The mature seeds of most angiosperms contain endospermic tissue, which surrounds the embryo and provides a source of nutrients. In the seeds of monocotyledonous plants, the endosperm may serve as the principal storage tissue while in dicotyledonous seeds the endosperm layer might be completely depleted during seed maturation (Howell, 1998). The seeds of Arabidopsis retain a single cell layer of endospermic tissue (Pritchard et al.,

2002; Liu et al., 2005). The maternally derived seed coat (testa) tissue surrounds the endospermic layer. The mechanism of radicle penetration of the endosperm and/ or testa tissue surrounding the embryo therefore differs according to species type. Seeds with no endosperm tissue undergo a one-step germination process which involves the concurrent rupture of the testa and elongation of the radicle. The exogenous application of the plant hormone abscisic acid (ABA) does not inhibit testa rupture but has been shown to inhibit radicle elongation (Schopfer & Plachy, 1984; Finch-Savage and Leubner-Metzger 2006). A two-step germination process involving the sequential rupture of the endosperm and testa has been described in a number of endospermic seeds including *Trollius* species (Hepher and Roberts 1985), *Chenopodium* species (Karssen 1976), and *Nicotiana* and *Petunia* species (Krock et al. 2002; Leubner-Metzger et al. 1995; Petruzzelli et al. 2003). This two-step germination process is also observed in seeds with a thin layer of endosperm such as Arabidopsis (single endosperm cell layer) and *Lepidium sativum* (1-2 cell layers of endosperm) (Liu et al., 2005; Muller et al., 2006). An inhibitory role for abscisic acid has been identified in the rupture of endosperm, but not testa tissue in these seeds (Müller et al., 2006). An antagonistic interaction between ABA and the phytohormone giberellic acid (GA) is suggested by the counteractive effects of GA on this inhibition (Kucera et al., 2005).

Radicle growth occurs prior to the penetration of the structures surrounding the embryo. Three main mechanisms have been proposed to drive this process: the relaxation of cell wall structures, weakening of the endosperm and/ or the testa or by the accumulation of solutes and resultant increased turgor pressure of the cells (Welbaum and Bradford, 1990; Bewley, 1997). It has been hypothesised that the mobilisation of storage reserves in radicle cells might result in the accumulation of solutes. The resultant decrease in the water potential of the radicle cells might increase the water uptake of the cells and drive the cell extension required for radicle emergence. However there is very little solid evidence for this occurrence (Bewley, 1997). In *Brassica napus* seeds with little endosperm tissue no significant accumulation of solute or change in hydraulic conductivity was detected prior to radicle emergence; the main contributing factor appeared to be the increased extensibility of cell walls and a decrease in turgor yield threshold (Schopfer and Plachy, 1985). Welbaum and Bradford

(1990) demonstrated that extension of the radicle during the germination of muskmelon seeds is predominantly driven by weakening of the endosperm; no accumulation of solutes or turgor pressure was observed but the removal of the endosperm and testa allowed the water content of the radicle to increase by 50%.

1.1.4 Postgerminative growth

Post-germinative growth is primarily supported by stored reserves of carbohydrate, protein or lipid accumulated in the seed during embryogenesis (Bewley and Black, 1994). In oil seeds such as Arabidopsis, as much as 40% of the seed storage reserve content is composed of oil stored in the cotyledons. During germination and seedling establishment fatty acids are released from triacylglycerol stores through the action of lipase and converted to carbohydrate through the sequential activity of the glyoxylate cycle and gluconeogenesis . The soluble carbohydrates produced may be transported throughout seedling tissue supporting seedling growth until the seedling has become established as a self-sufficient photoautotrophic organism (Beevers, 1961; Crawford, 1977; Bewley and Black, 1994). The majority of storage reserve mobilisation occurs in Arabidopsis in the few days following imbibition of the seed, peaking with the emergence of the radicle and cotyledons (Eastmond and Graham, 2001). The transcription of a number of genes encoding enzymes involved in these processes have been shown to peak in the first couple of days post-imbibition and then decrease rapidly (Comai et al., 1989; Eastmond and Graham, 2000; Eastmond et al., 2000; Germain et al., 2001). The mobilisation of storage reserves and the initiation of germination appear to be distinct developmental events which are independently regulated. The β-oxidation mutant ketoacyl CoA thiolase-2 undergoes germination but then demonstrates developmental arrest which may be rescued by the application of exogenous carbohydrate (Germain et al., 2001). ABA has been identified as a negative regulator of germination (Kucera et al., 2005); the application of exogenous ABA to germinating Arabidopsis seeds repressed the expression of genes involved in storage mobilisation but an accumulation of sucrose as a result of catabolism of stored resources was observed. The prevention of germination by ABA led to the suggestion that two

programmes were operational, one involved in the mobilisation of storage reserves, largely unaffected by ABA; and a developmental control pathway blocked by ABA (Pritchard et al., 2002). Although preparation for seedling establishment has begun during this time period, Arabidopsis seedlings still remain responsive to environmental perturbations; treatment with ABA or elevated levels of carbohydrate have been shown to arrest early seedling development, even in seeds which have already germinated (Zhou et al., 1998; Lopez-Molina and Chua, 2000; Dekkers et al., 2004).

1.2 Regulation of germination and seedling development

The release of seed dormancy and promotion of germination is subject to control by a number of factors. Of primary importance is the growth potential of the embryo and the physical restraints imposed by the structures surrounding the seed. However, environmental factors such as light, temperature, chemical application (e.g. KNO_3) and storage conditions interact with plant hormones, such as abscisic acid (ABA), gibberellic acid (GA), ethylene, brassinosteroids (BR), auxin and cytokinins, to influence gene expression in a complex network of interactions which regulates the developmental process (Derkx and Karssen, 1993; Bewley, 1997; Koornneef et al., 2002; Kucera et al., 2005; Finch-Savage and Leubner-Metzger, 2006). A major challenge presented by the numerous factors with a role in developmental regulation is to identify principal regulators within this network and to describe their interaction with other pivotal regulators (Bentsink and Koornneef, 2002).

1.2.1 Light as a regulator of germination and early seedling development

Light is a critical factor in seed germination and early seedling development. It has been demonstrated that approximately 20% of the genome in *Oryza sativa* and Arabidopsis seedlings (Jiao et al. 2005) is regulated by white light. Four types of light receptors have been characterised; phototropins, cryptochromes,

phytochomes and an unidentified receptor for ultraviolet B (UV-B) light (Chen, 2004; Jiao et al., 2007). The photoreceptors are characterised according to the wavelength of light perceived; UV-B light (282 – 320 nm) is perceived by an unknown receptor, UV-A and blue light (320 – 500 nm) through cryptochromes and phototropins, and red or far-red light is perceived through phytochrome receptors (Sullivan, 2003). A germination-specific role for phototropin receptors has not been identified, but phototropins have been implicated in the control of directional growth towards a light source. This essentially allows photosynthesis to occur at an optimal rate through mechanisms involving stomatal opening and chloroplast movement (Chen et al., 2004). The Arabidopsis phytochrome-interacting bHLH protein, PIL5 has been shown to interact with GA and ABA to negatively regulate germination (Oh et al., 2006; 2007). PIL5 has been shown to repress genes for GA biosynthesis and the catabolism of ABA; and to indirectly induce genes involved in GA catabolism and ABA biosynthesis. The direct induction of GA repressor genes by PIL5 has also been described. A phytochrome-mediated light signalling cascade degrades PIL5, resulting in the promotion of germination (Oh et al., 2006; 2007).

Photomorphogenesis is a developmental programme activated when seedlings are exposed to light. Seedlings grown in the dark undergo a period termed *skotomorphogenesis* where seedlings are characterised by an etiolated phenotype with long hypocotyls and unexpanded cotyledons which are protected by an apical hook. This growth phase occurs in the natural environment to allow the seedling to grow through the soil layer and emerge into the light. Upon exposure to the light, the seedlings commence the photomorphogenet growth phase where hypocotyl elongation ceases, cotyledons unfold and become photosynthetic, the shoot apical meristem becomes active and true leaves begin to form (Howell, 1998). Phytochrome response to the ratio of red light (660 nm) to far-red light (730 nm) is known to influence photomorphogenesis; phyotchromes may take on one of two spectrally distinct forms, one absorbing in the red light range, Pr and one absorbing in the far-red light range, Pfr. The presence of Pfr induces a signalling cascade that results in the induction of genes involved in photomorphogenesis (Casal et al., 2003; Rockwell et al., 2006). The effects of UV-B light have been shown to promote early seedling development by inducing the cessation of hypocotyl elongation and promoting cotyledon expansion (Kim

et al, 1998). Although the UV-B response element has not been identified, it has been demonstrated that the receptor system differs from that of the cytochromes and phytochromes but that low-level doses of UV-B enhance a phytochrome B de-etiolation response (Boccalandro, 2001). Cryptochromes also have a regulatory role in the photomorphogenic process and are frequently known to act in conjunction with phytochromes (Chen et al., 2004). It has been suggested that the blue-light induced inhibition of hypocotyl elongation is mediated by a cryptochrome response that suppresses the levels or sensing of auxin and GA (Folta et al., 2003).

1.2.2 Phytohormone regulation of germination and early seedling development

The major plant hormones involved in the control of seed germination are abscisic acid (ABA), gibberellic acid (GA), ethylene, brassinosteroids (BR), auxin and cytokinins (Bewley, 1997; Koornneef et al., 2002; Kucera et al., 2005; Finch-Savage and Leubner-Metzger, 2006; Nemhauser et al., 2006). GA, BR and auxins are involved in the control of plant size and expansion along longitudinal axis, eythlene and cytokinins are implicated in the control of cell expansion and ABA is a negative regulator of germination, acting in an antagonistic manner to GA and BR (Nemhauser et al., 2006). The roles of, and interactions between phytohormones are areas of considerable research interest. A brief overview of the main regulatory activities of each hormone is provided below.

1.2.2.1 Abscisic acid regulation of development

The role of ABA as a regulator of seed dormancy and germination is well characterised. In general, ABA is thought to induce and maintain a dormant state of the seed. Correspondingly, ABA is a negative regulator of the release of seed dormancy and the promotion of germination. Mutants deficient in ABA biosynthesis, sensing or signalling show reduced periods of dormancy and

exhibit precocious germination, while ABA mutants with a hypersensitive response to ABA demonstrate a prolonged period of dormancy on ABA concentrations that do not normally affect the wild-type (reviewed in Kucera et al., 2005). ABA levels are highest during seed development and decline to low levels as the seed matures and enters a phase of desiccation (Berry and Bewley, 1992). The presence of ABA during seed development is thought to act to prevent precocious germination (Black, 1991; Hilhorst, 1995, Karssen, 1995). ABA production during seed maturation may originate either from the seed or from the surrounding maternal tissue. It appears that ABA function during this time is dependent on the source; ABA produced by the developing seed acts to control dormancy and prevent vivipary, while ABA produced from maternal tissues affects the rate of embryo growth and seed yield (Karssen et al., 1983; Koornneef and Karssen, 1994; Frey et al., 2004). A deficiency in ABA during seed development results in the production of a mature seed with a thin layer of testa, facilitating the penetration of the radicle during germination (Toorop et al., 2000).

The last stage of ABA biosynthesis in Arabidopsis is catalysed by abscisic aldehyde oxidase 3 (AAO3), which is also the focus of a feed-back loop regulating ABA biosynthesis (Xiong et al., 2001). Disruption of this step in biosynthesis results in lower ABA levels in the seed and correspondingly a reduced period of dormancy (Gonzalez-Guzman et al., 2004). Conversely, over-expression of genes involved in ABA biosynthesis, resulting in the increase of ABA content within the seed can lead to a prolonged dormancy period and the delay of germination (Koornneef et al., 2002; Leubner-Metzger, 2003; Nambara and Marion-Poll, 2003). A similar phenotype of prolonged dormancy is observed when catabolism of ABA is disrupted (Kushiro et al., 2004).

1.2.2.2 Gibberellic acid regulation of development

GAs have been demonstrated to have a role in the promotion of seed germination. The biosynthesis of GAs and GA precursors is observed during seed development for a number of species (Lange and Graebe, 1993; Lange 1998) and has been shown to be essential for development to proceed (Swain and

Singh, 2005; Kim et al., 2005). Although the precise function during this developmental stage remains unknown, GA biosynthesis during seed development is implicated in embryo growth, the prevention of seed abortion, fruit development and the assimilation of nutrients (Swain et al., 1997; Hays et al., 2002; Koornneef et al., 2002; Kucera et al., 2005). Two main functions have been proposed for GA-mediated promotion of germination; the first is the action of GA to increase the growth potential of the embryo and the second is to overcome the mechanical restraints imposed by the endosperm and testa by weakening the tissues surrounding the embryo (Koornneef et al., 2002, Leubner-Metzger, 2003; Kucera et al., 2005). A role in the regulation of the rate of germination has also been proposed for GA. The germination of seeds with no photo-dormancy response was shown to occur with no requirement of GA. However, the application of GA to these seeds increased the rate of germination (Leubner-Metzger, 2001, 2002). Numerous experiments have suggested an antagonistic relationship between ABA and GA levels, whereby GA requirement is modulated by endogenous levels of ABA within the seed. These observations include the inhibition of endosperm weakening and rupture by ABA (Toorop et al., 2000; da Silva et al., 2004; Müller et al., 2006) and the description of temperature responses mediated by both ABA and GA signalling (Gonai et al., 2004).

1.2.2.3 Ethylene regulation of development

The plant hormone ethylene has been shown to affect plants in a variety of ways including fruit ripening, prevention of stem and root elongation, senescence, wound and pathogen response and in the promotion of seed germination (Howell, 1998). Considerable research has been performed into the dramatic morphological alterations observed in dicotyledonous seedlings grown in the dark in the presence of ethylene. The observed alterations have been termed the *triple response* which includes the inhibition of hypocotyls and root elongation, radial swelling of the epicotyl and root cells, and the absence of the normal geotrophic response (Crocker et al., 1913). Increased ethylene levels are thought to alter the distribution of auxin across the shoot axis, resulting in unequal

growth and the increased coiling of the apical hook. The triple response and exaggerated coiling of the apical hook associated with increased ethylene levels are thought to aid emergence of the seedling from the soil under normal growth conditions; should the seedling encounter an obstacle, the increased production of ethylene promotes radial swelling of the stem, providing extra mechanical support for the seedling. The apical hook is prevented from uncoiling, thereby minimising damage to cotyledons and the altered geotropic response allows the seedling to find an alternative route to the surface (Goeschl 1966; Howell, 1998).

A role in the promotion of germination in non-dormant seeds has been implicated for ethylene. The completion of germination is associated with a peak in ethylene levels, but ethylene biosynthesis is also observed during imbibtion and prior to radicle emergence (Kucera et al., 2005). Although the ethylene-mediated release of dormancy is known, a number of examples have been described where ethylene is able to promote germination of non-dormant seeds, but is not sufficient alone to release dormancy, indicating that interactions occur with other regulatory factors (Kucera et al., 2005).

The induction of the triple response is reliant on the ability of the seedling to sense and respond to ethylene levels. Inhibitors of ethylene biosynthesis, sensing and response prevent the physiological changes observed in the triple response, allowing the triple response to be used as a tool for screening seedlings for ethylene response mutants (Ecker, 1995). A number of constitutive-response mutants have been identified which demonstrate the triple response even in the absence of exogenous ethylene. These mutants may be classified according to the response to ethylene synthesis inhibitors; the constitutive response phenotype of ethylene over-producers is reduced in response to synthesis inhibitors while true constitutive-response mutants are unaffected by inhibitors. Ethylene insensitive mutants deficient in aspects of ethylene biosynthesis, signalling or perception fail to demonstrate the triple response phenotype, even in the presence of exogenous ethylene (Howell, 1998).

1.2.2.4 Brassinosteroid regulation of development

The growth promoting brassinosteroid (BR) group is known to contain over 40 naturally occurring BRs, the most biologically active of which is brasssinolide (BL) (Bishop and Koncz, 2002). BRs are known to induce cell expansion, and are especially active in young tissue (Clouse and Sasse, 1998). Although mutants with decreased levels of BRs display delayed germination, BRs do not appear to be essential for germination to proceed as the BR mutants *de-etiolated2* (*det2*) and *brassinosteroid insensitive1* (*bri1*) eventually germinate without BR (Steber and McCourt, 2001). Interactions between BRs and other regulators of plant development are indicated by the convergence of auxin and BR signalling pathways in the promoters of shared genes (Nemhauser et al., 2004). A role for BR regulation of seed germination was proposed by Leubner-Metzger (2001) whereby BR was observed to positively affect seed germination through enhancing the growth potential of the embryo prior to rupture of the seed coat in tobacco (*N. tabacum*) seeds.

BRs have also been implicated in the regulation of photomorphogenesis in Arabidopsis and tomato seedlings; dark-grown BR-deficient mutant seedlings do not show the expected etiolated phenotype and instead exhibit short hypocotyls with expanded cotyledons (Clouse, 1997). Li et al. (1996) described the rescue of a constitutive photomorphogenic mutant *det2* by the application of BL. Further investigation revealed that the *det2* mutant encoded a protein that showed homology to an enzyme central to mammalian steroid synthesis, indicating that *det2* was defective in the synthesis of BSs, without which the seedling experienced premature photomorphogenesis. Seedlings of BS mutants also exhibit dwarfism in the light but may be rescued by the application of exogenous BS (Szekeres et al., 1996).

1.2.2.5 Auxin and cytokinin regulation of germination

A role for auxins has been implicated in a number of aspects of plant growth and development including cell elongation, vascular differentiation, lateral branching of roots and shoots, phototropism and gravitropism (Droog et al., 1993). During

embryogenesis, auxin levels are thought to organise cells to correctly direct future cell fate (Fischer-Iglesias and Neuhaus, 2001). The major naturally occurring auxin found in plants is indole-3-acetic acid (IAA) (Hobbie, 2007). Increased levels of IAA have been shown to occur during germination of seeds of Scots pine (Ljung et al., 2001) and bean seeds (Bialek et al., 1992). Auxin-resistant mutants of *N. plumbaginifolia* demonstrate reduced dormancy periods and interactions with GA are suggested (Rousselin et al., 1992). It is unclear whether auxin levels play a role in the regulation of germination or are simply associated with events that occur during germination. Ogawa et al. (2003) demonstrated that the application of exogenous GA induced the expression of a number of genes associated with auxin. In conjunction with the previously described association with ethylene in the triple response these results suggest that auxin is indirectly or directly involved in seed germination (Howell, 1998; Kucera et al., 2005)

Cytokinins have been identified in developing seeds and are shown to accumulate in endosperm tissue, suggesting a role in the promotion of embryonic cell division during germination (Mok and Mok, 2001). Early studies on lettuce embryos demonstrated that the application of cytokinins was able to overcome ABA-mediated inhibition of germination (Bewley and Fountain, 1972) and a number of studies have described the ability of cytokinins to break seed dormancy, indicating a positive regulatory role in development (Cohn and Butera, 1982).

1.2.3 Plant hormone signalling interactions

The hormone-balance theory proposed by Wareing and Saunders (1971) suggests that simultaneous expression of ABA and GA is required to control the progression of seed germination. Numerous examples exist which describe interactions between hormone levels during development (reviewed in Kucera et al., 2005; Nemhauser et al., 2006). These interactions have led to the proposal of a central growth regulation pathway for the integration of hormone signalling. The DELLA protein family has been shown to be responsive to GA, ABA, auxins, ethylene and ethylene precursors, leading to the suggestion of a role as

repressors of the proposed central signalling pathway (Archard et al., 2003, 2006; Fu and Harberd, 2003, Nemhauser et al., 2006). However, Nemhauser et al. (2006) investigated the transcriptional response to seven plant hormones including ABA, GA, BR, ethylene, auxin, cytokinin and jasmonic acid and found no evidence of overlap in the early response of plants to the various hormones; very few common target genes were identified in hormone response. The authors concluded that based on current information, no common transcriptional response to hormone signalling may be observed and that responses are specific and independent of cross-talk between hormone signalling. It was suggested that further localised investigations over a longer time period might reveal more common signalling mechanisms (Nemhauser et al., 2006).

1.2.4 Temperature as a regulator of germination and early seedling development

Temperature plays an important regulatory role in the timing of seed germination in the natural environment, allowing a seed to germinate only when the conditions are optimal for survival. Exposure of dry seeds to low temperatures under dark conditions, or to high temperatures during summer may be required for synchronous germination to occur (Ali-Rachedi et al., 2004; Steadman, 2004). The altered response of a number of plant hormone levels to temperature variations has been described. For example, the exposure of seeds to low temperatures which promote germination has been shown to enhance GA biosynthesis and response (Yamaguchi and Kamiya, 2001; Yamauchi et al., 2004). The Blue Micropylar End 3 (BME3) transcription factor has been identified in Arabidopsis seeds in response to cold stratification. The transcription factor was suggested to have a role in the GA biosynthesis, thereby promoting germination as a result of exposure to cold temperatures (Liu et al., 2005). The SPATULA (SPT) transcription factor found in Arabidopsis has been identified as a repressor of seed germination and as a regulator of seed response to cold stratification; the loss of function of SPT allows germination to proceed without stratification. Interaction between cold response and photoreceptor

signalling was demonstrated by the light-dependent stability of the transcription factor (Penfield et al., 2003).

1.2.5 Nutrient sensing in germination

Nutrient sensing programs have been described for carbohydrate sensing in bacteria and yeast systems, allowing cells to become aware of and respond to perturbations in environmental conditions (Postma et al., 1993; Rolland et al., 2002). Although the study of nutrient sensing mechanisms is complicated by the metabolism and transport processes of nutrients which precede the elicited response, it has been shown that division of plant meristem cells requires the establishment of a suitable nutritional state (Holsbeeks et al., 2004; Francis and Halford, 2006). The existence of Arabidopsis mutants, such as the Cape Verde Island mutant, which exhibits a seed dormancy not relieved by the application of exogenous GA indicates that alternative factors, which may involve nutrient sensing, are required for germination to proceed (Ali-Rachedi et al., 2004; Penfield et al., 2006). In addition to its role as a nutrient providing a nitrogen supply to plants, nitrate has been implicated as a signalling molecule which affects various aspects of plant development and metabolism (Scheible et al., 1997; Wang et al., 2003). Nitrate levels have been shown to affect germination in Arabidopsis; the application of exogenous nitrate stimulates the phytochrome A mediated promotion of germination and a role in the release of a dormant state has been suggested for nitrate (Batak et al., 2002; Alboresi et al., 2005). A role in the regulation rather than promotion of germination has also been suggested for nitrate. Dormancy levels observed in Arabidopsis seeds are decreased when mother plants are grown under conditions that favour the accumulation of nitrate (Alboresi et al., 2005). These results suggest that the nutritional state of the seedling might also be counted amongst factors known to regulate germination.

1.3 Metabolic regulation of gene expression

1.3.1 Metabolite signalling in bacteria

Bacteria are reliant on the immediate environment for the availability of nutrients to support growth and development. As such, bacteria must be able to respond to environmental fluctuations through rapid alterations in gene expression levels in order to metabolise available resources or to prevent synthesis of nutrients already available in the environment. One of the means in which this may be accomplished is through metabolite sensing and signalling mechanisms within the cell. One of the earliest examples of metabolic regulation of gene expression was described for the *lac* operon in *E. coli* (Jacob and Monod, 1961). *E. coli* uses glucose as a source of carbon under normal conditions and, in the absence of lactose, the *lac* repressor blocks transcription of the operon. In the presence of an alternative carbon source, the *lac* operon is activated by the binding of lactose to the repressor which prevents repression of the operon and allowing transcription to proceed (Gilbert and Muller-Hill, 1966; Oehler, 1990). The lactose pathway is an example of a pathway operated by induction although a number of other control mechanisms are employed by bacterial cells. The metabolism of tryptophan is an example of a pathway operated by repression where tryptophan activates the repressor protein so that is binds to the operator and prevents transcription of genes involved in tryptophan metabolism (Lewin, 2000). A further mechanism in the regulation of bacterial gene expression is attenuation, commonly used in the biosynthesis of amino acids. The formation of alternative structures by the mRNA of the operon either allows or prevents transcriptional progression. Transcription is inversely regulated by the abundance of the amino acid product of the operon (Yanofsky, 1981).

The presence of glucose has been shown to decrease the expression of many bacterial operons. This occurs through a mechanism termed *catabolite repression* which reflects the preference of bacterial metabolism for glucose by the inhibition of alternative metabolic pathways. Glucose acts to reduce levels of cyclic-AMP (cAMP) in the cell (Greenstein and Greenstein, 1996). It has been shown that the expression of catabolite-regulated operons is inversely correlated

to cAMP levels. In *E.coli*, cAMP binds to the product of the catabolite activator protein (CAP) gene which is a positive control factor required for the initiation of transcription at dependent promoters (Kolb et al., 1993). There are two effects of CAP binding; one is to affect the binding of RNA polymerase, the other alters the structure of the DNA. The direct contact between the α-subunit of RNA polymerase and CAP has been shown to activate transcription. A second effect of CAP binding is an alteration in the structure of the DNA itself; a bend of 90° based around the TGTGA promoter consensus sequence occurs, which has been hypothesised to aid the contact between bound CAP and the α-subunit of RNA polymerase (Lewin, 2000).

The potential for organic acid regulation of gene expression in bacteria is described by the effects of acetate on gene expression. Many types of microorganisms are able to utilise acetate as a carbon source; global gene expression profiling identified a number of differentially expressed genes between *E. coli* grown on acetate as a sole carbon source and media containing glucose. The 370 genes repressed by growth on acetate included a number of genes associated with glucose utilisation cell replication, translation and transcription, while the 354 genes up-regulated in response to acetate included genes associated with the glyoxylate and TCA cycles and also gluconeogenesis (Oh et al., 2002).

1.3.2 Metabolite signalling in yeast

A well documented example of the metabolic regulation of gene expression in yeast is the sugar-mediated induction and repression of transcription of a number of genes. The importance of catabolite repression of gene expression has been described for the hexokinase-mediated response to glucose in *Saccharomyces cerevisiae* (Rose et al., 1991). One of the earliest microarray experiments described the large-scale re-organisation of gene expression in *Saccharomyces cerevisiae* in response to glucose-depletion with the induction of genes involved in carbohydrate storage and the TCA/glyoxylate cycle and the repression of genes associated with protein synthesis (DeRisi et al., 1997). The regulation by fatty acids of a number of yeast genes involved in the glyoxylate cycle and β-

oxidation has also been described (Schöler and Schüller, 1993; McCammon, 1996; Caspary et al., 1997; Maeting et al., 1999). Genes involved in acetate metabolism, and potential regulators of gene expression, have been identified in yeast mutants unable to utilise acetate as a carbon source (McCammon, 1996; Dennis et al., 1999). A fungal acetate regulatory gene which encodes a transcription factor was shown to be homologous to yeast proteins involved in carbon metabolism (Todd et al., 1998).

1.3.3 Sugar signalling in plants

Perhaps the best characterised example of metabolic regulation of transcription in higher plants is the extensive range and number of genes responsive to sugar signalling (Koch, 1996). One of the earliest investigations into the metabolic regulation of gene expression in higher plants described the repression of photosynthetic gene transcription in maize protoplasts. The potency of sugar signalling was demonstrated by the overriding of other forms of genetic regulation by sugars, including developmental stage and light (Sheen, 1990). Since then, research on metabolic signalling in plants has focused largely on carbohydrate levels. An abundance of carbohydrate has been shown to induce genes involved in carbon storage and utilisation while a reduction in sugar levels positively regulates genes associated with reserve mobilisation, photosynthesis and carbon transport (Koch, 1996; Sheen, 1990, Grierson et al., 1994, Gibson, 2005). Sugar signalling is also known to play a role in the timing of developmental stages as well as an involvement in the formation and number of organs produced (Gibson, 2005).

Sugar sensing and signalling responses have been demonstrated for both hexoses and for sucrose. Chevalier et al. (1996) demonstrated that ribosomal protein gene expression is influenced by sugar concentration. Repression of gene expression was observed under sugar starvation conditions and under treatment with non-metabolisable sugars while treatment with fructose, glucose and sucrose induced gene expression. Numerous examples exist in the literature of genes that respond to the presence of sucrose alone including the induction of the patatin storage gene in potato and the induction of nitrate reductase in

Arabidopsis (Grierson et al., 1994; Cheng et al., 1992). The reversible conversion of sucrose and UDP to UDP-glucose and fructose is catalysed by sucrose synthase (Chourey and Nelson, 1976). Sowokinos and Varns (1992) demonstrated that the expression of the gene encoding sucrose synthase responded to the presence of sucrose but not glucose. These findings were reinforced by the demonstration that a potato sucrose synthase gene was induced by the presence of sucrose, but not glucose (Fu and Park, 1995). A number of sugar signalling pathways have been proposed in plants including a hexokinase-sensing system, a hexose transport associated sensor and a sucrose-specific signalling pathway (Smeekens and Rook, 1997). The extent and complexity of carbohydrate regulation of gene expression in higher plants demonstrates the potential power of metabolite signalling (Koch, 1996). It is possible that further investigation into metabolic regulation of gene expression might reveal similar ranges of transcriptional regulation for other metabolites.

1.3.4 Fatty acid signalling in plants

Although less well characterised than sucrose-mediated gene expression, a role in the regulation of gene expression in higher plants has been described for a number of fatty acids. Jasmonic acid is amongst the best characterised fatty acid derivatives and has been implicated in the regulation of a large number of genes involved in pollen and seed development, ozone response, defence against insect attack and wound responses (Farmer, 1994; Blechert et al., 1995; Creelman and Mullet, 1997; Reymond and Farmer, 1998; Thomma et al., 1998). A role in the regulation of transcription for the 18-carbon fatty acid oleic acid has been identified for defence response genes and also in the induction and inhibition of protein activities in plants (Fox et al., 2000; Wang and Wang, 2001; Chandra-Shekara et al., 2007). The 12-carbon fatty acid lauric acid has been shown to induce the transcription of genes associated with both the degradation and synthesis of fatty acids in *B. napus* (Eccleston and Ohlrogge, 1998).

1.3.5 Organic acid signalling in plants

A number of studies have described the potential for organic acids to act as metabolic regulators of gene expression in plants. Elevated levels of citrate and isocitrate have been shown to increase expression of the *AOX1* gene encoding the mitochondrial alternative oxidase (Vanlerberghe and McIntosh, 1996). The repression of genes associated with glycolysis by pyruvate has been described for starved sycamore cells (Ho et al., 2001). The organic acid acetate, in its activated form of acetyl-CoA, is a pivotal compound in cellular metabolism and forms the ultimate precursor for many cellular compounds including amino acids, sugars, fatty acids and hormones. Acetyl-CoA *in vivo* is usually obtained from the catabolism of fatty acids, sugars and amino acids but plants are also able to utilise exogenously supplied acetate through the activity of the glyoxylate cycle (Canvin and Beevers, 1961; Vickery and Vickery, 1981). Given the central role that acetyl-CoA plays in metabolism, it is an attractive hypothesis that acetate is a potent regulator of gene expression. A number of studies have suggested a role in the regulation of transcription for acetate in plants. One of the earliest investigations into the role of metabolites as signalling molecules investigated the repression of photosynthetic genes in maize protoplasts by sucrose, glucose and acetate (Sheen 1990). Based on differences in the behaviour of photosynthetic gene promoters it was concluded that acetate was a potent inhibitor of gene expression and that a signalling pathway distinct from that involved in carbohydrate signalling was employed (Sheen, 1990). The repressive effects on the maize phosphoenolpyruvate carboxylase gene fused to the GUS (β-glucuronidase) reporter gene were described for glucose and acetate, with acetate repressing gene expression to a greater extent (Kausch et al., 2001). Acetate has been implicated in the induction of the glyoxylate cycle genes isocitrate lyase and malate synthase (Graham et al., 1992).

1.3.6 Cross-talk between signalling pathways

As sessile organisms, plants have to sense and respond in an appropriate fashion to environmental perturbations and other factors such as development. This involves conveying information through a series of signalling pathways. As the elucidation of signalling pathway components progresses, more and more intersections between pathways are uncovered which has led to the consideration of cellular signalling as an interconnected network of signalling pathways rather than a series of distinct events. Intersections between related signalling pathways are well known, for example numerous examples exist of cross talk between pair-wise combinations of the ethylene, jasmonic acid and salicylic acid pathogen defence signalling pathways (reviewed in Kunkel and Brooks, 2002). Interactions between two or more signalling pathways have been described as 'matrix effects' and are extremely well characterised for interactions between carbohydrate and a number of other signalling pathways including nitrogen (Coruzzi and Bush, 2001; Corruzzi and Zhou, 2001; Price et al, 2004), plant hormones (Zhou et al., 1998; Yuan and Wysocka-Diller, 2006) and light (Thum et al., 2004) amongst others (Smeekens, 2000, Rolland et al., 2002, Halford and Paul, 2003). The co-ordinate repression of photosynthetic genes in a maize protoplast system by sugars and acetate suggests an intersection between organic acid and carbohydrate signalling (Sheen, 1990). Furthermore, the germination of a number of Arabidopsis acetate non-utilisation mutants on sucrose resulted in a range of altered seedling phenotypes compared to the wild-type, again indicating some degree of signalling cross-talk (Turner, 2003; Hooks et al., 2004).

1.4 Global profiling techniques

1.4.1 Genome profiling

The genome of an organism contains all of the information potentially required by an organism throughout its life cycle. Genome sequencing projects aim to describe the number, position, and function of all the genes possessed by an organism. The first genome sequence described was that of the phi-X virus (Sanger et al., 1977). The elucidation of the genetic sequences of several other organisms such as simian virus 40 (SV40) and the bacteriophages T7 and lambda demonstrated that technological advances allowed the genetic sequences of these, and theoretically, larger organisms to be determined (Reddy et al, 1978; Fiers et al., 1978; Sanger et al., 1982; Dunn and Studier, 1983). The method utilised by these techniques is however a relatively expensive means of elucidating the genome sequence; a reference sequence for the human genome was produced at a cost of approximately one billion US dollars (International Human Genome Sequencing Consortium, 2004). Advances in genome sequencing technology that allow automation and high throughput analysis have led to a decrease in the cost of sequencing over time. A corresponding increase in the number of sequencing projects and available sequence information may be observed; the Genome Project database (available online at NCBI, http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj) currently lists over 1700 complete and ongoing genome sequencing, assembly, mapping and annotation projects.

The aim of genome-scale gene annotation projects is to assign a function to each gene identified in the genome. Gene functions may be identified experimentally or using bioinformatics techniques. Experimental assignation of gene function typically involves the production of a mutant which either over- or under-expresses the gene of interest or the production of a labelled gene product. The resulting phenotype is observed from which a function for the gene is determined. Another commonly used gene identification technique exploits the abundance of sequence information available to researchers to assign functions to genes on the basis of similarities in their sequences.

1.4.2 Gene expression profiling

The information provided by genome sequencing projects describes the genetic complement of an organism. In order to identify genes which are actively contributing to an observed phenotype it is necessary to identify genes which are induced or repressed by examining associated gene transcript levels. Changes in gene expression have historically been determined in a variety of ways including differential display, RNA blots, ribonuclease protection assays and reverse-transcription polymerase chain reaction (Alwine et al., 1977; Zinn et al., 1983; Liang and Pardee, 1992; Higuchi et al., 1992; McCulloch et al., 1995). The main drawback of these analytical techniques is the limited number of gene expression events that may be determined in a single experiment. In order to identify complex patterns of gene expression behaviour, techniques allowing the simultaneous quantification of expression levels of numerous genes were developed including serial analysis of gene expression (SAGE), macro- and micro-arrays (Brown and Botstein, 1999).

SAGE technology uses available sequence information to identify genes expressed in a sample. A short complementary sequence tag of approximately 20 bases is constructed for each gene transcript present in a sample. The full complement of tags constitutes a tag library for the sample of interest. The tag library is then compared with available sequence information to determine the identify of the expressed genes present in a sample, and how many times each tag is found, allowing semi-quantification of gene expression (Velculescu et al., 1995). As SAGE does not require any *a priori* sequence information, genes and gene variants may be identified. This however, does mean that each transcript tag produced from a sample may not have a corresponding match in the sequence databases; 70 to 75% of the SAGE tags produced from *O. sativa* and Arabidopsis leaves did not match genes in the sequence databases (Matsumura et al., 1999; Jung et al., 2003).

DNA microarrays enable the simultaneous detection of transcript levels of thousands of genes. The hybridisation-based techniques utilises fluorescent dyes as a labelling device. Oligonucleotides (termed *probes*) corresponding to gene sequences are immobilised on an inert surface, such as a glass microscope slide, or a nitrocellulose membrane. Transcript RNA is isolated from the samples of

interest and labelled with a fluorescent dye in a reverse transcription reaction to produce a labelled cDNA copy of each active gene in the sample (termed the *target*). The labelled target is hybridised to the immobilised probes and the resultant fluorescent signals are then detected at the appropriate wavelengths using high resolution scanning devices. Some microarrays allow the hybridisation of more than one labelled target (e.g. from an experimental and a control sample) to the same microarray. This allows direct comparison between transcript levels and ratios of altered gene expression events may be recorded, which reflect the relative amounts of transcript RNA in each sample hybridised (Aharoni and Vorst, 2002). Microarray technology demonstrates very high sensitivity, detecting hybridisation signals at a copy number of just a few mRNA molecules per cell (Schena, 1996). One of the first published microarray experiments described the simultaneous differential expression measurement of 45 Arabidopsis genes (Schena et al., 1995). Since then, microarray design has advanced and microarray measurement of gene expression is commonly used to measure gene expression in a number of organisms. The increase in genome sequence information has resulted in the availability of microarrays which represent the majority of the genes in a genome for a number of organisms.

There are currently two main types of arrays in use; printed arrays and photolithographic arrays. Printed arrays may be constructed using either contact printing, where solid or split pins are used to deposit the probe DNA onto a solid medium, or non-contact printing where DNA probes are directly ejected in solution from a defined distance onto a solid medium. After the DNA probes are deposited on the microarray array they are immobilised on the surface of the support medium and the remainder of the surface is blocked to prevent further DNA binding to the microarray. Microarrays produced using photolithographic techniques involve oligonucleotide synthesis directly onto the surface of the support medium. A light shield is placed over all the probe sites that do not require a specific nucleotide addition and a modified nucleotide is washed over the surface of the support, linking to those nucleotides not shielded from the light source in this manner a known sequence of oligonucleotides may be built up at each spot on the microarray surface (Aharoni and Vorst, 2002).

1.4.3 Protein profiling

The term *proteomics* was first used in reference to large-scale protein identification through a combination of amino acid analysis and 2D-gel electrophoresis (Wilkins et al., 1996). The primary aim of a proteomic analysis is to comprehensively describe the protein complement of a biological sample but information contained within the proteomics field has expanded to include databases of protein sequences, protein structures and expression analysis (Heazlewood and Millar, 2003; Edwards and Batley, 2004). Commonly used protein identification techniques involve digestion of the sample using a cleavage agent such as trypsin followed by a mass spectrometry-based analytical technique. Two of the most frequently used detection techniques are Matrix-Assisted Laser Desorption/Ionisation Time-Of-Flight (MALDI-TOF) and a tandem mass spectrometer such as an ElectroSpray Quadrupole TOF (ESI-Q-TOF) (Heazlewood and Millar, 2003). MALDI-TOF is a high-throughput technique producing a peptide mass fingerprint for each protein which is then compared with a database of theoretical fingerprints to identify predicted patterns which match those identified in the sample while ESI-Q-TOF is used to produce a peptide sequence for each of the digest products (Heazlewood and Millar, 2003).

1.4.4 Metabolite profiling

The entire set of metabolites produced by an organism is termed the 'metabolome' and is functionally characteristic of the system at any given time point (Fiehn, 2002; Sumner et al., 2003). The ultimate aim of any study of the metabolome is to quantify all the metabolites present in a sample under a given set of circumstances (Weckwerth, 2003). Currently, the major limitation of metabolomics is the inability of any one analytical technique to describe every type of metabolite in a sample due to the large number and diversity of metabolites. It has been estimated that the number of metabolites synthesised by higher plants is in excess of 200,000 (Fiehn, 2002, Sumner et al., 2003, Weckwerth, 2003) while the total number of metabolites present is estimated at

up to 500,000 in total (Hadacek, 2002). It has been estimated that a single plant species such as Arabidopsis might synthesise up to 5000 metabolites (Bino et al., 2004; Von Roepenack-Lahaye et al., 2004).

This diversity and number of metabolites presents a problem for the comprehensive quantification of the metabolome. Accordingly, a number of different techniques for metabolite characterisation are available. Mass spectrometry (MS) techniques such as fourier transform ion cyclotron resonance (FT-ICR) and time of flight (TOF) may be used alone or in combination with chromatographic techniques such as gas chromatography (GC-MS), high performance liquid chromatography (HPLC-MS) and capillary electrophoresis (CE-MS). Nuclear magnetic resonance (NMR) is also used as a technique for rapid and selective metabolite detection. Each technique offers a different sensitivity, range or accuracy of metabolite measurement; as no single technique currently exists for the comprehensive description of the metabolome, techniques are often used in conjunction with each other to optimise metabolite detection coverage and accuracy. For example, mass spectrometry based analyses typically take longer to perform, but demonstrate good sensitivity and selectivity while NMR provides a more rapid means of analysis, but shows less sensitivity (Hall et al., 2002; Sumner et al., 2003; Oksman-Caldentey and Saito, 2005).

Metabolite characterisation has been classified into four categories; metabolite fingerprinting, metabolite target analysis, metabolite profiling and metabolomics (Fiehn, 2002). *Metabolite fingerprinting* is commonly used to screen members of a population to identify an altered metabolic state in response to a perturbation such as genetic modification. It is not necessary to identify each individual metabolite in a sample; rather metabolite peak profiles patterns are compared using techniques such as cluster analysis to identify similarities between sample profiles. *Metabolite profiling* aims to quantify a pre-defined set of metabolites; this may be arranged to investigate a particular class of metabolites, or to produce an overview of several different types of metabolites. *Metabolite target analysis* is designed to quantify one or a few specific metabolites of interest. This enables a very specific extraction and clean-up process to be utilised, allowing highly accurate metabolite identification and quantification. The aim of *metabolomics* is to describe as fully as possible in an unbiased fashion the metabolic complement of a sample so that changes in

metabolite abundance from different samples may be compared with one another. Metabolomic profiling takes into account all identified peaks, including those for which the exact identity is unknown, therefore a major part of metabolomic studies is dedicated to assigning metabolic identities to unidentified peaks (Fiehn, 2002).

1.4.5 Systems biology

The advent of global profiling techniques has allowed the description of activity at a number of different levels of higher plants in an entirely new fashion. The sequencing of entire genomes, for example that of *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000), has been followed by the rapid development of high throughput techniques to allow the simultaneous quantification of gene transcripts (Lockhart et al., 1996; Aharoni and Vorst, 2000; Girke et al., 2000), proteins (Shevchenko et al., 1996; Millar et al., 2001) and metabolites (Fiehn et al., 2000; Roessner et al., 2000). At a specific level, the description of as many components as possible provides information relating to the behaviour of that level of an organism under specified conditions. However, to understand the fluid and dynamic network of interactions that occur in response to perturbations of the system and developmental stages requires integration of information from different levels of the organism. A new discipline has emerged termed *systems biology*, which is based on integrating information from interdependent constituent parts of a system in order to understand the observed phenotype (Ideker et al., 2001b; Kitano, 2002; Trewavas, 2006). Handling the large amounts of data generated by profiling experiments has necessitated the development of computational techniques to analyse and model the data. The integration of computational and wet-lab techniques is a central component of systems biology; information is gathered from experimental procedures and a computational analysis used to produce a model which may be tested and refined through further wet-lab experiments. A number of systems biology approaches have been used to investigate biological interactions between transcriptome and proteome information in yeast (Futcher et al., 1999; Gygi et al., 1999; Ideker et al, 2001a) and between transcript and

metabolite data in higher plants to reveal information on pathogen response, nutritional stress and developmental changes (Mercke et al., 2004; Hirai et al., 2004; Urbanczyk-Wochniak et al., 2003, 2005).

1.5 Thesis synopsis

The primary purpose of this project was to investigate potential metabolic signals with the ability to influence gene expression in the model dicotyledonous plant *Arabidopsis thaliana*. Initially a wide range of tissues were sampled, including stem, leaves, siliques and tissues stressed with pathogen and light treatment. A method obtained from the literature for the combined extraction of metabolites, RNA and proteins was used but proved unsuitable for use with the majority of tissues investigated. Consequently, a subsection of plant development was investigated involving the transition from heterotrophic to predominantly autotrophic metabolism observed during germination and seedling establishment. A holistic experimental design was utilised in order to produce expression profiles for both gene expression and metabolite levels throughout development.

The methodology of this thesis predominantly focused on the production and analysis of transcriptomic data. Interactions with collaborators with expertise in metabolite quantification were exploited to produce metabolite profiles corresponding to transcript data. Central to the aims of this work was the measurement of gene expression levels using microarrays and subsequent analysis of the multivariant datasets. A microarray dataset produced by Dr. M. Hooks and Dr. K. Greville provided the means to gain expertise in the methodology, with direct intention to publish. The dataset is the foundation of chapters 3 and 4 of this work.

The initial microarray datasets analysed used acetate non-utilisation mutants isolated from screening on fluroacetate, a toxic analogue of acetate. Analysis of metabolite levels in the mutant indicated that an overall disruption to metabolism occurred in the mutant. Acetate has been identified as a signalling metabolite able to influence gene expression, and with a potential regulatory role in germination. The acetate non-utilising mutants identified from the FAc screen

demonstrated interesting growth phenotypes, especially when germinated on media containing elevated levels of sucrose. Transcriptome profiles were used to investigate the effects on gene expression in a mutant disrupted in acetate signalling.

The expertise in the analysis of multivariate datasets gained was used to expand the investigation of metabolite signalling from a single metabolite identified as a transcriptional regulator to the identification of multiple metabolites with the potential to affect gene expression. The metabolite and transcript profiles produced form the basis for chapter 5 of this work. Interactions with collaborators with expertise in the representation of relationships in complex datasets were used to visualise correlated gene transcript and metabolite levels, representing potential regulatory relationships during development. These correlations form the basis of chapter 6.

Chapter 2

Materials and Methods

2.1 General reagents and equipment

Unless otherwise stated, general laboratory chemicals and biological reagents of Molecular Biology grade or equivalent were obtained from Fisher Scientific U.K. (Leicestershire, U.K.), Sigma-Aldrich (Dorset, U.K.) and VWR International Ltd. (Leicestershire, U.K.). All pipette tips, microcentrifuge tubes and Falcon tubes were obtained from Starstedt (Leicestershire, U.K.). Petri dishes were obtained from Barloworld Scientific Ltd. (Staffordshire, U.K.). Liquid volumes of less than 1000 µl were sampled and dispensed using Biohit Proline pipettors (Biohit Plc., Helsinki, Finland). Unless otherwise stated, all centrifugation steps were carried out in a MSE HAWK 15/05R Benchtop Centrifuge. A MSE mistral 3000i Benchtop Centrifuge was used to centrifuge 15 ml and 50 ml Falcon tubes (both centrifuges from Measuring and Scientific Equipment Ltd. Sanyo, Sussex, U.K.). All photography was performed using a Lumicron LDC-825Z3 digital camera (Lumicron Technology UK Ltd., Berkshire, U.K.). All other equipment used is referenced at the first use.

2.2 Consequences of disrupting acetate metabolism in Arabidopsis seedlings

2.2.1 Biological material

Arabidopsis thaliana ecotype Columbia: Col-7, and a corresponding T-DNA acetate non–utilisation mutant (*acn1-2*) (Weigel et al., 2000, Hooks et al. 2004)

were used to investigate the transcriptome-level effects of disruption of acetate metabolism.

2.2.2 Plant growth and maintenance

Arabidopsis seeds were collected from the siliques of dried mature plants. Stem and seed pod material were initially sieved to isolate the seeds. The remaining material was pulled gently over a clean sheet of paper. Seeds were stored in glass vials in the dark at room temperature. Arabidopsis seeds were surface sterilised prior to sowing. A stock sterilising solution was prepared consisting of 1 g of crushed Covclor 1,000 chlorine tablet (Coventry Chemicals Ltd., Coventry, U.K.) dissolved in 8 ml of dH_20. A drop (approximately 30 µl) of 1% (v/v) Tween 20 was added. A dilute solution was prepared by combining 1 ml of the stock solution with 9 ml 100% industrial methylated spirits (IMS). The diluted solution was left at room temperature for 5 minutes to allow a precipitate to form. The solution was centrifuged at 2,500 rpm for 5 minutes to pellet the precipitate and the supernatant removed prior to use.

Arabidopsis seeds were weighed into a 1.5 ml Eppendorf tube and sterilised with 1 ml of the diluted sterilising solution for 30 minutes with gentle agitation at room temperature. The sterilising solution was carefully poured off and the seeds washed twice with 1 ml 95% IMS for 1 minute. The IMS was removed and a pipette tip was used to spread the seeds around the walls of the Eppendorf. The seeds were left to air dry in a laminar flow cabinet for approximately two hours.

Arabidopsis seeds were germinated on plates containing ½-strength MS salts (Murashige and Skoog, 1962) and 0.8% (w/v) agar. The media were adjusted to pH 5.7 with 0.1 M KOH, added to the agar and autoclaved. Plates were poured in a laminar flow cabinet when the media had cooled to approximately 50°C. Surface sterilised Arabidopsis seeds were scattered on ½ MS plates that were sealed with Micropore™ surgical tape (3M United Kingdom PLC, Berkshire, U.K.) to prevent contamination. Plates were wrapped in foil and stored at 4°C for 4 days in order to stratify germination. The plates were placed in a growth

room in a random configuration under continuous light at 70 µmol of photons/m^2/s at 20°C.

2.2.3 Seed harvesting

2.2.3.1 Col-7 and acn1-2 *seedlings for metabolite and RNA extractions*

In order to obtain conclusive results regarding the behaviour of mutant compared to wild-type seedlings from the metabolite and transcript profiling, it was necessary to obtain equally developed seedlings. A comparison of the germination and growth phenotypes of the *acn1-2* seedlings showed that the *acn1-2* mutant appeared to germinate at the same rate as the wild-type, but established slightly quicker (Hooks et al., 2004). After four days in growth under continuous light both seedling types had attainted a principal growth stage of 1.0 where the cotyledons had fully greened and expanded (Boyes et al. 2001). However, the *acn1-2* seedlings could attain this stage of development up to 12 h before the corresponding wild-type parent. Therefore, in addition to visually staging seedling development, we selected batches of seedlings with approximately the same average fresh weight. The average fresh weights of 50 seedlings from plates used for RNA extraction were 20.1 ± 2.1 and 21 ± 1.2 for Col-7 and *acn1-2*, respectively, at PGS 1.0.

2.2.3.2 Seedlings used for seed production

Seedlings that were to be used as mature plants for seed propagation were sterilised and germinated as described above. After 10-14 days, the seeds were transferred to seed trays containing an sterile 4:1 mixture of B&Q Multipurpose Compost (B&Q plc., Hampshire, U.K.) and J. Arthur Bower's Vermiculite (William Sinclair Horticulture Ltd., Lincolnshire, U.K.). Immediately prior to sowing, the compost was treated with 0.2 g/L of the insecticide Intercept (Ropax, Lancs., U.K.). Seedlings were grown under continuous light at 70 µmol of photons/m^2/s and 20°C until required. The soil was kept moist by watering with

dH_20 every few days. When the primary shoot was produced, plastic collars were fitted to the plants, to force shoot growth in one direction. A few days later, plastic tubes were fitted into the collars to ensure that shoots were self contained. This prevented cross-fertilisation from occurring and facilitated the harvesting procedure. When the plant reached maturity and siliques were yellow or brown in colour, watering was stopped and the plant dried for approximately two weeks prior to harvesting. Seeds were harvested by transferring the aerial section of the plant to a paper bag, which was stored at room temperature for a minimum of 4 weeks to allow the material to dry out completely.

2.2.4 Hot phenol chloroform RNA extraction

Ribonucleases (RNase) are highly stable enzymes whose activity can affect the quality and quantity of the extracted RNA, so must be removed from the RNA extraction process. All glassware, spatulas and bottles used in the preparation of RNA were baked at 180°C for eight hours to remove RNase contamination. Electrophoresis tanks were cleaned with 0.5% (w/v) SDS, rinsed with RNase-free dH_20, rinsed with 70% ethanol and allowed to dry before use (Sambrook et al., 1989). Solutions were prepared using autoclaved H_20 and, where possible, were treated with 0.1% (v/v) diethyl pyrocarbonate (DEPC) overnight and autoclaved. All dH_2O used in RNA extractions was treated with 0.1% (v/v) DEPC in the same manner. Solutions containing Tris were not treated in this way, as DEPC reacts rapidly with amines; these solutions were prepared using RNase-free dH_20 (Sambrook et al., 1989). Molecular biology grade water (Sigma-Aldrich) guaranteed RNase and dexyribonuclease (DNase) free was used to re-suspend RNA pellets and was also used in all microarray reactions. Aliquots of water were taken and DEPC treated as described above prior to use.

2.2.4.1 RNA extraction

Phenol: chloroform: isoamyl alcohol (IAA) (25:24:1 v/v/v) solution was prepared. Equal amounts of phenol and chloroform:IAA (24:1 v/v) were mixed together and equilibrated by extraction with 0.1 M Tris-HCl (pH 7.6). The solution was stored under an equal volume of 0.01 M Tris-HCl (pH 7.6) at 4˚C in the dark. Immediately prior to use, β-mercaptoethanol was added to the extraction buffer (25 mM Tris-Cl, pH 8.0, 25mM EDTA, pH 8.0, 75 mM NaCl, 1% (w/v) SDS) to a final concentration of 7.8% (v/v). A pestle and mortar were pre-cooled with liquid nitrogen and the sample tissue (up to 500 mg fresh weight) was ground to powder, to which 900 µl extraction buffer were added and further ground to powder. Phenol: chloroform: isoamyl alchohol (25:24:1 v/v/v) was pre-heated to 65˚C and 900 µl added to the sample which was further ground until liquid and transferred to a 1.5 ml Eppendorf tube. The mortar was washed with 100 µl extraction buffer and 100 µl phenol: chloroform: isoamyl alchohol (25:24:1 v/v/v) added to the sample. The sample was centrifuged at 13,000 rpm at 4˚C for 10 minutes. The aqueous layer was removed to a new Eppendorf and 1 volume of phenol: chloroform: isoamyl alchohol (25:24:1 v/v/v) added. The sample was centrifuged at 13,000 rpm at 4˚C for 10 minutes. The aqueous layer was removed to a new Eppendorf and 10 M LiCl added drop by drop to a final concentration of 2 M. The sample was incubated at 4˚C overnight to precipitate the RNA. The sample was centrifuged at 13,000 rpm at 4˚C for 15 minutes to pellet the RNA. The supernatant was carefully removed and the pellet washed with 500 µl 2M LiCl, stored at 4˚C. The Eppendorf was briefly vortexed and the sample was centrifuged at 13,000 rpm at 4˚C for 10 minutes and the supernatant removed. The pellet was resuspended in 300 µl RNase-free dH_2O and 0.1 volume of 3M NaOAc, pH 5.5 was added. The sample was vortexed and 2.5 volumes of 100% Ethanol were added. The sample was incubated at -20˚C for 1 hour to allow the RNA to precipitate. The sample was centrifuged at 13,000 rpm at 4˚C for 15 minutes to pellet the RNA. The supernatant was removed and the pellet washed with 70% Ethanol. The supernatant was removed and the pellet allowed to air dry at room temperature for approximately 15 minutes. The RNA was resuspended in 30 µl RNase-free dH_2O for 1-2 hours on ice.

2.2.4.2 Quantification of RNA

RNA was quantified using a UniCam UV 500 spectrophotometer and Vision 32 Software (both from Thermo Spectronic New York, U.S.A.). The absorbance of the sample at 260 nm and 280 nm was measured. The amount of RNA was determined using the absorbance at 260 nm where an optical density of 1 unit corresponded to an RNA concentration of approximately 40 µg/ml. The purity of the RNA was determined by the ratio of the readings at 260 nm and 280 nm. A ratio of between 1.6 and 2.0 was obtained for good quality RNA; anything below this was indicative of contamination with protein (Sambrook et al., 1989).

2.2.4.3 Visualisation of RNA

The quality of the extracted RNA was visually examined using a denaturing formaldehyde agarose gel with an ethidium bromide stain. A gel was prepared with 1.5% agarose, 1X MOPS, pH 7.0 (from a stock solution of 10x MOPS: 0.2 M MOPS, 80 mM NaOAc pH 5.2 and 10 mM EDTA, pH adjusted to 7.0 with 10 M NaOH and stored at 4°C in the dark) completed to volume with RNase-free dH_2O. The gel was heated in a microwave (Daewoo, Antrim U.K.) to melt the agarose and then left to cool to approximately 50°C, 37% formaldehyde was added to a final concentration of 2.5% (v/v) and the gel poured into the mould and left to set for approximately 1 hour. A 2 µg aliquot of RNA was diluted to a total volume of 8 µl using RNase-free dH_2O. The sample running buffer (0.9% (v/v) ethidium bromide, 23% (v/v) 10x MOPS, 6.2% (v/v) formaldehyde and 69% (v/v) deionised formamide) was prepared and 6µl added to the RNA sample. The samples were denatured at 65°C for 2.5 minutes and chilled on ice. The samples were run at 40v for 4 hours with a 1X MOPS running buffer and the RNA visualised using a TM-20 UV Transiluminator (Ultra Violet Products Ltd., California, U.S.A.) and photographed using a K61B Mitsubishi Video Copy Processor (Mitsubishi, Tokyo, Japan).

2.2.5 Affymetrix microarray data processing

The Affy service at the Nottingham Arabidopsis Stock Centre (NASC) was used to monitor RNA quality, produce and hybridise labelled cDNA to the ATH1 GeneChip Microarrays (Craigon et al., 2004). The ATH1 GeneChip array is designed so that each Arabidopsis gene located on the microarray is represented by a set of eleven probe pairs, each consisting of a perfect match (PM) probe and a mismatch (MM) probe. PM probes are designed to precisely match the sequence of a specific gene transcript while MM probes differ from the transcript sequence by one base. Hybridisation intensity values for MM probes are considered to be background or non-specific hybridisation of target and are subtracted from the hybridisation intensity values for PM probes to produce an accurate representation of the amount of labelled target hybridised to the probe (Lipshutz et al., 1999).

2.2.6 Affymetrix signal detection, quantitation and normalisation

In order to determine whether a hybridisation signal was detected for a gene transcript the PM and MM probe pair intensity values were used to calculate a discrimination score for each probe on the array by comparing the target specific intensity difference (PM-MM) of the probe pair to the overall intensity of the probe pair (PM+MM). The discrimination score was compared to a detection threshold; if the discrimination score was higher than the detection threshold then the signal from the gene transcript was identified as *present*. Conversely, if a probe pair scored lower than the threshold then the probe was detected as being *absent*. If the discrimination score for a probe fell very close to the threshold then the probe was determined as *marginal* (Afffymetrix Statistical Algorithms Reference Guide, 2001).

The relative gene expression levels of probes identified as present were represented by a quantitative signal intensity value calculated using the One-Step Tukey's Biweight Estimate which produced a robust weighted mean for each probe pair set. The MM intensities are subtracted from the PM signal intensities and each resultant value is logged. The median PM – MM value for the probe

set was calculated and individual probe pairs were weighted according to how closely the adjusted value is to the median value. Weighted intensities are used to calculate the mean of the probe set so that PM – MM values closer to the median value have a greater effect on the weighted mean than those further away, thereby minimising the effect of outliers. The weighted mean was transformed to linear scale to produce the quantitative signal intensity for the probe pair set (Affymetrix Statistical Algorithms Reference Guide, 2001).

A basic normalisation procedure was applied to the signal values obtained from each GeneChip microarray. This aim of this procedure was to normalise signal intensity distribution on each microarray so that they could be compared with one another. All of the signal values from the microarray were ordered and the top and bottom 2% removed. The mean was calculated for the remaining 96% of the signal values and a scaling factor value calculated to adjust the mean to 100. Each of the signal intensity values was multiplied by the scaling factor value to normalise signal intensities on an array-by-array basis (GeneChip Expression Analysis technical Manual, 2004). The quality of the resultant data was demonstrated by the closely correlated signal intensity values between microarray sample replicates (figure 2.1).

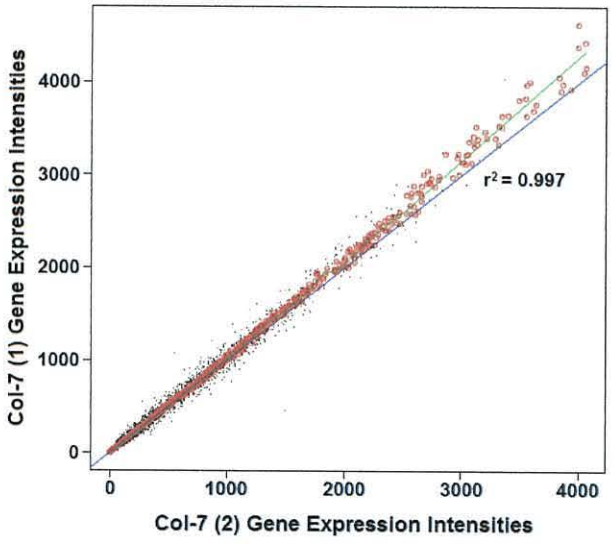


Figure 2.1: Scatterplot to show the correlation of signal intensity values obtained for two Col-7 replicate Microarrays. The high r^2 value of 0.997 indicates that the data are highly reproducible. Similar correlation coefficients were obtained for pair-wise comparisons between all replicated microarrays.

2.2.7 Data normalisation using GeneSight

The GeneSight v. 4.1 analytical software (BioDiscovery, Inc) was used to normalise, visualise and analyse the transcriptome datasets produced for Col-7 and *acn1-2* . The software provided a pre-set "Affy" data transformation sequence for use with transcriptome profiles generated with Affymetrix microarrays. The first step in the transformation sequence removed all signal intensities deemed unsuitable for analysis. In the case of Affymetrix datasets this refers to spots with absent or weak signal flagged as 'A' or 'M'. Secondly, negative signal intensities have to be made positive in order to allow a logarithmic transformation to be applied later in the data transformation sequence and so probes with signal intensity values less than 0 were omitted from the analysis in this step. Incorrectly high ratios can result from genes with very low expression values; consequently probes with low signal intensity values were raised to a minimal threshold of 1.0 in the third step of the transformation sequence (Cui and Churchill, 2003). In the fourth step of the transformation sequence signal intensities were normalised in order to eliminate technical variation between the microarrays by dividing the signal intensity values for the remaining genes on the microarray by the value of the genes in the 75^{th} percentile of signal intensities. The fifth step in the transformation sequence was the application of a logarithmic transformation (base 2) which stabilises variance in the dataset. The differences in gene expression values were calculated between the designated experimental and control microarrays and the signal intensities from replicated spots in each microarray were combined and their variance calculated

2.2.8 Identification and removal of outliers

In order to identify gene expression values with consistent ratios between *acn1-2* and Col-7 a dataset was created so that each of the three *acn1-2* replicate microarrays were compared with a single Col-7 replicate microarray (table 2.1).

Table 2.1: A dataset was created to identify genes with consistent expression ratios between *acn1-2* and Col-7 replicate microarrays. The dataset was composed of three groups to allow average expression values for each *acn1-2* replicate compared with a common control sample to be determined

Group	Experimental Replicates	Control Replicates
1	*acn1-2* (1)	Col-7 (1)
	acn1-2 (2)	Col-7 (1)
	acn1-2 (3)	Col-7 (1)
2	*acn1-2* (1)	Col-7 (2)
	acn1-2 (2)	Col-7 (2)
	acn1-2 (3)	Col-7 (2)
3	*acn1-2* (1)	Col-7 (3)
	acn1-2 (2)	Col-7 (3)
	acn1-2 (3)	Col-7 (3)

The gene signal intensities were transformed using the preset "Affy" transformation sequence detailed in the previous section with the omission of the final step which combined replicate signal intensity values. The ratios of all gene expression values identified as present in each of the six microarray replicates were calculated between *acn1-2* and Col-7. An analysis of variance test (ANOVA) was applied to the dataset to identify expression ratios significantly different between the three groups ($p < 0.05$). A group of 14256 genes was identified which were marked as present in each of the six microarray replicates and which demonstrated consistent expression ratios between *acn1-2* and Col-7.

2.2.9 Identification of genes differentially expressed between Col-7 and *acn1-2*

Historically a 2-fold change in gene expression levels has been used as a suitable threshold to identify differentially expressed genes (Schena et al., 1996; DeRisi et al., 1997). A potential drawback to this method is that there are no associated values to indicate the level of statistical confidence of the genes identified as differentially expressed. An additional disadvantage is that the fold change threshold is an arbitrary value which might lead to the inclusion of either false-

positive or false-negative values in the differentially expressed gene set (Cui and Churchill, 2003; Breitling et al., 2004). The use of statistical tests to identify differentially expressed genes also has associated problems as it is possible that a relatively small change in gene expression levels might produce a statistically significant result but in reality have very little biological relevance if many other genes show much larger fold-changes in expression (Breitling et al., 2004). It was therefore decided to identify a set of genes differentially expressed between Col-7 and *acn1-2* using a combination of a fold change threshold and a statistical test. In this way a statistically confident group of differentially expressed genes was obtained which also demonstrated biologically relevant fold-changes in expression. The dataset described in the previous section (table 2.1) was used to identify genes differentially expressed in *acn1-2* compared to Col-7. The pre-set "Affy" transformation sequence was applied to the data with no modifications and all three experimental replicates in each group were compared in a pair-wise manner with a single control sample. Gene expression values were subsequently combined to produce an average expression value obtained for each gene in each group and were identified as differentially expressed only if the expression values were significantly different in all three groups.

2.2.9.1 Identification of an appropriate fold-change threshold

An appropriate fold-change threshold level to identify differential gene expression was determined using a combination of a bootstrapping algorithm and hierarchical clustering (Kerr and Churchill, 2001). Following convention, a bootstrapping algorithm was initially used to identify differentially expressed genes with a minimum 2-fold increase or decrease in expression in each of the three groups in the dataset (table 2.1) with a confidence level of 95% (Schena et al., 1996; DeRisi et al., 1997). A total number of 130 genes common to all three groups of the dataset were identified as differentially expressed at this level. A divisive hierarchical clustering algorithm was used to identify relationships between the six individual microarray replicates when the 130 genes identified as differentially expressed were removed from the dataset. Divisive hierarchical clustering algorithms initially assign all data values to one group. The most

related components of the group are identified (in this case, microarray experimental type) and the group is split into two accordingly. The resulting two groups are split further until each cluster contains a single component (Speed, 2003). Hierarchical clustering of the gene-set when the 130 genes differentially expressed with a minimum of 2-fold change in expression were removed showed that experimental samples clustered according to the source tissue, indicating that the gene set contained a sufficient number of differentially expressed genes between *acn1-2* and Col-7 to allow microarray replicates to be distinguished according to the sample hybridised (figure 2.2). Due to the stochastic nature of the clustering algorithm used, a number of different clustering arrangements were possible upon each application of the algorithm. The algorithm was therefore applied twenty times and the cluster arrangement determined upon the profile observed in more that 95% of the resultant dendrograms.

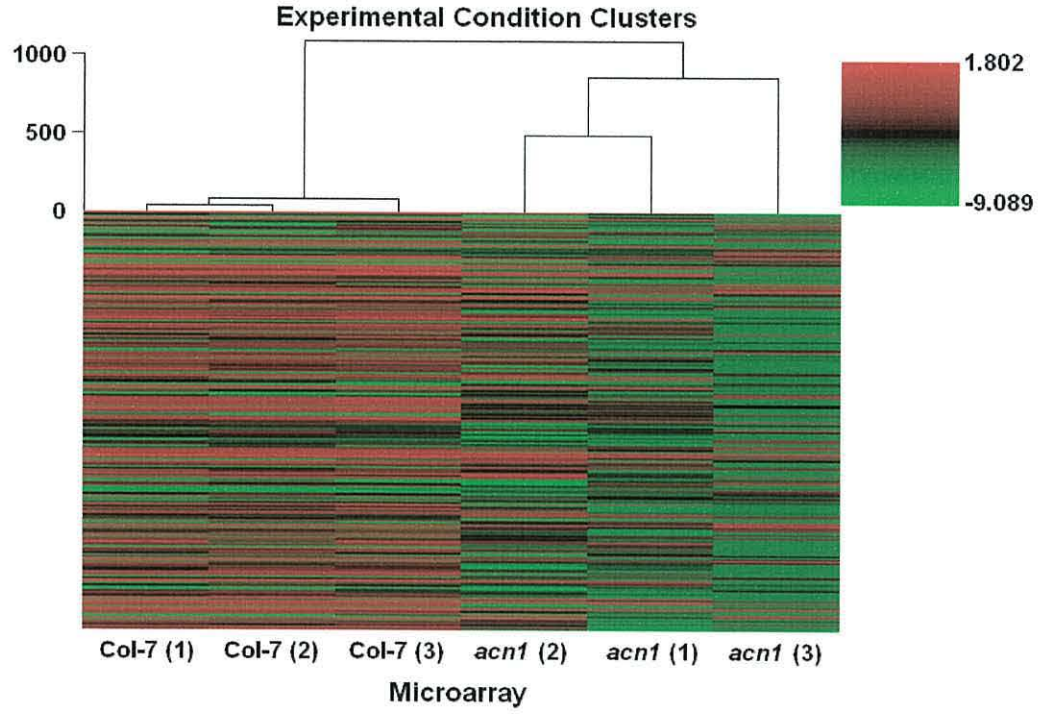


Figure 2.2: Hierarchical clustering of microarray replicates when the 130 genes identified as differentially expressed with a minimum of 2-fold change in expression were removed from the dataset. Microarray replicates have clustered together according to source tissue, indicating that the group of genes contains a sufficient number of differentially expressed genes so that sample types may be distinguished from each other.

The lowest fold-change threshold for the identification of differentially expressed genes was determined to ensure that any false negative results resulting from a 2-fold change in expression threshold were included in the final differentially expressed gene list. The threshold for differential gene expression was decreased in increments of 0.1-fold and hierarchical clusters produced from the gene set resulting from each threshold. At a fold-change threshold of 1.3-fold a set of 296 genes were identified as differentially expressed. Hierarchical clustering of the microarray replicates when these genes were removed showed that clustering of experimental replicates was disrupted, indicating that the gene set no longer contained a sufficient number of differentially expressed genes to discern between the two sample types (figure 2.3).

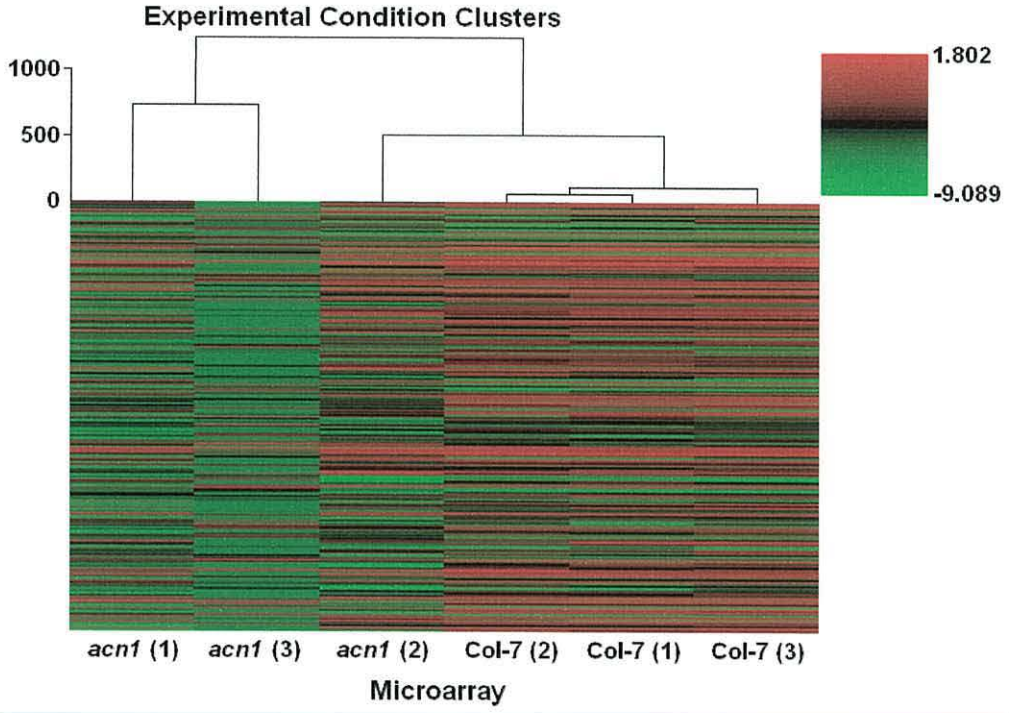


Figure 2.3: Hierarchical clustering of microarray replicates when the 296 genes identified as differentially expressed with a minimum of 1.3-fold change in expression were removed from the dataset. Microarray replicates do not cluster together according to source tissue, indicating that the gene set does not contain a sufficient number of differentially expressed genes to distinguish between the two sample types.

A set of 252 genes were identified as differentially expressed with a minimum change in expression of 1.4-fold and a confidence level of 95% (Kerr and Churchill, 2001). Hierarchical clustering of experimental replicates when these genes were removed showed that experimental replicates clustered together (figure 2.4). Decreasing the fold-change threshold below 1.4 resulted in disruption of the clustering of experimental replicates (figure 2.3). It was therefore decided that a 1.4-fold change in gene expression would be an appropriate threshold to identify differentially expressed genes between *acn1-2* and Col-7.

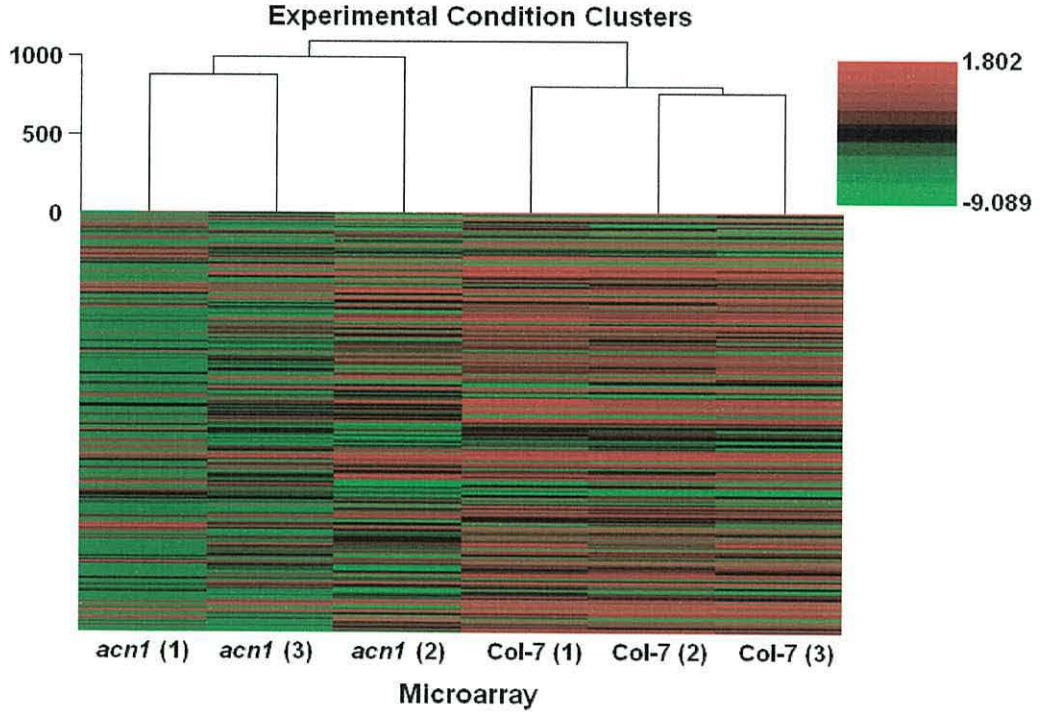


Figure 2.4: Hierarchical clustering of microarray replicates when the 252 genes identified as differentially expressed with a minimum of 1.4-fold change in expression were removed. Microarray replicates cluster together according to source tissue, indicating that the gene set does not contain a sufficient number of differentially expressed genes to distinguish between the two sample types.

2.2.9.2 Statistical integration of fold-change threshold

A t-test was used to identify genes with significantly different expression levels between *acn1-2* and Col-7 from the group of 14256 genes with consistent gene expression ratios described previously. A set of 2829 genes were identified as differentially expressed ($p < 0.05$). A subset of these genes that showed a minimum change in expression of 1.4-fold was identified in each of the three groups of the dataset (table 2.1) with a confidence level of 95% (Kerr and Churchill, 2001). The genes common to all three groups were identified to produce a set of 201 genes that were identified with statistically different expression levels between *acn1-2* and Col-7 but which also showed a minimum change in expression levels of 1.4-fold (Appendix B).

2.2.10 Identification of common promoter motifs in co-expressed genes

A gene may be described on a fundamental level as active if it produces an mRNA transcript which is later translated in to a polypeptide gene product. Not all genes are active at all times; some genes are constitutively expressed whereas others are only expressed in response to a particular stimulus, or at a specific developmental stage. Clearly some control of gene transcription must occur during these processes. In higher eukaryotes, the main level of regulation appears to occur at the transcriptional level mediated through the binding of regulatory factors to the promoter elements of genes (Schwechheimer and Bevan, 1998). It should be noted that the regulation of gene expression occurs at multiple levels in addition to the binding of promoter elements. A number of other processes may act to modulate the expression of the final gene product. For example, exons fused from alternative reading frames in the same gene may produce two distinct proteins (Ruas and Peters, 1998) while a single protein might encode more than one enzymatic reaction (Velayos et al., 2000). Changes to the structure of the DNA itself, even before transcription has occurred, may also have an affect on which genes are expressed. Commonly observed processes are those of DNA methylation and histone acetylation. DNA methylation

normally occurs at cytosine residues in particular 5'-CG-3' sequences and may induce silencing of gene expression (Chim et al, 2000). Histone acetylation involves the addition of acetyl groups to side chain groups of lyseine residues in histones H3 and H4, and lowers the affinity of histones for DNA, generally promoting gene expression. The acetylation event is also thought to be able to increase gene expression by decreasing methylation activity, as demonstrated in insulin-like growth factor II receptor in mouse (Hu et al., 2000). RNA processing also promotes diversity in the final end product metabolites. Splicing of a single coding sequence prior to RNA transcript formation may give rise to multiple mRNAs from one single coding sequence, each with different properties as each exon displays different properties (Reddy, 2001). Post-translational modification of proteins, again, promotes further diversity (Bachmir et al., 2001). This modification may be mediated by multiple factors, such as light, or internal metabolite levels. The site of polyadenylation of the molecule, to produce the poly A tail can affect the final product. For example, a change in the site of polyadenylation on the B lymphocyte in humans alters the final product from being a membrane-bound to a secreted antibody. Poly-A tails also affect the stability of the molecule, with longer tails increasing the stability of the molecule and consequently increasing its chance of translation. An additional way in which transcript diversity may be increased is by the process of RNA editing. This may involve the modification, addition or deletion of bases in the transcript. For example, the mRNA sequence for apoloprotein B-100 found in liver cells codes for an amino acid polypeptide containing 4563 amino acids. In intestinal cells, the same mRNA sequence undergoes RNA editing where a CAA triplet is converted to an UAA triplet producing a truncated polypeptide of 2153 amino acids (McKee and McKee, 2003).

The sequences bound by transcriptional regulators may enhance, repress or silence gene expression levels. The DNA sequence elements bound by transcription factors are known as CAREs (*cis* - acting regulatory motifs) and generally range from five up to twenty nucleotides in length. Although some CAREs may be found within the gene itself, the majority are located upstream from the transcription start site (Rombauts et al., 2003). The regulation of gene transcription has been described in terms of hierarchical arrangement of regulatory factors with CAREs as the elementary control mechanism. The

organisation of CAREs into clusters within individual promoters is considered as a means by which transcriptional activation of genes may be finely controlled (Rombauts et al., 2003). The compilation and comparison of the CAREs present in a promoter sequence, or set of promoter sequences of co-expressed genes might therefore reveal novel information about gene regulation.

2.2.11 Motif searching in co-expressed genes

The datasets produced from global gene expression analysis are often composed of hundreds or even thousands of genes. Grouping together of genes with similar expression profiles provides a way in which these large datasets may be easily visualised (Stekel, 2003). Following the simplified assumption that genes that behave in a similar manner might share a common regulatory mechanism, the next cognitive step is to examine these groups of genes for the regulatory machinery. The regulation of gene expression in eukaryotic organisms primarily occurs at the level of transcription through the binding of transcriptional regulators to DNA motifs located in the promoter regions of genes (Schwechheimer and Bevan, 1998).

A number of algorithms have been designed to extract and process potential DNA motif information from the promoters of co-expressed genes. The algorithms may be classified into three main types; enumeration, deterministic optimisation or probabilistic optimisation algorithms (D'haeseleer, 2006). Enumeration or word-counting algorithms consider promoter sequences to be a continuous line of text with DNA motifs hidden within as 'words'. An example of a motif prediction programme using this type of algorithm can be found in the Regulatory Sequence Analysis-tools motif prediction program (van Helden et al., 2000). In a set of related promoter sequences, motifs of a defined length are searched for and their frequency calculated. The frequency of the identified motifs is compared with the expected frequency of the motifs in a background model consisting of unrelated sequences. The flexibility of the identified motif sequences is increased by the inclusion of symbols from the IUPAC code (Appendix A) and similar motifs are grouped together to form a consensus motif sequence (Rombauts et al, 2003, D'haeseleer, 2006).

Deterministic optimisation algorithms such as that used in the MEME motif prediction program (Bailey and Elkan, 1995) search for a single *n*-mer DNA motif in a promoter sequence. For each *n*-mer sequence in the target sequences, the probability is calculated whether the *n*-mer was generated by a biologically relevant motif rather than a background sequence. Taking the weighted averages across these probabilities produces a more refined motif model. Further iterations of the algorithm alternate between refining the motif model and calculating the probability that the latest DNA motif is generated by the background sequence, until a convergence to the best fitting motif is reached (D'haesleer, 2006).

Probabilistic motif detection algorithms are designed to align conserved promoter regions to produce a multiple alignment in a set of unaligned related promoter sequences. A randomly selected set of motif sequences of specific lengths is used to initialise a motif model that is compared with all potential motif sequences in the target sequences. Motif sequences are removed or added to the motif model based on the weighted probability that they belong to the motif. The model is updated and the probabilities recalculated. Over time, the best fitting combinations are assigned to their positions to produce a consensus motif (Rombauts et al, 2003, D'haeseleer, 2006). An example of a probabilistic motif detection program is the one used in this work, MotifSampler (Thijs et al., 2002).

2.2.12 MotifSampler sequence scanning program

The sequence-scanning program used in this work was MotifSampler (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, Thijs et al., 2002), which uses a modification of the Gibbs sampling method described by Lawrence et al. (1993) to identify DNA sequence motifs which are over-represented in the promoter regions of co-expressed genes. Although a number of motif prediction programs exist, this program was selected for a number of reasons. A pre-compiled higher order background model consisting of 172 carefully selected non-related intergenic sequences from Arabidopsis (totalling 341,248 bp) was constructed for use by the program authors with the MotifSampler program

(Thijs et al., 2001). A background model of order *m* means that the probability of finding a nucleotide at a specific position in the motif is dependent on the *m* preceding bases in the sequence. A 3rd-order background model was used in this work, as the sequences under investigation were noisy. Noise in this context results from the large sections of promoter sequence that do not contain the motif sequence of interest. The stochastic nature of the clustering algorithms that were used to identify the groups of co-expressed genes means that not all sequences in a set are expected to contain the same motif sequences. The promoter sequences that do not contain the motifs are another source of noise in the analysis. Thijs et al. (2001) showed that the MotifSampler algorithm, when run with a 3rd-order background, performed consistently better than a single-order background model under increasing noise conditions on a test set of Arabidopsis genes. The higher-order background model also returned more motifs than the single-order model although there was little difference in the number of correctly predicted motifs returned by the various models.

Although the fundamental assumption of the MotifSampler is that at least one DNA motif is present in an otherwise noisy background sequence, the promoter sequences under investigation are not all expected to contain the same motif. It is also possible that each promoter sequence contains more than one copy of a motif (Thijs et al., 2002). MotifSampler addresses this problem by allowing a number of runs to be carried out and also by allowing more than one motif to be found in each promoter sequence. The positions of the motifs identified in previous runs were masked so that each new run would produce a different result. Although this setting allowed for different consensus motifs to be retrieved each time, over-represented motifs with highly conserved sequences were retrieved more frequently than poorly conserved infrequent sequences (Thijs et al., 2002, Rombauts et al., 2003).

It is desirable to perform large numbers of runs on a single set of promoter sequences. On-line motif prediction programs often do not have the computational power to handle data of this size. MotifSampler offers the opportunity to download a stand-alone version of the program to allow for as many searches as required. The user-friendly and intuitive interface and highly flexible parameter settings are also major advantages of this program.

2.3 Transcriptomic investigation of cross talk in organic acid and carbohydrate signalling

2.3.1 Biological material

Transcriptome profiles were produced using Affymetrix ATH1 GeneChip microarrays as described in the previous section for Col-7 and *acn1-2* seedlings harvested at principle growth stages 1.0 and 0.7 (Boyes *et al.*, 2001) from standard ½ MS agar plates without and supplemented with 100 mM sucrose respectively. Three replicates consisting of seedlings pooled from three agar plates were produced for each sample type. The transcriptome profiles were used to compare the global gene expression responses of Col-7 and *acn1-2* to sucrose. No osmotic adjustment was made to sucrose-treated seedlings in order to permit differential expression of stress- and development-regulated genes. Each sample consisted of seedlings pooled from three individual plates.

2.3.2 Identification of sucrose responsive genes

RNA was extracted, quantified and the quality inspected visually as described in the previous section. Standard Affymetrix procedures were applied to quantify and determine the quality of the data. Two datasets were produced to identify the sucrose responsive genes in both Col-7 and *acn1-2* (table 2.2). Each dataset was constructed to contain three groups so that each of the sucrose-treated samples was compared with a single control sample. Gene expression values were subsequently combined to produce an average expression value obtained for each gene in each group.

Table 2.2: Datasets I and II were created to identify genes differentially expressed in response to growth on 100 mM sucrose in (I) Col-7 and (II) *acn1-2*. Each dataset was composed of three groups to allow average expression values for each sucrose-treated replicate compared with a common control sample to be determined. Genes were identified as sucrose-responsive if the expression values were significantly different in each of the three groups.

Dataset I: Identification of sucrose-responsive genes in Col-7		
Group	Experimental Replicates	Control Replicates
1	Col-7 (1) plus 100 mM sucrose	Col-7 (1)
	Col-7 (2) plus 100 mM sucrose	Col-7 (1)
	Col-7 (3) plus 100 mM sucrose	Col-7 (1)
2	Col-7 (1) plus 100 mM sucrose	Col-7 (2)
	Col-7 (2) plus 100 mM sucrose	Col-7 (2)
	Col-7 (3) plus 100 mM sucrose	Col-7 (2)
3	Col-7 (1) plus 100 mM sucrose	Col-7 (3)
	Col-7 (2) plus 100 mM sucrose	Col-7 (3)
	Col-7 (3) plus 100 mM sucrose	Col-7 (3)

Dataset II: Identification of sucrose-responsive genes in *acn1-2*		
Group	Experimental Replicates	Control Replicates
1	*acn1-2* (1) plus 100 mM sucrose	*acn1-2* (1)
	acn1-2 (2) plus 100 mM sucrose	*acn1-2* (1)
	acn1-2 (3) plus 100 mM sucrose	*acn1-2* (1)
2	*acn1-2* (1) plus 100 mM sucrose	*acn1-2* (2)
	acn1-2 (2) plus 100 mM sucrose	*acn1-2* (2)
	acn1-2 (3) plus 100 mM sucrose	*acn1-2* (2)
3	*acn1-2* (1) plus 100 mM sucrose	*acn1-2* (3)
	acn1-2 (2) plus 100 mM sucrose	*acn1-2* (3)
	acn1-2 (3) plus 100 mM sucrose	*acn1-2* (3)

A combination of fold-change threshold and significance test parameters was used to identify a set of sucrose-responsive genes in Col-7 and *acn1-2*. A fold-change threshold for differential expression of genes was identified as described in the previous section. Briefly, an arbitrary value of 2.0-fold change in expression was set as the threshold for differential gene expression. Genes which satisfied this criterion in each of the three groups were identified and samples clustered on this basis. These genes were removed from the dataset and the remainder of the genes used to hierarchically cluster sample types. If the resultant dendrogram showed that sucrose-treated microarrays and control microarrays clustered into two groups then it was assumed that the identified

gene set was primarily composed of differentially expressed genes and the fold-change threshold decreased to increase the number of genes identified as differentially expressed. If the dendrogram produced showed that microarrays did not cluster according to sample type it was assumed that the gene set contained a sufficient number of genes not differentially expressed to disrupt the clustering. Due to the stochastic nature of the hierarchical clustering algorithm, clustering was performed twenty times for each fold-change threshold; a gene set was identified as primarily composed of sucrose-responsive genes if replicate microarray types clustered together in at least 95% of the resulting dendrograms. The fold-change threshold was decreased in 0.1-fold increments for each dataset to identify thresholds for differential gene expression levels (sucrose-treated vs. control) as 1.3-fold for Col-7 and 1.4-fold for *acn1-2* with a confidence level of 95% (Kerr and Churchill, 2001). A t-test (*p*-value of 0.05) was used to identify statistically significant sucrose-responsive genes in Col-7 and *acn1-2*. The differentially expressed genes common to all three groups in a dataset were identified as sucrose-responsive for either Col-7 or *acn1-2*. Genes which were identified as differentially expressed using both the determined fold-change threshold and the t-test were identified as sucrose-responsive for Col-7 and *acn1-2*. Therefore a group of 3580 sucrose responsive genes was identified for Col-7 (based on a minimum change of 1.3-fold in expression and a t-test, $p<0.05$) and a group of 3255 sucrose-responsive genes identified for *acn1-2* (using a minimum change of 1.4-fold in expression and a t-test, $p<0.05$).

2.4 Transcriptome profiling of Arabidopsis seedlings

2.4.1 Biological material

Arabidopsis thaliana ecotype Columbia: Col-0 was used for the transcriptome, analyses performed in this section. Seeds were sterilised, germinated and grown as described previously. Day 0 was defined as the day the seeds were placed in the growth room. Seeds harvested on Day 0 were sampled 1 hour after transfer to the growth room. Seeds harvested on Days 1 through to Day 8 were harvested

at 24-hour time intervals from this point. Seedlings were inspected under the microscope before harvesting to ensure uniformity of developmental stages selected for tissue extraction. Harvested material was stored at -80°C for no longer than 2 weeks before extraction.

2.4.2 RNA extraction

Two methods of RNA extraction were used to produce the transcriptome profiles in this section. A protocol designed for the integrated extraction of RNA, metabolites and proteins (Weckwerth et al., 2004) was used for seedlings harvested from Days 2 to 8. This protocol did not work well on seedlings harvested on Days 0 and 1, and so a borate-based extraction protocol (Penfield et al., 2004) was used for these samples.

2.4.2.1 Integrated RNA, metabolite and protein extraction

Total RNA was extracted from the samples based on an integrated RNA, metabolite and protein extraction protocol (Weckwerth et al., 2004). Only the steps relevant to the RNA are detailed here. A pestle and mortar were pre-cooled with liquid nitrogen and the sample tissue was homogenised under liquid nitrogen. To the sample, 2 ml of a single phase solvent mixture of methanol/ chloroform/ dH_2O (2.5:1:1 v/v/v), stored at -20°C, was added and the sample was mixed at 4°C for 30 minutes. The sample was centrifuged at 14,000 rpm at 4°C for 10 minutes. The supernatant was removed and 1 ml of methanol/ chloroform (1:1 v/v), stored at -20°C was added and the sample vortexed. The sample was centrifuged at 14,000 rpm at 4°C for 10 minutes and the supernatant removed. The pellet was incubated with 1 ml water-saturated phenol and 1 ml RNA extraction buffer (0.05 M Tris, pH 7.6, 0.5% SDS, 1% β-mercaptoethanol) at 37°C for 1 hour. The sample was centrifuged at 14,000 rpm at 4°C for 10 minutes and the phenol phase of the supernatant removed and discarded. Proteins were removed from the remaining sample by precipitation with 200 µl

chloroform. The sample was centrifuged at 14,000 rpm at 4°C for 15 minutes and the supernatant removed to a clean Eppendorf. The RNA was precipitated with 40 µl of glacial acetic acid and 1 ml of absolute ethanol at 4°C for 30 minutes. The sample was centrifuged at 14,000 rpm at 4°C for 15 minutes and the supernatant removed. The RNA pellet was washed once with 3 M sodium acetate and twice with 70% ethanol and re-suspended in RNase-free dH_2O.

2.4.2.2 Borate-based RNA extraction

RNA was isolated using a borate-based extraction protocol (Penfield et al., 2004). Polyvinylpyrrolidone, DTT and IGEPAL were added to the RNA extraction buffer (200 mM sodium borate decahydrate, 30 mM EGTA, 1% (w/v) SDS, 1% (w/v) sodium deoxycholate, pH adjusted to 9.0 with 10 M NaOH. The solution was treated with 0.1% (v/v) diethyl-pyrocarbonate (DEPC) overnight and autoclaved) to final concentrations of 2% (w/v), 10 mM and 1% (v/v) respectively immediately prior to use. A pestle and mortar were pre-cooled with liquid nitrogen and the sample tissue was ground to powder, to which 500 µl RNA extraction buffer was added and further ground to powder. The sample was then transferred to a 15 ml conical bottom centrifuge tube, allowed to thaw and a further 500 µl extraction buffer was added. To the thawed sample, 40 µl proteinase K was added and the sample incubated at 42°C for 90 minutes. From the sample, 1 ml was removed to an Eppendorf tube and KCl added to a final concentration of 160 mM. The Eppendorf was inverted several times to mix the sample and then incubated on ice for 60 minutes. The sample was then centrifuged at 13,000 rpm at 4°C for 20 minutes. The supernatant was transferred to a 2 ml Eppendorf and 360 µl 8M LiCl was added. The sample was incubated at -20°C for at least 2 hours, to precipitate the RNA which was collected by centrifugation at 13,000 rpm at 4°C for 20 minutes. The supernatant was removed and the RNA pellet re-suspended in 100 µl RNase-free dH_2O. The RNA was further purified using the RNeasy Plant Mini Kit (QIAGEN Ltd. West Sussex, U.K.) according to the manufacturer's instructions and the RNA eluted in 30 µl of RNase-free dH_2O. Total RNA extracts were visualised using agarose gel electrophoresis as detailed previously (figure 2.5).

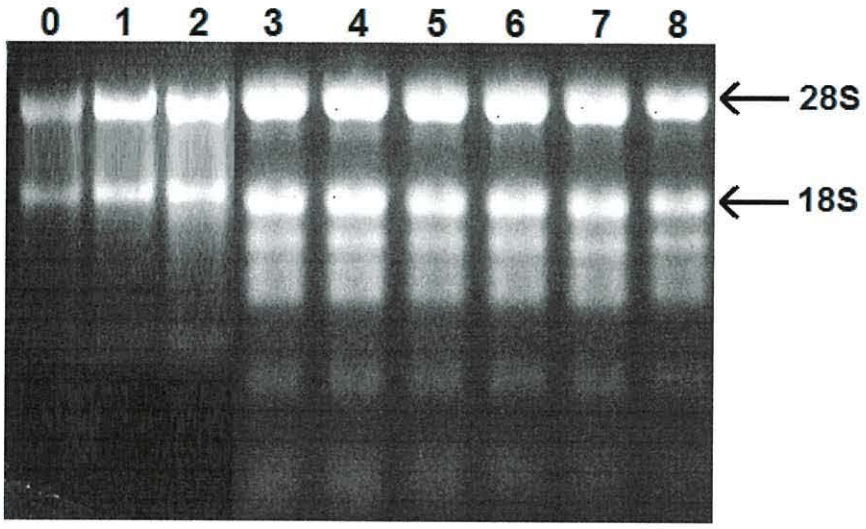


Figure 2.5: Typical RNA extractions corresponding to each time point harvested, visualised on a denaturing formaldehyde gel (0) day 0, (1) day 1, (2) day 2, (3) day 3, (4) day 4, (5) day 5, (6) day 6, (7) day 7, (8) day 8.

2.4.3 Printed microarray preparation

Printed 70-mer oligonucletoide microarrays obtained from the laboratory of Professor David Galbraith at the University of Arizona were used to produce transcriptome profiles for the seedling stages detailed above. Two array versions, 1.0 (26000 elements) and 3.0 (29000 elements) were used in these experiments, which correspond to the versions of the Qiagen DNA probe set used in the printing of the microarrays. Since transcriptome profiles were produced from more than one microarray print version, only those genes common to all microarrays were used in this analysis, which corresponded to a common set of 22890 genes.

The boundaries of the microarray were marked on the non-printed side of the microarray using a diamond-tipped pencil. The microarray slides were suspended with the printed side facing down for 8 seconds over a 250 ml beaker containing water heated to 60°C. The microarrays were snap dried with the printed side face-up on an inverted heating block at 65°C for 5 seconds and left to cool on the bench for 1 minutes. This process was repeated 4 times for each slide to re-hydrate the oligonucleotide probes. An ultraviolet (UV) oven (Ultra-Violet Products Ltd., Cambridgeshire, U.K.) was run at 120 mJ/cm^2 to pre-warm

the UV tubes. The microarray slides were placed probe side up on moistened 3MM paper and exposed to 120 mJ/cm^2 to immobilise the oligonucleotides spotted onto the slides. The microarray slides were washed in 1% (w/v) SDS for 5 minutes on a rotary shaker. The slides were washed in 100% Ethanol for 30 seconds with gentle agitiation to remove the SDS. Microarray slides were dried by centrifuging at 1000 rpm for 5 minutes and stored in a cool dry place.

2.4.4 Production of labelled cDNA

The fluorescence properties of the Cy^{TM}5-dCTP (Amersham Biosciences, Buckinghamshire U.K.) dye used to produce labelled cDNA from RNA extracts are affected by light. Therefore, all operations involving Cy5-dCTP were carried out with minimal exposure to light and samples were wrapped in foil to prevent any photo-degradation of Cy5-dCTP.

For each labelling reaction 1 µg of Oligo dT_{20} primer was added to 100 µg of total RNA in a total reaction volume of 20.5 µl. The RNA and primer were incubated at 65°C for 10 minutes, room temperature for 10 minutes and chilled on ice for 2 minutes. A reaction mix of 8 µl of 5x First-Strand Buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM MgCl2), 4 µl of dNTP mix (5 mM of dA-, dG-, dTTP and 2 mM dCTP) and 4 µl of 0.1 M DTT was added to 2 µl Cy5-dCTP, 25 nM. The reaction components were pre-warmed at 42°C for 2 minutes and 1.5µl of (200 U/µl) Superscript II reverse transcriptase (Invitrogen, California, U.S.A.) was added. The reaction components were incubated in the dark for 2 hours at 42°C. After 1 hour, a further 1 µl of Superscript II reverse transcriptase was added to the reaction. The reverse transcription reaction was terminated by the addition of 10 µl of 1M NaOH to the reaction mix. After incubation at 65°C for 10 minutes 10µl of 1M HCl was added. To adjust the pH to the optimal pH for cleaning columns, 200 µl of TE, pH 7.2 (1mM EDTA and 10mM Tris-HCl, pH 7.2), was added. The labelled cDNA was purified using a QIAquick PCR purification kit (QIAGEN Ltd. West Sussex, U.K.) according to the manufacturer's instructions with the addition of an extra wash step using 200 µl of the provided washing buffer after the first washing step. Bound cDNA was

eluted from the QIAquick spin column in a volume of 50 µl of elution buffer. The labelled cDNA was kept on ice and protected from light.

2.4.5 Quantification of labelled cDNA

The quantity of the labelled cDNA and incorporation of Cy5-dCTP were assessed using a UniCam UV 500 spectrophotometer and Vision 32 Software (both from Thermo Spectronic New York, U.S.A.). Measurements were taken at 260 nm (DNA absorbance) 280 nm (protein absorbance), 650 nm (Cy5 absorbance) and a range of values up to 850 nm (to determine the baseline absorbance). Using the absorbance values, the amount of protein contamination was determined, as described above. The extinction coefficient of Cy5 (250,000) was used to calculate the total amount of ssDNA (µg) where an optical density of 1 unit corresponded to a ssDNA concentration of approximately 33 µg/ml. Figure 2.6 shows a typical example absorbance spectrum from labelled cDNA produced from RNA for samples harvested at day 6.

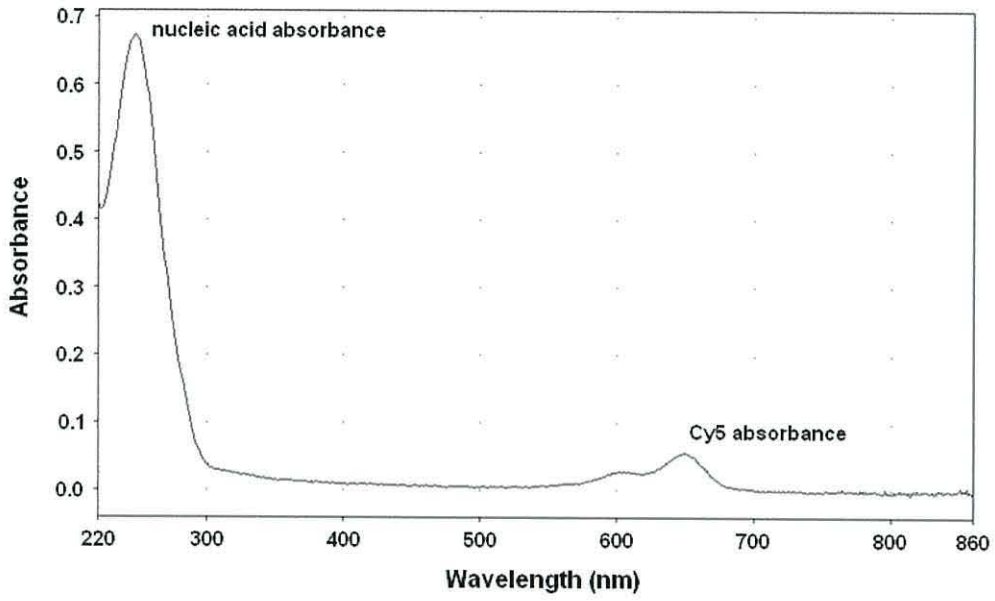


Figure 2.6: A typical UV spectrum to determine the quality and quantity of labelled cDNA produced from reverse transcription reactions for each sample to by hybridised to a microarray. The ratio of the peak observed at 260 nm (nucleic acid absorbance) to the peak at 650 nm (Cy5 absorbance) was used to determine the quantity and quality of the labelled cDNA.

2.4.6 Hybridisation of labelled cDNA to microarrays

The efficiency of fluorescent dye incorporation into the labelled cDNA product is known to vary according to the type of dye used (Stekel, 2003). The use of a single fluorescent dye in this experiment reduces the technical variation between microarrays, aiding the direct comparison of hybridisation intensities between individual arrays. A minimum of three microarrays were produced for each time point sampled, allowing generalisation of technical variability such as microarray manufacture, general handling, hybridisation and washing of the microarray slides. Two hybridisation protocols were used in the preparation of the microarrays used in this section. A manual hybridisation protocol was used for the majority of the microarrays produced using a hybridisation chamber (MWG Biotech Ltd. Ebersberg Germany). An automatic hybridisation protocol using the MAUI® Hybridization System (BioMicro Systems, Inc. Utah, U.S.A.) was used for a subset of the microarrays in order to allow a comparison between the two methods.

2.4.6.1 Microarray blocking

A blocking solution of 5x SSC (3 M NaCl and 0.3 M sodium citrate), 0.1% SDS (w/v) and 10 mg/ml bovine serum albumin (BSA), completed to volume with sterile dH_2O was prepared in a 50 ml Falcon tube and pre-heated to 42°C. Microarrays were incubated in the blocking solution in the dark for 45 minutes at 42°C, rinsed 5 times with sterile dH_2O at room temperature and dried by centrifuging at 1000 rpm for 15 minutes at room temperature.

2.4.6.2 Manual hybridisation

Sterile dH_2O was added to the edges of a hybridisation chamber (MWG Biotech Ltd. Ebersberg Germany) to create a high humidity environment. The microarray slide was placed in the chamber and the labelled cDNA target applied in two drops of approximately 12 µl to the printed surface of the microarray

using a pipette. A hydrophobic 22 x 60 mm cover-slip (Sigma-Aldrich Company Ltd., Dorset, U.K.) was cut in half and positioned on top of the microarray so that the hybridisation solution was free to move across the entirety of the printed surface. The chamber was sealed and covered with foil to protect the contents from light and incubated in a water bath at 42˚C for 16 hours.

2.4.6.3 MAUI hybridisation

Automatic hybridisations were carried out using the MAUI® Hybridization System (BioMicro Systems, Inc. Utah, U.S.A.). The system consisted of two sections: a reusable base unit which provided a constant hybridisation temperature and allowed the hybridisation solution to be mixed throughout the incubation period and an adhesive gasket containing the labelled target in a hybridisation buffer which was fixed to the microarray slide. An adhesive gasket was applied to the microarray slide and sealed at the edges. The gasket contained two air bladders that ensured continuous mixing of the hybridisation solution in the chamber. Hybridisation solution was introduced to the mixer unit through the fill port. A Gilson Microman® M-100 positive displacement pipette (Gilson Company, Inc., Ohio, U.S.A.) was used to inject the hybridization solution into the fill port of the gasket until hybridisation solution appeared at the vent-port at the opposite end of the chamber. Excess hybridisation solution was removed and the chamber sealed with adhesive plastic strips. The microarray slide with the attached mixer unit was placed on the base unit pre-heated to 42˚C. The hybridisation solution was mixed every 3.3 seconds during the 16 hour incubation period.

2.4.6.4 Washing array

The cover-slips used in both the manual hybridisations and the MAUI mixer units were removed by immersing the microarray in a wash buffer of 2x SSC and 0.1% SDS (w/v) pre heated to 42˚C. The microarray slides were washed in a 50 ml Falcon tube containing 50 ml of 2x SSC and 0.1% SDS (w/v) at 30˚C for 5

minutes with constant agitation. The microarrays were then washed in Falcon tubes containing a 1x SSC solution followed by a 0.5x SSC solution at room temperature for 5 minutes each with constant agitation. The microarrays were dried by centrifuging at 1000 rpm for 5 minutes and stored in a cool, dry dark place prior to scanning.

2.4.7 Microarray image acquisition

The microarrays were scanned at 10 μm resolution using an Affymetrix 428™ Array Scanner (Affymetrix, Inc. California, U.S.A.). False-coloured images were produced from laser-induced fluorescence of the labelled cDNA on the microarray slide and saved as image files. The quality of the image was reviewed visually and the signal strength of the microscope adjusted accordingly to maximise signals from weak spots whilst ensuring that the signal from strong spots remained within the intensity range.

2.4.8 Microarray image analysis

ImaGene® 5.0 image analysis software (BioDiscovery, Inc., California, U.S.A.) was used to produce a quantified signal intensity for each microarray spot. Images of hybridised microarrays were imported into ImaGene where each spot was located and assigned an identity according to its position on the microarray. Background signals were subtracted and spots designated as poor hybridisation events were discounted from future analysis. In order to address the problem of negative spots, signal intensities below a set value of 20 were raised to that value. A standard normalisation procedure was applied to the quantified gene expression values obtained for each printed microarray to facilitate comparisons among individual microarrays (Affymetrix GeneChip Expression Analysis technical Manual, 2004). In brief, the top and bottom 2% of the signal values were removed and the mean calculated for the 96% of the values remaining. A value, the scaling factor, was calculated to adjust the mean of the remaining

values to 100. Each of the signal intensities on the array was then multiplied by the scaling factor to normalise signal intensities on an array-by-array basis.

2.4.9 Normalisation of gene expression using reference genes

Quantitative real time PCR (Q-RT-PCR) is a sensitive, specific and rapid means to quantify transcript abundance from active genes and is therefore commonly used to verify the results of microarray experimentation (Gachon et al., 2004). When present in solution as an unbound dye SYBR Green exhibits a low fluorescence; upon binding to double stranded DNA fluorescence is increased by more than 1000-fold. As the concentration of DNA increases, the fluorescent signal measured increases proportionally, allowing the accumulation of PCR amplification product to be monitored. (Lewin, 2000). A standard curve produced from RNA, purified plasmid dsDNA or cDNA may be used to determine the original concentration of the target in samples.

Three genes were selected to confirm the microarray results, *ACTIN2* (At3g18780), *ribosomal protein S9* (At1g74970) and *ribosomal protein L32* (At5g46430). *ACT2* is commonly used as a reference gene to validate microarray gene expression (Laval et al., 2002) and was, therefore, used to normalise gene expression values before comparison. Each reference gene was spotted between 13 and 27 times throughout the microarrays, according to the version used. Primers for the reference genes were designed according to the following characteristics: a sequence length of 18-24 bases, a sequence G-C content of 40-60%, closely matching annealing temperatures for each primer in the range of 55 to 65°C. The sequences were designed to have non-complementary 3' ends, to avoid primer-dimer formation and to be non-self complementary, to avoid secondary structure formation (e.g. hairpins) (Innis and Gelfand, 1990, Saiki, 1989). Primers were designed using the Primer 3 programme (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi.) using DNA sequences obtained from the Arabidopsis Information Resource (TAIR) (www.arabidopsis.org) and synthesised by MWG Biotech Ltd. (Ebersberg, Germany). A list of the primers used for the reference genes is given in table 2.3.

Table 2.3: Primer sequences used for quantification of reference genes.

Gene		Primer Sequences (5' – 3')
Act2	Forward	CTAAGCTCTCAAGATCAAAGGCTTA
(At3g18780)	Reverse	ACTAAAACGCAAAACGAAAGCGGTT
RPL32	Forward	ATGTTGGTTACGGATCTGATAAG
(At5g46430)	Reverse	TGTTGAGATATTGTGAGCAATCT
RPS9	Forward	CTATCGTGATGCCAAGGAGTA
(At1g74970)	Reverse	GTCTGCACTTACCTTCAGGA
Rubisco	Forward	CTATGGTCGCTCCTTTCAACGG
(At1g67090)	Reverse	TGCAACCGAACAAGGGAAGC

2.4.9.1 DNase treatment of RNA extracts

Genomic DNA contamination was found in some samples. If genomic DNA was amplified along with the cDNA in the quantitative real time PCR reactions, incorrect values would have been obtained for the reference genes used for microarray normalisation. The genomic contamination was removed with the Sigma Deoxyribonuclease I Amplification Grade (AMP-D1) kit. A reaction was set up using 1 µl of the provided DNase I Solution (1 unit/µl DNase I, 50% glycerol, 10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$ and 10 mM $MgCl_2$), 1 µl of the provided 10x Reaction Buffer (200 mM Tris-Hcl, pH 8.3 and 20 mM $MgCl_2$) and 1µg RNA completed to 10 µl total volume with RNase-free dH_2O. The reaction was incubated at room temperature for 15 minutes. After the incubation period, 1 µl of the provided Stop Solution (50 mM EDTA) was added and the sample incubated at 70°C for 10 minutes. The sample was chilled on ice prior to cDNA synthesis. Successful reverse transcription reactions were confirmed by PCR (figure 2.7).

PCR reactions were carried out in 200 µl thin-walled PCR tubes in a PTC-150 MiniCycler (MJ Research Inc., Massachusetts, U.S.A.). Each PCR reaction consisted of 1x Mg-free reaction buffer, 2 mM $MgCl_2$, 200 µM dNTPs, 1.5 units of Taq polymerase (storage buffer of 50 mM Tris-Hcl, pH 8.0, 100mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 50% glycerol and 1% Triton® X-100) and 0.5 µM of each primer in a total reaction volume of 25 µl. A sample volume

of 5 µl was used. PCR cycling conditions were a denaturing step of 94˚C for 5 minutes; 35 cycles of a denaturing step of 95˚C for 30 seconds, an annealing temperature step for 30 seconds and an elongation step of 72˚C for 2.5 minutes; a final elongation step of 72˚C for 10 minutes and a holding temperature of 4˚C. Agarose gel (1.5%) electrophoresis was used to confirm the success of each reaction. Agarose was added to 1 x TAE buffer and melted in a microwave (Daewoo, Antrim U.K.). When the gel had cooled to approximately 50˚C, ethidium bromide was added to a final concentration of 0.2 µg/µl. The gel was poured into a mould and left to set for approximately 1 hour. A 1x TAE buffer was used to run the gels. DNA sample running buffer was added to the samples and an aliquot of the sample run on the gel. Molecular markers were run alongside the samples to enable sizing of the PCR products. Two different sized markers were used, depending on the size of the PCR products. For smaller PCR products a 100bp DNA ladder (Fermetas U.K., York, U.K.) was run, while for larger PCR products a 1kb DNA ladder was used (Promega Corporation, Southampton, U.K.).

Rubisco primers were used to monitor genomic contamination; the primers were designed across an intron resulting in two different PCR product sizes for genomic DNA (584 bases) and cDNA (386 bases). Negative controls consisting of all reaction components but no DNA template were performed at each stage of the process.

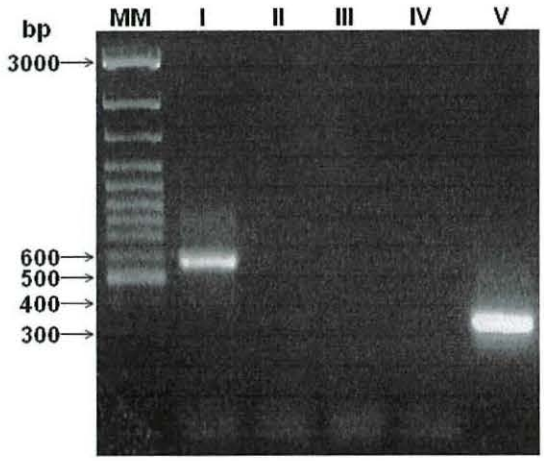


Figure 2.7: A typical PCR to confirm successful cDNA synthesis from RNA (extracted on day 2). *Rubisco* primers were used in order to check for genomic DNA contamination of samples as the predicted sizes for amplification products differ; genomic 584, cDNA 386 (MM Fermentas 100 bp ladder). (I) Positive PCR control, genomic DNA, (II) negative PCR control, (III) negative cDNA synthesis control, (IV) negative DNase treatment control, (V) day 2 cDNA.

2.4.9.2 Reverse transcription

Omiscript® reverse transcriptase (QIAGEN Ltd., West Sussex, U.K.) was used to produce cDNA from the DNase-treated RNA. A reaction mix of 1 x of the provided Buffer RT, 0.5 mM each dNTP, 1 µM Oligo dT_{15} primer, 10 units of RNasin (Promega Corporation Southanpton, U.K.), 4 units of Omniscript Reverse Transcriptase, 2 µg of template RNA was made up to 20 µl total volume with Rnase-free dH_2O. The reaction components were incubated at 37°C for 1 hour and an aliquot added to the PCR mix.

2.4.9.3 Production of purified plasmid DNA

Purified plasmid DNA target was used to produce a standard curve for quantitative PCR. A PCR product for cloning into plasmids was generated for each reference gene using conditions as described previously. A TOPO TA Cloning® kit (Invitrogen, Carlsbad California, U.S.A) was used to directly insert a PCR product for each reference gene into a plasmid vector. From the TOPO®

Cloning Reaction, 2 µl was transferred to a vial of One Shot® Chemically Competent *E. coli*, mixed gently and chilled on ice for 5 to 30 minutes. The cells were heat-shocked at 42°C for 30 seconds and immediately chilled on ice. To the cells, 250 µl of room temperature S.O.C. medium was added. The tube was capped tightly and shaken horizontally at 200 rpm and 37°C for 1 hour.

LB growth media was prepared using 1% (w/v) Tryptone, 0.5% (w/v) yeast extract, 0.02 M NaCl. The solution was pH adjusted to 5.7 with 5N NaOH and autoclaved. For plates, the media was added to 0.8% (w/v) agar before autoclaving. Appropriate antibiotics were added after the media had been cooled to 55°C. Selective LB plates containing ampicillin at a concentration of 0.1 mg/ml were pre-warmed to 37°C. Two volumes (20 µl and 50 µl) from each transformation were each added to 20 µl of S.O.C. medium and applied to selective plates, which were then incubated at 37°C overnight. Two dilutions of cells were used to ensure that at least one plate would produce evenly spread colonies. From the several hundred resulting colonies, 10 isolated colonies were selected for each reference gene, transferred to LB-ampicillin (0.1 mg/ml) plates using a sterile toothpick and incubated overnight at 37°C. A sterile toothpick was used to transfer some of the resulting colony to a standard PCR reaction mix to confirm successful transformation events (figure 2.8).

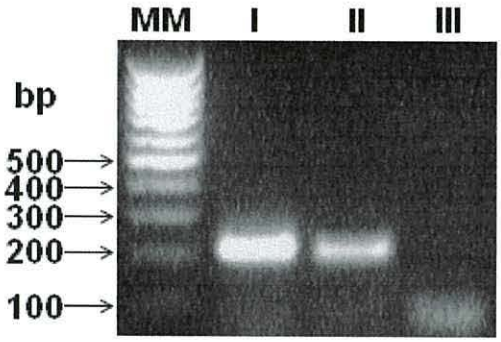


Figure 2.8: PCR confirmation of successful transformation of TOP10 vector with an *ACT2* PCR fragment (MM Fermentas 100 bp ladder). (I) Plasmid transformed with *ACT2* PCR fragment, (II) positive PCR control, genomic DNA, (III) negative PCR control.

Plamsid DNA was prepared from successfully transformed colonies following a modified protocol of Birnboim and Doly (1979). Two colonies were selected for each reference gene construct and a sterile 200 µl pipette tip used to inoculate 5 ml of LB media containing 500 µg of ampicillin. The media were incubated at 37°C overnight with vigorous shaking at 200 rpm. A 1.5 ml aliquot was transferred to an Eppendorf tube and centrifuged for 1 minute at 14,000 rpm. The medium was removed by aspiration to leave the pellet of bacteria as dry as possible. When completely dried, the pellet was re-suspended in 100 µl of an ice-cold Solution I (50 mM glucose, 10 mM EDTA and 25 mM Tris-HCl, pH 8.0, stored at -20°C), vortexed, and incubated at room temperature for 5 minutes. Immediately prior to use, Solution II (0.2 M NaOH and 1% (w/v) SDS) was prepared and 200 µl added to the Eppendorf, which was closed and inverted rapidly three times. The sample was chilled on ice for 5 minutes, and 150 µl of an ice-cold Solution III (60 ml of 5 M potassium acetate was added to 11.5 ml of glacial acetic acid and 28.5 ml of dH_2O to give a final concentration of 3 M with respect to potassium and 5 M with respect to acetate) was added. The cap was closed and the Eppendorf vortexed gently in an inverted position for 10 seconds and chilled on ice for 5 minutes. The sample was centrifuged for 5 minutes at 14,000 rpm and 4°C, and the supernatant transferred to a fresh Eppendorf. An equal volume of phenol/chloroform was added to the supernatant and vortexed. The sample was centrifuged for 2 minutes at 14,000 rpm and the supernatant transferred to a fresh Eppendorf. Two volumes of ice-cold 100% Ethanol were added to the supernatant, vortexed and incubated for 2 minutes at room temperature. The sample was centrifuged for 10 minutes at 14,000 rpm, the supernatant removed and the Eppendorf inverted on a paper towel to allow the liquid to drain away. To the dry pellet, 500 µl of ice-cold 70% Ethanol was added. The sample was vortexed briefly and centrifuged for 5 minutes at 14,000 rpm. The supernatant was removed and the pellet was dried at room temperature. The pellet was re-suspended in 20 µl of TE buffer (pH 8.0) containing 0.4 µg DNase-free pancreatic RNase (Merck Biosciences, Nottinghamshire, U.K.). The sample was vortexed briefly and incubated at 37°C for 15 minutes. The DNA was stored at -20°C.

Purified plasmid DNA was quantified using a UniCam UV 500 spectrophotometer and Vision 32 Software (both from Thermo Spectronic, New

York, U.S.A.). The absorbance of the sample at 260 nm and 280 nm was measured. The quantity of DNA was determined using the absorbance at 260 nm where 1 optical density unit corresponded to a DNA concentration of approximately 50 µg/ml. The sample was checked for protein contamination by calculating the ratio of the readings at 260 nm and 280 nm. A ratio of between 1.6 and 2.0 was obtained for good quality DNA; anything below this was indicative of contamination with protein (Sambrook et al., 1989).

2.4.9.4 Quantitative real-time PCR

PCR reactions were set up on 96-well plates using the CAS3200 Robotic Liquid Handling System and the associated Corbett Robotics v4.0 software (Corbett Life Science, New South Wales, Australia). Target reference gene expression was quantified using a Stratagene Mx4000 Multiplex Quantitative PCR System and the associated v.4.20 software (Stratagene, California, U.S.A.). Reactions were carried out in a total volume of 25 µl using Brilliant® SYBR® Green QPCR Master Mix (Stratagene, California, U.S.A.) with primer concentrations of 0.2 µM as described in Love et al. (2005). Each reaction was performed on 5 µl of a 1/10 dilution of cDNA. A standard curve was produced from DNA purified from plasmids as described above. A five-fold dilution series of purified plasmid DNA was produced from an initial standard quantity of 20 ng.

PCR conditions were a denaturing step of 95°C for 5 minutes; 40 cycles of a denaturing step of 95°C for 30 seconds, an annealing temperature step for 30 seconds and an elongation step of 72°C for 30 seconds. A dissociation curve was produced to ensure that detected SYBR-Green fluorescence was due to the labelling of specific PCR products and not the formation of primer-dimers. Dissociation curve conditions were 1 cycle of 95°C for 1 minute followed by 40 cycles of 30 seconds each of 1°C increments from 55°C.

PCR plates were set up so that parallel reactions were carried out in duplicate for each sample well. Reactions for the production of a standard curve were also carried out in duplicate. No template controls (NTC) containing reaction mix only were included on the PCR plate. The PCR cycle where the measured SYBR-Green fluorescence signal was determined to be above that of the

background at a statistically significant level was termed the threshold cycle (Ct). The Ct values produced for each sample were used to determine copy numbers of the cDNA targets. A standard curve was produced by relating the known initial amount of target in each standard curve reaction to the corresponding Ct value. The standard curve was used to determine the initial target quantity in the samples based on the Ct values obtained for each sample. In order to confirm that the observed fluorescence for each sample was due to amplified product, the dissociation curves were analysed to determine if any other peaks appeared other than the amplicon peak. The absence of additional peaks indicated that no significant primer-dimer formation had occurred. The gene intensities for the reference genes *RPS9* and *RPL32* were normalised relative to *Actin2* intensity values for both the quantitative PCR and microarray values and compared with each to ensure consistency between the measurement values.

2.4.10 Metabolite extraction

Metabolites were extracted based on the integrated RNA, metabolite and protein extraction protocol detailed in section 2.4.2.1 (Weckwerth et al., 2004). Only the steps relevant to the extraction of metabolites are detailed here. A pestle and mortar were pre-cooled with liquid nitrogen and the sample tissue was homogenised under liquid nitrogen. To the sample, 2 ml of a single phase solvent mixture of methanol/ chloroform/ dH_2O (2.5:1:1 v/v/v), stored at -20°C, was added and the sample was mixed at 4°C for 30 minutes. The sample was centrifuged at 14,000 rpm at 4°C for 10 minutes. The supernatant was removed to a new Eppendorf and chilled on ice. To the remaining pellet, 1 ml of a single phase solvent mixture of methanol/ chloroform (1:1 v/v), stored at -20°C was added and the sample vortexed. The sample was centrifuged at 14,000 rpm at 4°C for 10 minutes and the supernatant removed to a new Eppendorf and chilled on ice. The two supernatants collected were combined and 500 µl dH_2O was added to separate the hydrophobic phase from the hydrophilic phase. The sample was centrifuged at 14,000 rpm at 4°C for 5 minutes and the two phases separated using a pipette. The separated phases were stored at -80°C for no longer than two

weeks. Samples were dehydrated and stored under desiccation at room temperature during transit.

2.4.11 ^{1}H-NMR quantification of metabolites

Samples were re-hydrated in a total volume of 4 mL of 5 mM EDTA, pH 8.0 immediately prior to NMR analysis. A number of primary metabolites were quantified for each time point using ^{1}H-NMR (Moing et al., 2004). NMR metabolite analytical methods are generally less sensitive than MS methods for the detection and quantification of metabolites; however, several techniques have been developed for the purpose of metabolic fingerprinting (Belton et al., 1998; Noteborn et al 2000; Ward et al., 2003). A total of 27 peaks were identified from the ^{1}H-NMR spectra corresponding to a variety of known and unknown metabolites including soluble carbohydrates, amino acids and organic acids. The quantification of the identified metabolites was verified by generating a mix of 5 mg of each of eight samples selected at random. Based on the values obtained for each of the eight samples, an expected value was calculated for each metabolite and compared with the metabolite levels measured in the mix of eight samples. The measured values in the random mix corresponded very closely to the values expected from the previously measured metabolite levels, demonstrating the high quality of the metabolite data measurements.

Chapter 3

Consequences of disrupting acetate metabolism in Arabidopsis seedlings

3.1 Introduction

Seed germination and the subsequent establishment of photosynthetically competent seedlings represents a unique stage in plant development as a shift from heterotrophic to predominantly autotrophic metabolism occurs. Immediately post-germination and prior to the establishment of photosynthetic organs, development is dependent on the nutrient reserves contained within the seed itself. In Arabidopsis seeds as much as 40% of storage content is stored as oil in the cotyledons. During this period of germination and seedling establishment fatty acids are released from triacylglycerol stores through lipase activity and activated to form acyl Co-A esters. β-oxidation activity in the glyoxysome breaks down the acyl Co-A esters to acetyl Co-A, the activated form of acetate (Eastmond and Graham, 2001). Acetyl Co-A is incorporated into the glyoxylate cycle and ultimately converted to sucrose thereby supporting seedling growth until the seedling is established as a self-sufficient photoautotrophic organism (Beevers, 1961, Crawford, 1977).

In its activated form of acetyl Co-A, acetate is involved in many other aspects of plant metabolism as the catabolic product of a number of compounds including fatty acids, amino acids and glucose and may be metabolised to form many different cellular compounds (Hooks et al., 2004). In addition, acetate has been shown to affect gene expression in a number of organisms including bacteria, yeast and plants (Sheen, 1990; Graham et al., 1992; Hartig et al., 1992; Oh et al., 2002). A possible role in the regulation of the transition from heterotrophic to autotrophic metabolism in seedlings has also been suggested for acetate (Sheen, 1990). Furthermore, free acetate has served as an invaluable tool in order to introduce isotopically labelled carbon into certain metabolic pathways.

The glyoxylate cycle, by which acetyl-CoA from β-oxidation is converted into organic acids, was elucidated by feeding studies with radiolabelled acetate (Canvin & Beevers, 1961).

A series of acetate non-utilisation mutants (*acn*) were identified based on seedling establishment on monofluoroacetic acid, a toxic analogue of acetate (Hooks et al., 2004). The *acn1* mutant was identified as a acetate non-utilising mutant; the exogenous application of [^{14}C]acetate showed that general metabolism was uninterrupted but seedlings were compromised in their ability to incorporate label into soluble carbohydrates. ACN1 was identified as a short-chain acyl-CoA synthetase responsible for the activation of acetate to acetyl CoA for entry into the glyoxylate cycle (Turner et al., 2005). The toxic acetate analogue fluoroacetate was used to screen for EMS-mutagenized Arabidopsis mutants and a mutant was isolated, *acn1-1*. A tDNA-mutagenized population was screened and a further two fluoroacetate resistant mutants were identified with highly similar characteristics to *acn1-1*. Reciprocal crosses of the tDNA mutants to the EMS mutant *acn1* showed that the mutants were allelic. Therefore, the EMS mutant allele was referred to as *acn1-1* and the tDNA-tagged alleles as *acn1-2* and *acn1-3*. Southern blotting of genomic DNA using probes specific to an unque EcoRI site within the tDNA insert was used to confirm that only one tDNA insert was present in *acn1-2* and *acn1-3* (Turner et al., 2005).

The *acn1-2* mutant was used to investigate the effects of disrupting acetate metabolism in Arabidopsis seedlings. As ACN1 lies outside of the main heterotrophic flow of carbon, it is unclear what function it has and what the consequences of its elimination are. It is apparent that under optimal conditions growth of *acn1-2* is not compromised but subtle effects in metabolism and gene expression are induced. Microarrays were used to produce global gene expression profiles for *acn1-2* and the corresponding wild type seedlings, Col-7. The first objective toward ascertaining the importance of the function of ACN1 in developing seedlings was to determine the effects of disrupting the activation step upon metabolite levels and gene expression profiles.

3.2 Metabolite profiling of the *acn1-2* mutant

3.2.1 Global acetate consumption appears normal in *acn1-2*

Seedlings of the *acn1-2* mutant have been shown to be less tolerant of acetate following germination and growth on agar plates containing acetate than related *acn* mutants (Hooks et al., 2004). Disrupting the conversion of acetate to acetyl CoA might result in an accumulation of acetate which would lead to acidification of the cytosol and ultimately death of the seedlings. Accordingly, the relatively high tolerance of *acn1-2* to germination and growth on butyrate compared to acetate indicated that butyrate may be activated by another enzyme and again demonstrated the importance of ACN1 in converting acetate to acetyl CoA. It is known that acetate metabolism is not completely abolished in *acn1-2*; [2-^{14}C] acetate feeding studies on similarly staged seedlings showed that non-carbohydrate compound classes incorporated label at levels close to those observed in Col-7 except for the ethanol insoluble and organic acid fractions (Turner et al., 2005). This suggested that acetate metabolism continued to some degree although it remained unclear whether it was reduced in *acn1-2*. The organic acid fraction obtained from *acn1-2* showed relatively higher levels of label which may have come from higher levels of un-metabolised free acetate, indicating a reduction in acetate assimilation. The uptake and fate of acetate by *acn1-2* was compared to Col-7 using NMR to quantify the resonance corresponding to the C2 carbon of [2-^{13}C]acetate-fed seedlings (figure 3.1). The feeding experiment was performed with additional sucrose as seedlings fed acetate alone took up very little label, showing a barely detectable ^{13}C signature after even 8 h, and also began to disintegrate during harvest. In the presence of sucrose, the uptake of [2-^{13}C]acetate was increased and seedling integrity was preserved. The degree of physical deterioration at any given time was similar between *acn1-2* and Col-7 during the incubation times (up to 24 hours).

The levels of labelled acetate remained consistent between *acn1-2* and Col-7. Labelled acetate increased from undetectable levels to a peak of 40-50 μmol g^{-1} FW within 2 hours and remained relatively constant at levels between 30-40 μmol g^{-1} FW for up to 24 h (data not shown). In order to determine whether

acetate was metabolised in *acn1-2*, the ^{13}C acetate was chased by isotopically normal acetate and the fate of labelled acetate followed for a series of samples. The levels of labelled acetate were decreased within two hours to approximately 10 μmol g^{-1} FW. In the subsequent two hours a small decrease in levels was observed.

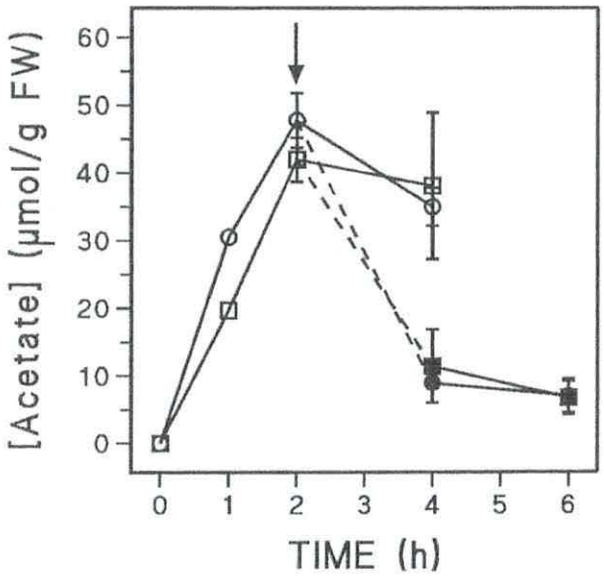


Figure 3.1: Acetate utilisation by *acn1-2*. Symbols: □, ■ Col-7; ○, ● *acn1-2*. Darkened symbols represent acetate levels after [2-^{12}C]acetate chase. Seedlings harvested at principal growth stage 1.0 were fed with 4 mM [2-^{13}C]acetate plus 20 mM sucrose at the times indicated. The chase experiments were performed by removing seedlings at 2 hours and transferring them to a solution with 4 mM isotopically normal acetate plus 20 mM sucrose (indicated by an arrow on the chart). Symbols and error bars represent the averages ± Standard deviation, respectively of values from 3 independent experiments. Only one measurement was done for the 1 h and 24 h time points.

3.2.2 Disruption of the flow of carbon from acetate into glutamine

The fate of the labelled [2-^{13}C]acetate was followed to monitor the appearance of label in other metabolites. The predominant sink for labelled carbon was glutamine where label was observed in all three carbons of the side chain (figure 3.2 A). However, the amount of label incorporated into all three carbons was reduced by approximately 50% in *acn1-2* after 4 hours. The levels of the three labelled carbon atoms of glutamine were measured in the same extracts used for acetate quantification (figure 3.2 B-D). Signals corresponding to the C3 and C4 carbons of the side chain of glutamine were initially observed in extracts of unfed seedlings for both *acn1-2* and Col-7. In the wild type seedlings a substantial increase in the amount of label in C3 and C4 from 3 to 8 μmol g^{-1} FW, 4 h after feeding while little if any appearance of label was observed in C3 and C4 carbons in the mutant. Initially, the C2 label was not apparent, but over time, a signal was measured which increased to approximately 4 μmol g^{-1} FW after 4 hours of feeding. The intensities of the measured signals for each carbon remained stable for up to 24 hours (data not shown). A decrease of around 20% in intensity was observed for each carbon, probably due to a decrease of acetate in the media. The ^{13}C acetate was chased by isotopically normal acetate at two hours and an increase in glutamine levels was observed over the subsequent two hours, followed by a decrease in levels.

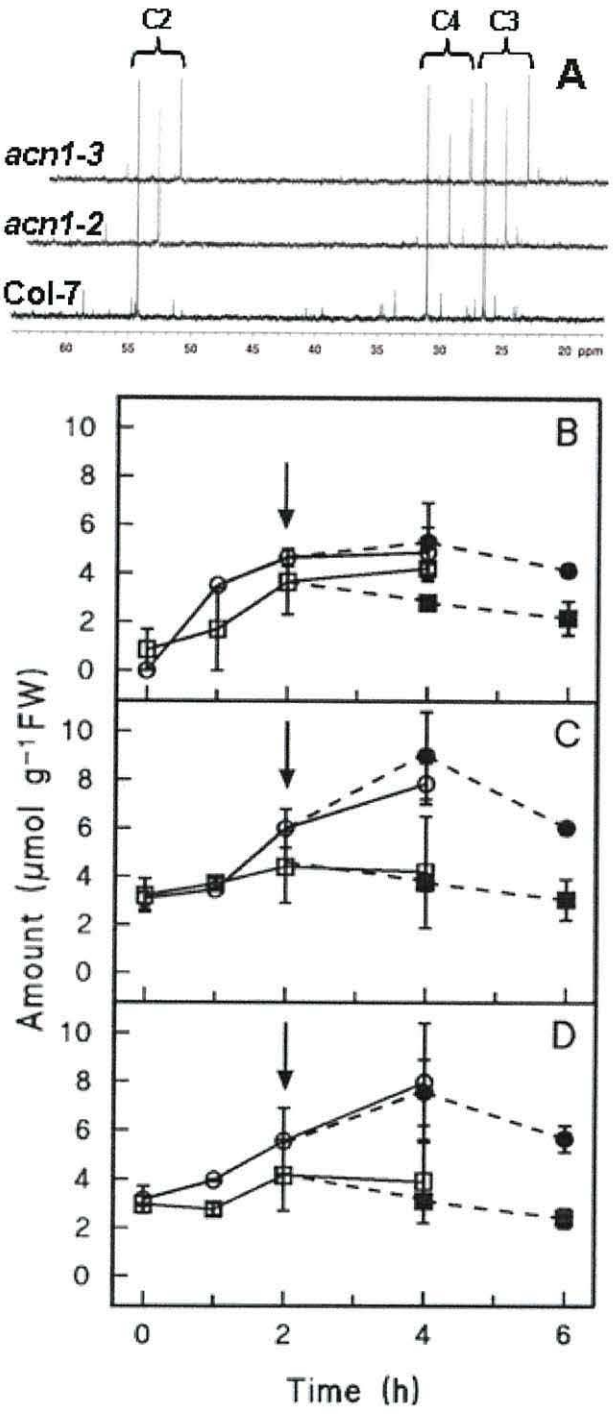


Figure 3.2: Quantification of ^{13}C-labelled glutamine. Symbols: □, ■ Col-7; ○, ● *acn1-2*. Label was observed in all three carbons of the side chain of glutamine (A). The levels of the three labelled carbons of the side chain of glutamine were determined (B-D). Darkened symbols represent C3-Gln levels after the [2-^{12}C]acetate chase (indicated by the position of the arrow in each chart). The values represent the average and ranges of two independent experiments.

3.2.3 Depression of major metabolite levels in *acn1-2*

Seedlings in this experiment were harvested at principal growth stage 1.0 (Boyes et al., 2001) at which point stored carbohydrate and lipid reserves are still being mobilised as a carbon source, but seedlings have developed some photosynthetic capacity. Although a decrease of label in individual compounds may be interpreted as a depression in metabolism, this may not be a general effect. Consequently, ^{1}H-NMR was used to profile a number of metabolites to investigate the general effects of the *acn1-2* mutation on metabolism in seedling development (Moing et al., 2004). A total of 28 distinct resonances were quantified corresponding to a variety of known and unknown metabolites including soluble carbohydrates, amino acids and organic acids (table 3.1). The levels of acetyl-CoA were determined using HPLC (Larson et al., 2001). In general, the absolute levels of amino acids quantified on a nmol/g FW basis were of the same order as those observed in seeds of the Arabidopsis ecotype Wassilewskija (Li et al. 2003) or in seed germination series for other species, such as the legumes soybean and lupin (Martínez-Villaluenga et al. 2006). Most notably, the amounts of glutamine quantified by ^{13}C-NMR and ^{1}H-NMR methods corresponded closely. Measurements by ^{13}C labelling quantified glutamine levels at approximately 8000 nmol g^{-1} FW while ^{1}H-NMR profiling showed levels of glutamine of around 9000 – 10,000 nmol g^{-1} FW. The comparable levels indicate that glutamine labelling had reached a steady-state by 4 hours after feeding with labelled acetate and that it was likely that labelled acetate had also attained steady-state levels within this time. Therefore, measurements of acetate from the two methods may be compared directly and it may be concluded that wild-type Arabidopsis seedlings have the capacity to accumulate acetate (from an exogenous source) 1000 times greater than observed in untreated seedlings.

Table 3.1: Comparison of the major metabolite levels in *acn1-2* and Col-7. ^{1}H-NMR was used to profile a number of metabolites to investigate the general effects of the *acn1-2* mutation on metabolism in seedling development

Metabolite*	Col-7	*acn1-2*	P†
	(nmol g^{-1} FW)		
Alanine	275 ± 84	106 ± 66	0.0036
Arginine	383 ± 287	119 ± 67	0.0541
Glutamate	2430 ± 1869	600 ± 337	0.0469
Glutamine	9520 ± 5100	2543 ± 1298	0.0164
Isoleucine	30 ± 14	12 ± 4	0.0234
Leucine	158 ± 60	66 ± 26	0.0095
Proline	268 ± 293	55 ± 47	0.0906
Threonine	413 ± 225	123 ± 48	0.0216
Valine	103 ± 43	36 ± 15	0.0111
Choline	471 ± 202	163 ± 95	0.0114
Trigonelline	24 ± 12	9 ± 9	0.0199
Fructose	1165 ± 601	465 ± 196	0.0290
Glucose	1316 ± 573	522 ± 318	0.0163
Rhamnose	29 ± 20	9 ± 7	0.0436
Sucrose	507 ± 215	193 ± 98	0.0135
Citrate	220 ± 100	135 ± 59	0.0729
Formate	160 ± 80	48 ± 21	0.0164
Fumarate	64 ± 57	75 ± 68	0.3979
Lactate	198 ± 113	68 ± 10	0.0307
Malate	465 ± 220	225 ± 128	0.0366
Acetate	45 ± 6	16 ± 8	0.0011
Acetyl-CoA‡	38 ± 35	88 ± 18	0.0200
	Relative Units		
Unknown D8.0	10 ± 6	6 ± 6	0.1636
Unknown Q7.9	18 ± 11	6 ± 6	0.0280
Unknown S7.37	22 ± 8	9 ± 3	0.0099
Unknown D5.69	93 ± 52	36 ± 12	0.0346
Unknown Q5.18	7 ± 3	3 ± 3	0.0244
Unknown D3.12	233 ± 120	116 ± 33	0.0407
Unknown M1.85	108 ± 61	44 ± 27	0.0309

*The letters and numbers of unknown peaks represent the type of signal and the chemical shift relative to TMS: D, doublet; M, multiplet; Q, quintuplet; S, singlet.

† The value is the probability based on the statistic from Student's T-tests with 9 degrees of freedom.

‡ Determined by HPLC according to Larson et al.

Metabolite levels were significantly decreased ($p < 0.05$) in *acn1-2* for all metabolites measured except fumarate, which demonstrated a small but insignificant increase in the mutant (table 3.1). The free amino acid content of the mutant was on average 33% of that observed in Col-7, ranging from 80% less proline to 68% less leucine. Organic acid levels in *acn1-2*, excluding fumarate, were approximately 42% of those in Col-7 and ranged from 70% less formate to 40% less citrate. The levels of soluble sugars in *acn1-2* were on average 37% of those in the wild-type. The unknown compounds were affected similarly to the other classes with levels decreasing between 40 and 60%.

3.3 Transcriptome profiling of the *acn1-2* mutant

3.3.1 Differentially expressed genes between *acn1-2* and Col-7

A total of 201 genes were identified as differentially expressed between *acn1-2* and Col-7 as detailed in chapter 2, section 2. A larger number of genes were repressed rather than induced in *acn1-2*; 130 genes showed a decrease and 71 genes showed an increase in *acn1-2* compared to Col-7. Differences in the observed fold-changes in expression levels were also seen; repressed genes showed a decrease in expression from 3.8 to a maximum of 52.8-fold while induced genes showed lower changes in expression in the range of 1.8 to 27.7-fold increases.

Due to the large amounts of data generated by global gene expression profiling experiments, it is useful to summarise information, or examine subsets of data to facilitate meaningful analysis. It has been suggested that the genes with the largest changes in expression levels are the most biologically relevant to identify differences between experimental samples, although this assumption does not take into account situations where a relatively small change in gene expression levels affects the expression in a number of other genes, in a cascade-effect. However, a large change in gene transcript levels often corresponds to a gene of importance (Breitling et al., 2004). In order to identify the genes most affected by the disruption to acetate metabolism in the *acn1-2* mutant, the top 20

up and top 20 down regulated genes were identified between *acn1-2* and Col-7 (table 3.2) and analysed using the gene ontology annotation tool (available at http://www.arabidopsis.org/tools/bulk/go/index.jsp, Berardini et al., 2004).

Table 3.2: The top twenty up and down regulated genes in *acn1-2* compared to Col-7. MIPS numbers are provided as gene identifiers and the corresponding fold changes in expression are listed (continued on page 82).

Genes up-regulated in *acn1-2* compared to Col-7		
Transcript	Gene Title	Fold change
At4g31070	pentatricopeptide (PPR) repeat-containing protein	27.7
At3g49460	60S acidic ribosomal protein-related	23.9
At1g49920	Zinc finger protein-related	23.7
At1g56680	glycoside hydrolase family 19 protein	21.2
At5g13350	auxin-responsive GH3 family protein	19.6
At1g22080	hypothetical protein	16.7
At1g27590	expressed protein	16.0
At5g19920	Transducin family protein / WD-40 repeat family protein	14.0
At2g04620	cation efflux family protein	14.0
At2g46780	RNA recognition motif (RRM)-containing protein	13.7
At3g28510	AAA-type ATPase family protein	12.7
At1g75230	HhH-GPD base excision DNA repair family protein	11.2
At2g17920	expressed protein	10.2
At1g66360	C2 domain-containing protein	9.4
At2g06480	hypothetical protein	9.0
At3g19330	polyadenylate-binding protein-related	9.0
At5g24655	expressed protein	8.4
At4g18780	cellulose synthase, catalytic subunit (IRX1)	8.1
At5g25950	hypothetical protein	7.5
At4g05240	ubiquitin family protein	7.5

Genes down-regulated in *acn1-2* compared to Col-7		
Transcript	Gene Title	Fold change
At2g36640	Late embryogenesis abundant protein (ECP63)	52.8
At1g33055	expressed protein	45.0
At1g72070	DNAJ heat shock N-terminal domain-containing protein	38.4
At1g48660	auxin-responsive GH3 family protein	34.4
At1g32560	Late embryogenesis abundant group1 domain-containing protein	30.6

At4g36600	Late embryogenesis abundant domain-containing protein	28.5
At2g33850	expressed protein	24.9
At4g21020	Late embryogenesis abundant domain-containing protein	24.9
At5g28630	glycine-rich protein	23.1
At1g10640	Polygalacturonase, putative / pectinase, putative	21.7
At5g62210	embryo-specific protein-related	20.1
At1g33265	expressed protein	19.1
At1g24090	RNase H domain-containing protein	17.0
At4g35480	Zinc finger (C3HC4-type RING finger) family protein	16.5
At3g45460	hypothetical protein	16.0
At4g36830	GNS1/SUR4 membrane family protein	15.5
At3g02480	ABA-responsive protein-related	15.1
At4g20800	FAD-binding domain-containing protein	14.9
At2g32790	ubiquitin-conjugating enzyme, putative	14.6
At5g39110	germin-like protein, putative	14.3

Analysis of the ontology of the genes induced in *acn1-2* compared to Col-7 showed that 65% had no assigned ontology in TAIR and 20% were involved in stress response. A gene associated with a hormone regulator, auxin was induced (At5g13350) along with a number of genes involved in the regulation of translation and transcription (At3g49460, At1g49920 and At5g19920).

Of the twenty genes with the largest decreases in expression levels sixteen had an assigned identity in the TAIR database, eight of which were involved with early seedling development. Analysis with the TAIR gene ontology tool showed that the majority of genes (65%) had no assigned ontology and the remaining genes did not appear enriched in a specific ontology. A number of genes associated with development showed a decrease in expression levels in *acn1-2* including genes encoding late embryogenesis abundant (LEA) proteins (At2g36640, At1g32560, At4g36600, At4g21020), seed storage proteins (At5g39110) and proteins associated with embryonic development (At5g62210). Genes associated with hormone regulators of plant development (auxin, At1g48660 and abscisic acid, At3g02480) and a transcription factor (At4g35480) were also repressed in *acn1-2*. A number of genes with no assigned identity were also among the top twenty repressed genes in *acn1-2*.

3.3.2 Clustering of genes differentially expressed between *acn1-2* and Col-7

A variety of clustering algorithms are available to visualise the behaviour of large datasets generated by high throughput analytical techniques, such as microarrays. Features and relationships inherent to the dataset are used to organise the data, facilitating the identification of patterns and related gene expression profiles (Stekel, 2003). The 201 genes differentially expressed between *acn1-2* and Col-7 were analysed using a combination of hierarchical and 2D-SOM (2-Dimensional Self Organising Map) clustering.

Hierarchical clustering algorithms are commonly used to identify relationships within a dataset. The application of hierarchical clustering does not require a user-specified number of clusters for the data to form, allowing this clustering technique to be applied in an exploratory manner to determine how many clusters are present in a dataset. Hierarchical clustering was applied to the 201 differentially expressed genes between *acn1-2* and Col-7 in order to determine the number of groups formed by the data (figure 3.3).

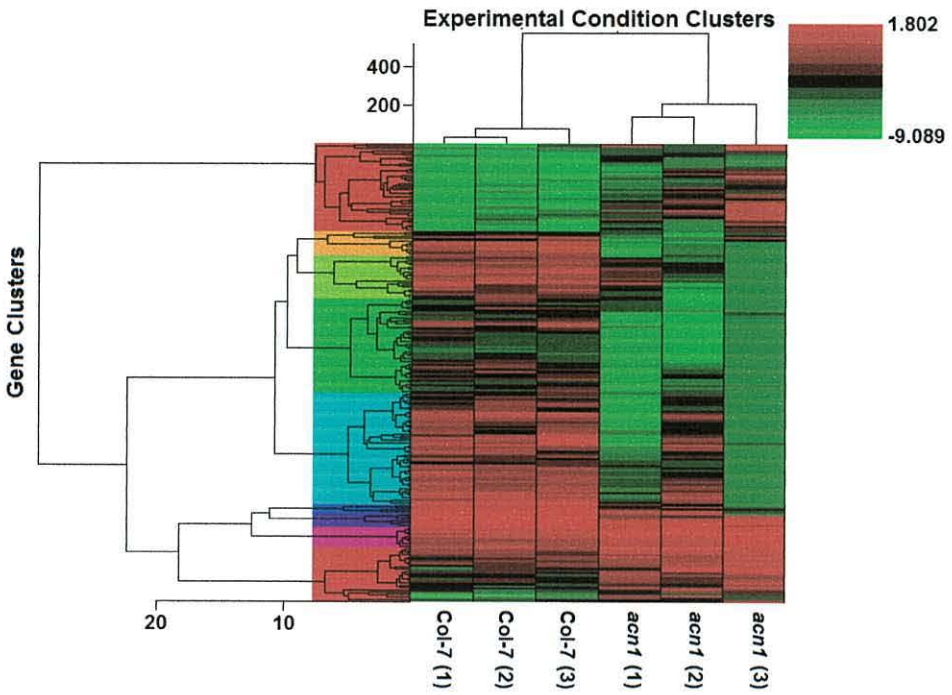


Figure 3.3: Hierarchical clustering of genes and experimental replicates based on the 201 genes differentially expressed between *acn1-2* and Col-7. The scale bars to the bottom and left of the clusters represents the distance between clusters and cluster components. The colour scale chart is representative of relative gene expression levels. Coloured blocks are used to highlight gene clusters, showing that the differentially expressed genes form eight or nine groupings.

Approximately eight or nine groupings were formed by the genes differentially expressed between *acn1-2* and Col-7. The number of clusters is determined by the user and as such is open to interpretation but was supported by the use of alternative clustering algorithms (data not shown). The behaviour of the differentially expressed genes was further examined using a 2D-SOM clustering algorithm. A 2D-SOM presents a dataset partitioned into a user-defined number of clusters with clusters arranged in a two-dimensional grid format where each cluster contains a set of genes that show similar expression patterns. Clusters are further arranged so that clusters showing similar behaviour are sited closer to each other than those that are not related. A gene is initially selected at random and assigned to a cluster. Additional genes are examined for related behaviour and placed with genes with similar expression patterns (Tamayo et al., 1999). The set of 201 genes identified as differentially expressed between *acn1-2* and Col-7 was clustered using a 2D-SOM with a total of nine

clusters, based on the results of hierarchical clustering of the differentially expressed genes (figure 3.4). Although hierarchical clustering indicated that a 2D-SOM composed of an eight component grid, the limitations of the software necessitated the use of a 3 x 3 grid. For subsequent analysis the genes in clusters 4 and 5 were grouped together as the genes displayed similar expression profiles and were observed to cluster together in repeated applications of the 2D-SOM algorithm.

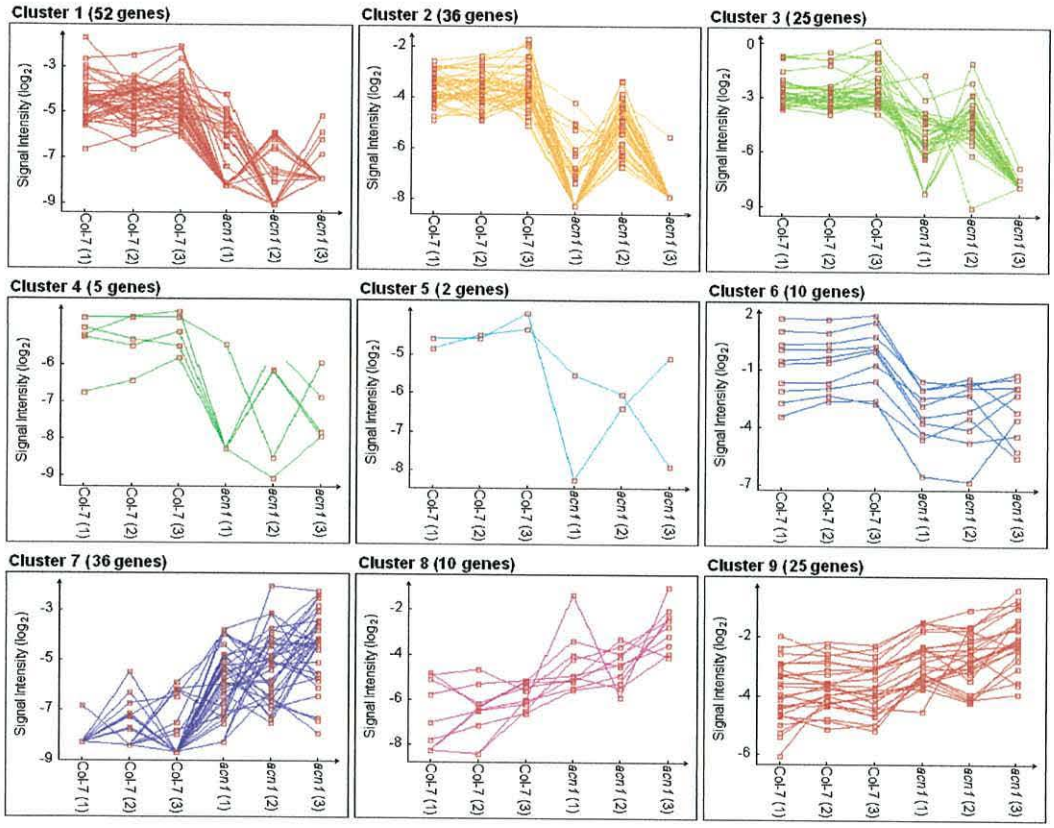


Figure 3.4: 2D-SOM clustering of the 201 genes identified as differentially expressed between *acn1-2* and Col-7. The signal intensity (log_2) of the genes is displayed for each of the *acn1-2* and Col-7 microarray replicates.

A brief description of the genetic complement of each cluster is provided in table 3.3. Gene descriptors available from TAIR were used to identify groups of genes in each cluster. A full list of the genes found in each cluster is given in Appendix B.

Table 3.3: Gene descriptions for the genetic complement of each of the clusters identified through 2D-SOM clustering of differentially expressed genes between *acn1-2* and Col-7. A brief description of gene type and the number of occurrences are listed (table continued on pages 87,88 and 89).

Down-regulated in *acn1-2* compared to Col-7		
Cluster	**Description of gene function**	**Number**
1	---	2
(52 genes)	expressed protein	8
	hypothetical protein	5
	1-aminocyclopropane-1-carboxylate oxidase, putative / ACC oxidase, putative	1
	annexin 6 (ANN6)	1
	auxin-responsive protein	2
	beta-expansin, putative (EXPB4)	1
	calcium-binding EF hand family protein	1
	cation/hydrogen exchanger, putative (CHX4)	1
	cupin family protein	1
	cyclin, putative (SDS)	1
	cytochrome P450 71B16, putative (CYP71B16)	1
	DEAD/DEAH box helicase, putative	1
	dehydration-responsive protein-related	1
	DNA topoisomerase II family protein	1
	DNAJ heat shock N-terminal domain-containing protein	1
	embryo-specific protein-related	1
	FAD-binding domain-containing protein	1
	germin-like protein, putative	1
	glycine-rich protein	1
	kelch repeat-containing F-box family protein	1
	kinase interacting family protein	1
	late embryogenesis abundant domain-containing protein	4
	MADS-box protein	2
	myb family transcription factor (MYB35)	1
	pentatricopeptide (PPR) repeat-containing protein	2
	polygalacturonase, putative / pectinase, putative	1
	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	1
	protein kinase family protein	2
	RNase H domain-containing protein	1
	terpene synthase/cyclase family protein	1
	transferase family protein	1
	wall-associated kinase, putative	1
2	aminotransferase class IV family protein	1
(36 genes)	DC1 domain-containing protein	1
	eukaryotic translation initiation factor-related	1
	expressed protein	7

	GNS1/SUR4 membrane family protein	1
	hypothetical protein	5
	leucine-rich repeat family protein	3
	myb family transcription factor (MYB43)	1
	myrcene/ocimene synthase, putative	1
	nodulin family protein	1
	pentatricopeptide (PPR) repeat-containing protein	2
	protease inhibitor/seed storage/lipid transfer protein family protein	2
	protein kinase family protein	3
	proton-dependent oligopeptide transport (POT) family protein	1
	serine carboxypeptidase S28 family protein	1
	sugar transporter, putative	1
	transcriptional factor B3 family protein	1
	ubiquitin-conjugating enzyme, putative	1
	vegetative storage protein 1 (VSP1)	1
	zinc finger (C3HC4-type RING finger) family protein	1
3	17.7 kDa class II heat shock protein 17.6A (HSP17.7-CII)	1
(25 genes)	ankyrin repeat family protein	1
	armadillo/beta-catenin repeat family protein / U-box domain-containing protein	1
	cation exchanger, putative (CAX8)	1
	DC1 domain-containing protein	2
	expressed protein	3
	Fe-S metabolism associated domain-containing protein	1
	histone H1-3 (HIS1-3)	1
	hydrolase, alpha/beta fold family protein	1
	hypothetical protein	1
	jacalin lectin family protein	2
	lipase class 3 family protein	1
	N-acetyltransferase, putative / hookless1 (HLS1)	1
	palmitoyl protein thioesterase family protein	1
	pathogenesis-related thaumatin family protein	1
	pentatricopeptide (PPR) repeat-containing protein	2
	peroxidase-related	1
	serine carboxypeptidase S10 family protein	1
	sterol desaturase family protein	1
	zinc finger (AN1-like) family protein	1
4	chaperonin, putative	1
(5 genes)	cytochrome P450 family protein	1
	expressed protein	3
5	pentatricopeptide (PPR) repeat-containing protein	1
(2 genes)	receptor-like protein kinase-related	1
6	ABA-responsive protein-related	1
(10 genes)	AWPM-19-like membrane family protein	1
	expressed protein	3
	late embryogenesis abundant group 1 domain-containing	1

	protein	
	leucoanthocyanidin dioxygenase, putative / anthocyanidin synthase	1
	plant defensin-fusion protein, putative (PDF1.1)	1
	rubber elongation factor (REF) protein-related	1
	short-chain dehydrogenase/reductase (SDR) family protein	1

Up-regulated in *acn1-2* compared to Col-7

Cluster	Description of gene function	Number
7	---	3
(36 genes)	expressed protein	7
	hypothetical protein	6
	pentatricopeptide repeat-containing protein	2
	peroxidase	1
	60S acidic ribosomal protein-related	1
	AAA-type ATPase family protein	1
	C2 domain-containing protein	1
	cation efflux family protein	1
	CBL-interacting protein kinase 13 (CIPK13)	1
	cellulose synthase, catalytic subunit (IRX1)	1
	glycine-rich protein	1
	glycoside hydrolase family 19 protein	1
	HhH-GPD base excision DNA repair family protein	1
	MADS-box family protein	1
	polyadenylate-binding protein-related / PABP-related	1
	protein phosphatase 2C, putative / PP2C, putative	1
	replication protein-related	1
	RNA recognition motif (RRM)-containing protein	1
	transducin family protein / WD-40 repeat family protein	1
	ubiquitin family protein	1
	zinc finger protein-related	1
8	---	3
(10 genes)	hypothetical protein	2
	apical meristem formation protein-related	1
	auxin-responsive GH3 family protein	1
	cytochrome P450 71A13, putative (CYP71A13)	1
	disease resistance protein (TIR-NBS-LRR class), putative	1
	pentatricopeptide (PPR) repeat-containing protein	1
9	---	1
(25 genes)	expressed protein	3
	hypothetical protein	4
	ankyrin repeat family protein	1
	DC1 domain-containing protein	1
	DNA-directed RNA polymerase family protein	1
	hexose transporter, putative	1
	homeobox transcription factor, putative	1

leucine-rich repeat family protein	2
microtubule associated protein (MAP65/ASE1) family protein	1
phospholipase D, putative (PLDZETA)	1
plastocyanin-like domain-containing protein	1
protein kinase family protein	1
splicing factor family protein	1
sulfate transporter	1
Toll-Interleukin-Resistance (TIR) domain-containing protein	1
transcriptional regulator (FUSCA3)	1
tyrosine specific protein phosphatase family protein	1
zinc-binding family protein	1

Analysis of the biological function ontology for the gene clusters using the online search facilities available from TAIR (Beradini et al., 2004) shows that a variety of genes are found in each cluster. As seen previously, a large prportion of the genes in each cluster were not annotated with biological function ontology. Of the genes with assigned biological ontologies, Cluster 1 was compromised of genes involved in developmental processes, protein metabolism, nucleic acid related and signal transduction. Cluster 2 contained genes for protein metabolism, signalling and transcription. Cluster 3 contained genes involved in responses to stress or external stimuli, development, signalling and transcription. Genes from clusters 4 and 5 were examined jointly as these genes clustered together intermittently upon repeated 2D-SOM clustering, but no enrichment of a particular type of gene was noted. Similarly, genes in cluster 6 did not show a bias towards any specific function or process. The lack of enrichment in certain clusters highlights the need for further work on assigning functions and annotations to genes in existing databases. The up-regulated clusters were also examined for enrichment of genes of a specific ontology. Analysis of cluster 7 showed that the majority of genes were involved in response to stress or external stimuli, signalling and nucleic acid related. The genes of cluster 8 did not show enrichment in any specific ontology. Analysis of cluster 9 showed the genes were mainly involved in transcription or development.

3.3.3 Promoter analysis of co-expressed genes

Following the simplified assumption that co-expressed genes share common regulatory mechanisms and that these mechanisms are active at the level of the promoter, the sequences of the promoters of co-expressed genes identified in the 2D-SOM displaying differentially expressed genes between Col-7 and *acn1-2* (figure 3.4) were analysed for the presence of common transcription factor binding sites as detailed in chapter 2, section 2. Intergenic sequences of up to 1500 bases upstream from the transcription initiation site were obtained using the sequence retrieval tool available from the regulatory sequence analysis tools (RSAT) website and translated into FASTA format (http://rsat.ulb.ac.be/rsat/, van Helden et al., 2000). The sequence scanning MotifSampler software was used in this work (chapter 2, section 2). The outputs from MotifSamper were assigned three scores: consensus score, information content and log-likelihood. The consensus score describes how conserved the motif was, the information content describes the difference between the motif model from the background model, and the log-likelihood score which is dependent on the strength of the motif and the total number of instances of the motif, describes the probability of the sequence not being generated by the background model.

The consensus sequences produced were compared and similar motifs grouped together. The grouped consensus sequences were aligned using the ClustalW service available from the European Bioinformatics Institute (EBI) website (available at http://www.ebi.ac.uk/clustalw/, Chenna et al., 2003) to produce a single consensus sequence. In this manner, a few consensus sequences were produced for each set of co-expressed genes (table 3.4). In order to determine which consensus sequences were more likely to be biologically relevant, the statistics associated with the motifs were examined. Motifs with high information content and log-likelihood scores, found in the majority of the sequences in each set of co-expressed promoter sequences were selected as being reliable.

The DNA sequence of a known or suspected CARE can be used to search a database of CAREs in order to identify homologous sequences. Although a number of specialised databases exist composed of *cis*-regulatory elements and promoter sequences, few are specific to plants (Rombauts et al., 2003). In order

to ensure correct identification of motifs, two plant specific databases were searched to assign identities to the motifs detected in the promoter sequences. The first database used was the PlantCARE database (plant *cis*-acting regulatory elements) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, Lescot et al., 2002), which contains the sequences of over 400 plant specific *cis*-elements obtained from the literature. The second database used was the PLACE database (plant *cis*-acting regulatory DNA elements), which also contained sequences acquired from the literature (http://www.dna.affrc.go.jp/PLACE/index.html, Higo et al., 1999). DNA homology searches were performed in both the PlantCARE and the PLACE databases as both databases rely on literature searches to update the database contents. Searching in both databases provides a more complete summary of the available information. A summary of the motifs identified in each cluster is given in table 3.4, along with identified motifs in the CARE databases.

Table 3.4: Consensus sequences of motifs found in each cluster using MotifSampler (Thijs et al., 2002). Motifs were submitted to both the PlantCARE (Lescot et al., 2002) and the PLACE (Higo et al., 1999) online *cis*-regulatory databases. The corresponding motifs found in PlantCARE ([a]) and PLACE ([b]) are listed and a brief description of the motif is given. Motif sequences are highlighted in **bold** in the sequences obtained from the databases. Where the sequences obtained from the databases were found in the reverse complement orientation, sequences are presented in *italics*. Corresponding CARES were not found in the databases for every submitted DNA motif (table continued on pages 92 and 93).

Cluster	Motif	Consensus from database	Description
1	ACTTNAAT	No match in database	-
	GTTG	***CAAC***TC [b]	Giberellin response in *O. sativa* seeds
	TCYGGTTT	***WAACCA*** [b]	MYB recognition site, dehydration responsive
	AATTCAG	No match in database	-
	AACCNG	No match in database	-
2	ATTATT	CA**AT(W)ATT**G[a]	Homoeodomain leucine zipper, regulation of development
		AATAAT [b]	Consensus sequence for plant polyadenylation signal (*Z. mays*)
	TAATAA	TG**TAATAA**TATATTTA TATT[a]	Light responsive element (*S. tuberosum*)

		TGG**TAATAA**[a]	*G. max* SEF1 binding site in seeds
	GAATCTG	***NGATT***[b]	ARR1 response regulator binding element
	CTCTGTATC	***GATA***[b]	Tissue specific light responsive element
3	AGAGAG	**AGAGAG**T[a]	GAG box light responsive element
		TC***TCTCTC***[b]	CT-rich motif gene expression enhancer
	TAAAG	**AAAG**[b]	Dof binding site, regulation of carbon metabolism (*Z. mays*)
		TAAAG[b]	StDof1 binding site, guard cell-specific gene expression (*S. tuberosum*)
	TTGATT	**NGATT**[b]	ARR1 response regulator binding element
	GGTGTC	No match in database	-
	TGCARKTG	**CANNTG**[b]	E-box of napA storage-protein gene of *B. napus* seed
		CANNTG[b]	MYC recognition site, dehydration and cold responsive
	AATCTA	**AATCTA**ATCT[a]	ATCT motif LAMP-like light responsive element
	TCAGYGTC	No match in database	-
4	ATATATAA A	**TATATAA**[a]	TATA box promoter element
	ACCTTNNG	No match in database	-
6	GCTANANG	No match in database	-
	ACACGTG	TG**ACACGTG**GCA[a]	G-box, light responsive element
		CACGTG[a]	ABRE motif, abscisic acid and drought responsive
	TCGTCNCA	No match in database	-
	GTTGTCGC	No match in database	-
	TAAAGA	**TAAAGA**TT[a]	TATA box promoter element
	CGGTCA	***TGACY***[b]	W box - wounding response
		TGAC[b]	W-box core element giberelllin and pathogenesis response
7	GAANAGC	No match in database	-
	TTTTGCTT	No match in database	-
	GTCNNTG	No match in database	-
	CATTGACT	***CAAT***[b]	Consensus sequence in *P. sativum* seed legA gene
		TTGAC[b]	W-box, salycilic acid responsive
		TGACT[b]	W-box element in isoamylase1 (Barley)
		TGACY[b]	W box, wounding response
		TGAC[b]	W-box core element

			giberelllin and pathogenesis response
8	TCGCGNGAC	No match in database	-
	ACAGTNGC	No match in database	-
	GCTTNG	No match in database	-
	TGACGT	C**TGACGT**CAG[a]	C-box, light responsive element
		acagaggg**TGACGT**ggcatca[a]	Cellcycle1b, cell cycle dependent transcription
		TGACGTCA[a]	Activator sequence, root specific expression
		TGAC[b]	W-box core element giberelllin and pathogenesis response
		TGACGT[b]	Enhancer of alpha amylase in cotyledons of germinated seeds (*V. mungo*)
		ACGTCA[b]	Regulation of H3 and H4 genes (Wheat)
		TGACG[b]	Auxin and/or salicylic acid responsiveness
		ACGT[b]	Dehydration responsive
	CCGANCGA	No match in database	-
9	CGANNAC	No match in database	-
	AAGAA	No match in database	-
	GAGAGA	TC***TCTCTC***[b]	CT-rich motif, gene expression enhancer
	GATTGATT	N**GATT**[b]	Response regulator (ARR1) binding element
	TCTGANTT	No match in database	-
	GCTTNNAGC	No match in database	-
	TCAACTCT	**CAACTC**[b]	Giberellin response (*O. sativa*)
		GANTTNC[b]	Myb binding site
	TTGATCTT	No match in database	-
	TCTTNNT	TCATA**TCTTCTT**CAG[a]	sbp-CMA1 light responsive element

A large proportion of the identified DNA motifs in the promoter sequences showed an involvement with seed and seedling development, as might be expected. Core promoter elements such as the TATA box were found in the promoter sequences of some of the clusters. Some of the DNA motifs identified in the promoter sequences for each cluster did not match any of the CAREs in either the PlantCARE or the PLACE database. These motifs provide an

opportunity to identify further potential regulatory mechanisms for the expression of genes affected by the disruption in acetate metabolism.

3.4 Transcript levels of genes associated with acetate metabolism

Turner et al. (2005) identified ACN1 as a peroxisomal acetyl-CoA synthetase, responsible for the activation of free acetate to acetyl-CoA for incorporation into the glyoxylate cycle. The expression levels of a number of genes involved in the downstream metabolism of free acetate after conversion to acetyl-CoA were investigated in the *acn1-2* mutant to determine the effects of disrupting acetate metabolism (the full complement of genes is listed in Appendix C). Table 3.5 shows the top twenty genes which demonstrated the largest increases and decreases in expression levels. None of the genes showed differential expression between *acn1-2* and Col-7 when the 1.4-fold change in gene expression threshold was used in conjunction with the t-test ($P > 0.05$), however, certain genes showed differential expression when only one cut-off threshold was applied.

Cytosolic malate dehydrogenase and phosphoenolpyruvate carboxylase kinase both showed an increase in *acn1-2* when compared to Col-7 of 2.9- and 2.0- fold, respectively. Malate synthase, isocitrate lyase, malate dehydrogenase, citrate synthase, ACN1 and aspartate aminotransferase were all down regulated in *acn1-2* when compared to Col-7 in the range of 5.5- to 1.4- fold change in expression.

Table 3.5: The top twenty up and down regulated genes from the targeted group involved in acetate metabolism. Genes highlighted in italics are identified as being differentially expressed in *acn1-2* when compared to Col-7 on one threshold, however failed to meet both thresholds. MIPS number identifiers are provided for each gene and the corresponding fold changes in expression are given.

Up Regulated Genes		
Gene ID	Transcript ID	Fold Change
At5g56720	*malate dehydrogenase, cytosolic, putative*	*2.9*
At1g08650	*phosphoenolpyruvate carboxylase kinase*	*2.0*
At3g03910	glutamate dehydrogenase, putative	1.2
At3g53180	glutamine synthetase, putative	1.1
At4g26970	Aconitase 3	1.1
At1g62800	aspartate aminotransferase, cytoplasmic isozyme 2 / transaminase A	1.1
At4g35830	Aconitase 1	1.1
At2g42600	phosphoenolpyruvate carboxylase kinase	1.1
At5g09660	malate dehydrogenase, glyoxysomal	1.1
At5g35630	glutamine synthetase (GS2)	1.1

Down Regulated Genes		
Gene ID	Transcript ID	Fold Change
At5g03860	*malate synthase, putative*	*5.5*
At3g21720	*isocitrate lyase, putative*	*4.3*
At3g53910	*malate dehydrogenase-related*	*3.5*
At2g42790	*citrate synthase (peroxisomal 3)*	*1.7*
At3g16910	*ACN1*	*1.5*
At2g30970	*aspartate aminotransferase, mitochondrial / transaminase A (ASP1)*	*1.4*
At4g37870	phosphoenolpyruvate carboxylase kinase	1.3
At1g68750	phosphoenolpyruvate carboxylase kinase	1.3
At2g22780	malate dehydrogenase, glyoxysomal, putative	1.3
At1g48470	glutamine synthetase, putative	1.2

3.5 Discussion

3.5.1 Assimilation of free acetate in *acn1-2* and Col-7

As a number of other acetyl-CoA synthetase activities are present in Arabidopsis (Ke et al. 2000; Turner et al. 2005), it was necessary to determine the effect of the disruption to ACN1 activity on the overall assimilation of acetate. Acetate metabolism is not completely compromised in the *acn1-2* mutant. However, the large decrease in [^{14}C]acetate label appearing in soluble carbohydrates in the *acn1* and *acn2* mutants compared to the wild-type indicates that the component of acetate metabolism affected represents the gluconeogenic contribution (Turner et al., 2005; Hooks et al., 2007). The fate of free acetate is dependent on the levels of acetate and the relative activities of glyoxysomal vs non-glyoxysomal assimilatory pathways involving the affinities of cytosolic acetyl-CoA synthetase and COMATOSE for acetate. The involvement of both COMATOSE and ACN1 in acetate assimilation into the glyoxylate cycle would necessitate the expenditure of 2ATP for every molecule of free acetate assimilated. This raises the question as to why a mechanism would evolve to link free acetate assimilation to the energy state of the cell and suggests that regulatory mechanisms would be required to control free acetate incorporation into the glyoxylate cycle. It may be concluded that the observed effects on gene expression and metabolite levels in *acn1-2* reflect a defect in a relatively narrow, but metabolically important, aspect of acetate assimilation.

3.5.2 Acetate cycling within seedlings

In seedlings the main source of acetyl-CoA is the degradation of stored fatty acid reserves. It is likely that citrate is exported and can be partitioned into several pathways including re-entering the glyoxysome to complete the cycle of reactions (Canvin and Beevers 1961), acetyl-CoA formation (Fatland et al. 2005) or transported into mitochondria (Pracharoenwattana et al. 2005). Other types of carbon exportable from the glyoxysome are succinate as part of the classic

glyoxylate cycle (Canvin and Beevers 1961), glyoxylate as amino acids (Cornah et al. 2004) and malate (Pracharoenwattana et al. 2007). A suggested model of acetate cycling within seedlings is presented in figure 3.5.

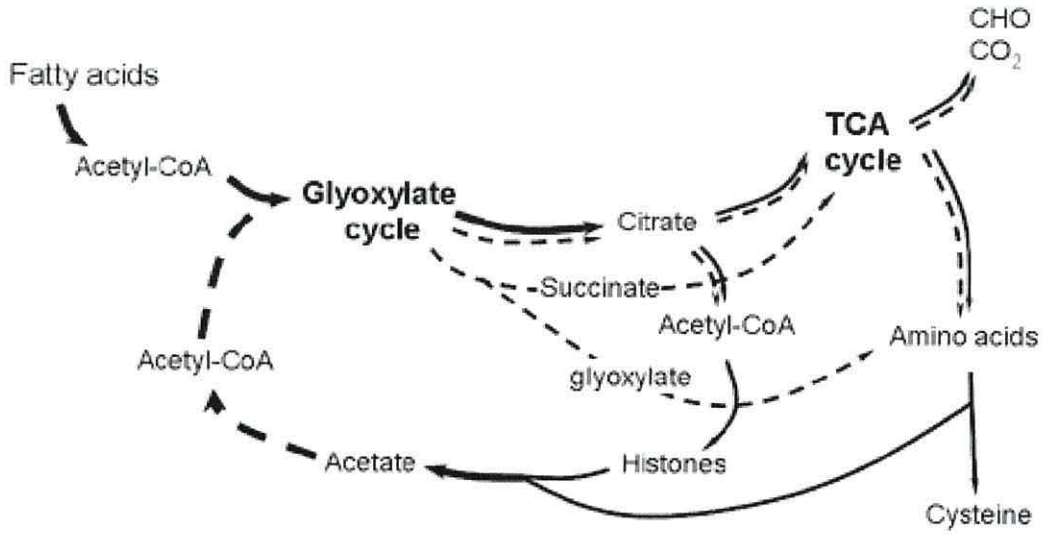


Figure 3.5: Model of acetate cycling in Arabidopsis during seedling development derived from a number of published reports based on labelled acetate feeding to seed and seedling tissues and from the characterisation of glyoxylate cycle mutants. Evidence suggests that in Arabidopsis seedlings mobilising fats, acetyl-CoA from β-oxidation is exported primarily as citrate, which has several metabolic fates. As the seedling develops, free acetate will be formed from the synthesis of cysteine and other sulphur containing compounds derived from it. The free acetate can either be reincorporated into the glyoxylate cycle or activated by the cytosolic or plastidial AcetCS. (CHO – carbohydrates, TCA cycle – tricarboxylic acid cycle)

The historical view of the glyoxylate cycle shows that malate is formed from glyoxylate and acetyl-CoA, which in turn is converted to oxaloacetate by malate dehydrogenase. Oxaloacetate reacts with another acetyl-CoA to form citrate. However, it has been suggested that in yeast and plants the primary function of glyoxysomal malate dehydrogenase is to reduce oxaloacetate to malate, thereby re-oxidising NADH formed during β-oxidation (Mettler and Beevers, 1980; van Roermund, 1995; Pracharoenwattan et al. 2007). The labelling of the carbons of the side chain of glutamine from 2[^{13}C]acetate feeding studies in wild-type seedlings provides an indication of the flow of organic acids from peroxisomes to

the mitochondria. The observed labelling pattern of C3 and C4 with a lower amount in C2 is possible only if glutamine is produced via citrate labelled at C3 and C5 (the C4 of citrate is removed in a decarboxylation reaction to form α-ketoglutarate). The described labelling of citrate may occur if glyoxysomal acetyl-CoA produces both malate and citrate via malate synthase and citrate synthase, respectively, but that malate is primarily incorporated into citrate via oxaloacetate and does not enter mitochondria. The entry of malate or succinate into the mitochondria would result in the random labelling of the C2 and C3 carbons of malate and the subsequent equivalence of label in the C2 and C3 of glutamine. The gradual appearance of label in the C2 of glutamine may be due to the circulation of carbon through the TCA cycle, including fumarase activity. The C3 and C4 carbons of glutamine showed lower levels of label in *acn1-2* compared to Col-7 which reflects the decreased levels of acetyl-CoA produced from labelled acetate entering citrate and malate. These results support the model of glyoxylate cycle activity and peroxisomal malate dehydrogenase activity as described by Mettler and Beevers (1980) and Pracharoenwattana et al. (2007), if malate exported from the glyoxysome is converted to oxaloacetate, through the activity of cytosolic malate dehydrogenase, which then re-enters the glyoxysome for partitioning between malate and citrate.

3.5.3 The biological relevance of ACN1

The biological relevance of an enzyme responsible for the conversion of acetate to acetyl-CoA for entry into the glyoxylate cycle is not immediately apparent during seedling establishment. The activity of ACN1 lies outside the catabolic processes involved in the mobilisation of stored resources and the β-oxidation of fatty acids which are central to metabolism during development. An obvious function for ACN1 is the recycling of acetate in the cytosol. However, the low levels of free acetate present in Arabidopsis presented with an exogenous source indicates that plant cells are able to assimilate acetate effectively, with only a small proportion of assimilated acetate entering peroxisomes through the activity of ACN1. Despite the assumed low involvement of ACN1 in acetate assimilation, a disruption to the activity of the enzyme has an effect on the levels

of a wide range of metabolites whereby the majority of metabolites measured decrease by the same proportion. This suggests that a disruption in *acn1* activity has a global effect on metabolism. The observed results may be due to a direct effect of a disrupted input into primary metabolism or due to an altered developmental state caused by the *acn1* mutation which indirectly affects metabolite levels.

A potential role for ACN1 in the production of acetyl-CoA from a cytosolic free acetate pool is possible. Although a number of mechanisms exist for the conversion of free acetate to acetyl-CoA within cells (Hooks et al., 2004), acetate levels were relatively consistent after chase feeding was performed, suggesting that processes exist for the replenishment of free acetate. Several mechanisms for acetate production have been described (Bao et al., 2000). Acetate is released during the synthesis of ornithine the biosynthetic precursor for polyamines. Although it is known that some higher plants contain N-acetylornithine deacetylase activity (Thompson, 1980), its contribution to free acetate formation in Arabidopsis seedlings is probably negligible with ornithine being produced via arginine decarboxylase (Hanfrey et al., 2001). The regulation of protein function by reversible acetylation produces free acetate, but estimating quantities released is difficult without detailed information about the types and quantities of modified proteins. Cysteine biosynthesis is also likely to substantially contribute to a cellular pool of free acetate. The levels of cysteine in dry seed of loblolly pine are essentially undetectable but increase dramatically to become the tenth most abundant free amino acid within several days of radical emergence (King and Gifford, 1997). Furthermore, the flux through cysteine would be high, since it serves as the route by which sulfur is assimilated into proteins and metabolic intermediates (Giovanelli et al., 1980) and a variety of sulfur-containing defense compounds (Rausch and Wachter, 2005). Therefore, it is highly likely that cysteine biosynthesis with contributions from protein deacetylation provides a steady and high flux of free acetate within developing seedlings.

3.5.4 A general depression of metabolism occurs in *acn1-2*

The observed metabolic changes in *acn1-2* compared to Col-7 demonstrated that a general effect of the *acn1-2* mutation is the repression of the levels of a number of primary metabolites (table 3.1). This result differs from that observed for the isocitrate lyase mutant, *icl1*, at the same developmental stage (Cornah et al., 2004). An increase in glucose levels is observed in the *icl* mutant while the levels of a number of amino acids, especially glutamine decreased. The more global disruption to metabolism observed in the *acn1-2* mutant reflects the importance of the acetate input step catalysed by ACN1. It is unclear why an increase in fumarate levels is observed in *acn1-2*. The reduction of carbon flow through the glyoxylate cycle might result in a decrease in succinate levels, which could have a corresponding effect on fumarate levels. As isocitrate lyase activity appears unnecessary in seedlings grown in the light, it is possible that succinate may not be the primary organic acid transported into mitochondria, with the result that fumarate levels would remain unaffected.

3.5.5 A number of genes are differentially expressed in *acn1-2*

In order to determine whether the general depression in metabolite levels observed in *acn1-2* compared to Col-7, the associated transcriptional consequences of disrupting acetate metabolism were investigated using Affymetrix GeneChip microarrays. A set of 201 differentially expressed genes between *acn1-2* and Col-7 was identified using a 1.4-fold change cut-off for differential expression in conjunction with a t-test ($p < 0.05$). Of the 201 differentially expressed genes, 132 were repressed and 69 induced by the mutation.

In order to investigate the genes most affected by the disruption to acetate metabolism, the twenty genes with the largest increases and decreases in gene expression in the mutant were identified (table 3.2). Of the genes repressed as a result of the disruption to acetate metabolism, half were identified as encoding late embryogenesis abundant (LEA) proteins. LEA proteins were first identified

during late embryogenesis in cotton and wheat and have since been shown to be present in a number of other plant species (Galau et al., 1986; Grzelczak et al., 1982; Wise and Tunnacliffe, 2004). A specific function has yet to be identified for these genes although correlations have been described between drought stress and LEA transcript abundance (Oliver et al., 2004). LEA transcript levels decrease on re-hydration indicating that LEA proteins have a role in dehydration tolerance, especially during seed desiccation (reviewed in Wise and Tunnacliffe, 2004; Tunnacliffe and Wise, 2007). LEA protein activity has also been associated with a number of other stress responses including osmotic stress (Espelund and Jakobsen, 1992; Naot et al., 1995; Xu et al., 1996; Chourey et al., 2003) and cold tolerance, suggesting a protective role for LEA proteins (Sutton et al., 1992; Welin et al., 1994; Tsuda et al., 2000; NDong et al., 2002; Oliver et al., 2004). The rapid decline of LEA protein transcript abundance within 24 hours post-imbibition has been described for *Ricinus communis* (Han et al., 1997). Furthermore, genes encoding seed storage and embryonic development proteins were also seen to be significantly repressed in *acn1-2* when compared to Col-7. The continued transcription of these genes associated with embryonic development post-germination suggests that their roles might be extended to regulate post-embryonic development and/ or germination.

A role in the regulation of the transition from heterotrophic to autotrophic metabolism has been suggested for acetate (Sheen, 1990). Accordingly, a comparison of the germination and growth phenotypes of the *acn* mutants showed that *acn1-2* germinated at the same rate as Col-7 but established slightly quicker, suggesting that the disruption to acetate metabolism has an effect on seedling establishment (Hooks et al., 2004). Examination of gene expression in seedlings harvested at the same principal growth stage 1.0 (Boyes et al., 2001), showed that a number of genes involved in developmental processes and associated with plant hormones known to regulate development were down-regulated in *acn1-2* compared to Col-7. This suggests the possibility of a potential regulatory role for acetate during seedling establishment.

3.5.6 Cluster analysis of differentially expressed genes

A combination of hierarchical and 2D-SOM clustering was used to analyse the gene expression profiles of the 201 genes differentially expressed between *acn1-2* and Col-7 (figures 3.3 and 3.4). Approximately eight clusters of genes were formed by the dataset. The clusters formed from genes repressed in *acn1-2* contained a number of genes typically associated with late embryogenesis and germination. Genes identified as responsive to negative regulators of germination such as abscisic acid were also identified in clusters composed of genes down-regulated in the mutant. The clusters of genes induced in *acn1-2* included a number of genes associated with promoters of seedling establishment such as auxin and developmental processes such as the formation of shoot apical meristem, which occurs towards the end of the seedling establishment period.

The promoter sequences were obtained for the clusters of co-expressed genes and analysed for the presence of over-represented DNA motifs which might possibly act as regulatory elements for transcriptional regulators to bind to. Unsurprisingly, seed specific promoter elements were identified in a number of clusters of genes. The clusters composed of genes repressed in *acn1-2* contained a large number of potential promoter elements associated with the regulation of development, seed storage mobilisation and hormones including abscisic acid, a negative regulator of development. The putative elements identified in the promoters of clusters composed of genes induced by *acn1-2* featured a number of elements associated with auxin and gibberellins which are positive regulators of seed germination along with elements associated with post-germinative development. These results suggest the potential for interactions between acetate and developmental regulators.

Acetate has been shown to affect gene expression in a number of organisms including bacteria, yeast and plants (Sheen, 1990; Graham et al., 1992; Hartig et al., 1992; Oh et al., 2002). The genes differentially expressed in *acn1-2* included a number of genes involved in signalling, transcriptional and translational control. This suggests that a disruption to acetate metabolism affects a number of signalling networks, underlying the central role that acetate plays in plant metabolism. The differential expression of a number of genes involved in developmental regulation including light, phytohormones and sugars indicates

the potential for acetate signalling to interact with a variety of other signalling pathways, aspects of which will be explored in more detail in the next chapter.

3.5.7 Expression of genes related to acetate metabolism is similar between *acn1-2* and Col-7

Acetate has been shown to induce glyoxylate cycle genes in *Ashbya gossypii*, *Aspergillus nidulans* and in plants (Graham et al., 1992; Bowyer et al., 1994; Todd et al., 1998; Maeting et al., 1999). In order to investigate the effects of disrupting ACN1 activity, the expression levels of several genes involved in glyoxysomal assimilation of acetate were investigated. Using the previously determined threshold for differential expression of a minimum 1.4-fold change in expression levels in conjunction with a t-test ($p < 0.05$) showed that none of the selected genes were differentially expressed between *acn1-2* and Col-7 (table 3.5). This indicates that mobilisation of stored lipid resources through the action of the glyoxylate cycle is mainly unaffected by the *acn1-2* mutation.

When a threshold of 1.4-fold change in gene expression was used alone as the criterion for differential expression, eight genes were identified. A gene encoding phosphoenolpyruvate carboxylase previously identified as repressed by acetate was slightly induced in *acn1-2* (Kausch et al., 2001). Genes previously identified as induced by acetate, malate synthase and isocitrate lyase, were slightly repressed in *acn1-2* compared to the wild-type (Graham et al., 1992). This suggests that the disruption to ACN1-mediated conversion of acetate to acetyl-CoA might have an effect on gene expression events whereby genes normally induced by acetate are slightly repressed and *vica versa* although the differential expression is only slight. The ACN1 gene transcript was amongst the down regulated genes in *acn1-2* compared to Col-7 with a 1.5-fold decrease in expression. The production of a partial transcript, not likely to be translated into a viable protein, for ACN1 in *acn1-2* has previously been described which explains the presence of a signal in this experiment (Turner, 2004).

3.5.8 Models of acetate utilisation based on *acn1-2* profiling

The information obtained from metabolite and transcriptome profiles of *acn1-2* and Col-7 was used to develop models of acetate metabolism in seedlings (figure 3.6). Acetate levels are lower in *acn1-2* than in Col-7, but the levels of acetyl-CoA are increased in *acn1-2*. The [^{13}C]acetate feeding of seedlings has revealed information regarding the mechanisms of acetate production. Figure 3.6 (A) describes the situation occurring in the wild-type where both acetate and fatty acids are incorporated into the glyoxylate cycle and shunted to the mitochondria or cytosol for further metabolism. As metabolism continues throughout development, acetate levels may be expected to increase. In the absence of ACN1 activity, as illustrated in figure 3.6 (B), fatty acid degradation is unaffected, but a reduced input of free acetate into the glyoxylate cycle occurs, resulting in an alteration in the levels of a number of metabolites. The synthesis of amino acids is lower in the *acn1-2* mutant, resulting in a decrease in the production of free acetate. Turner et al. (2005) described the activity of a number of cytosolic acetyl-CoA synthetases which might be expected to act on the acetate produced. The increased levels of acetyl-CoA observed in the mutant might be due to the continued production of acetate and subsequent conversion to acetyl-CoA from carbohydrate (either stored or products of photosynthesis) via the TCA cycle, or acetyl-CoA may accumulate in the cytosol due to slower subsequent metabolism.

The increased levels of acetyl-CoA observed in the mutant might be due to the export of acetate from the glyoxysomes which would not be reactivated to acetyl-CoA for re-entry into the glyoxylate cycle. It has been demonstrated that acetyl-CoA is unable to cross the peroxisomal membrane, but the presence of acyl-CoA thioesterases in Arabidopsis peroxisomes which are able to hydrolyse acetyl-CoA to acetate might provide a means of export (van Roermund et al., 1995; Tilton et al. 2000; Zolman et al. 2001). A similar increase in the levels of acetyl-CoA is observed when the *S. cerevisiae* acetyl-CoA hydrolase is expressed in the mitochondria of tobacco plants where it is likely that free acetate is exported from the mitochondria for reactivation in the cytosol (Bender-Machado et al. 2004). The lack of ACN1 activity in seedlings may be compensated for by the production of acetyl-CoA from stored resources via β-

oxidation. It is possible that the increased levels of cytosolic acetyl Co-A might promote development in optimal growing conditions by feeding directly into biosynthetic processes (figure 3.6 B). These results support the conclusion of Bender-Machado et al. (2004) that acetate and acetyl-CoA levels are tightly regulated.

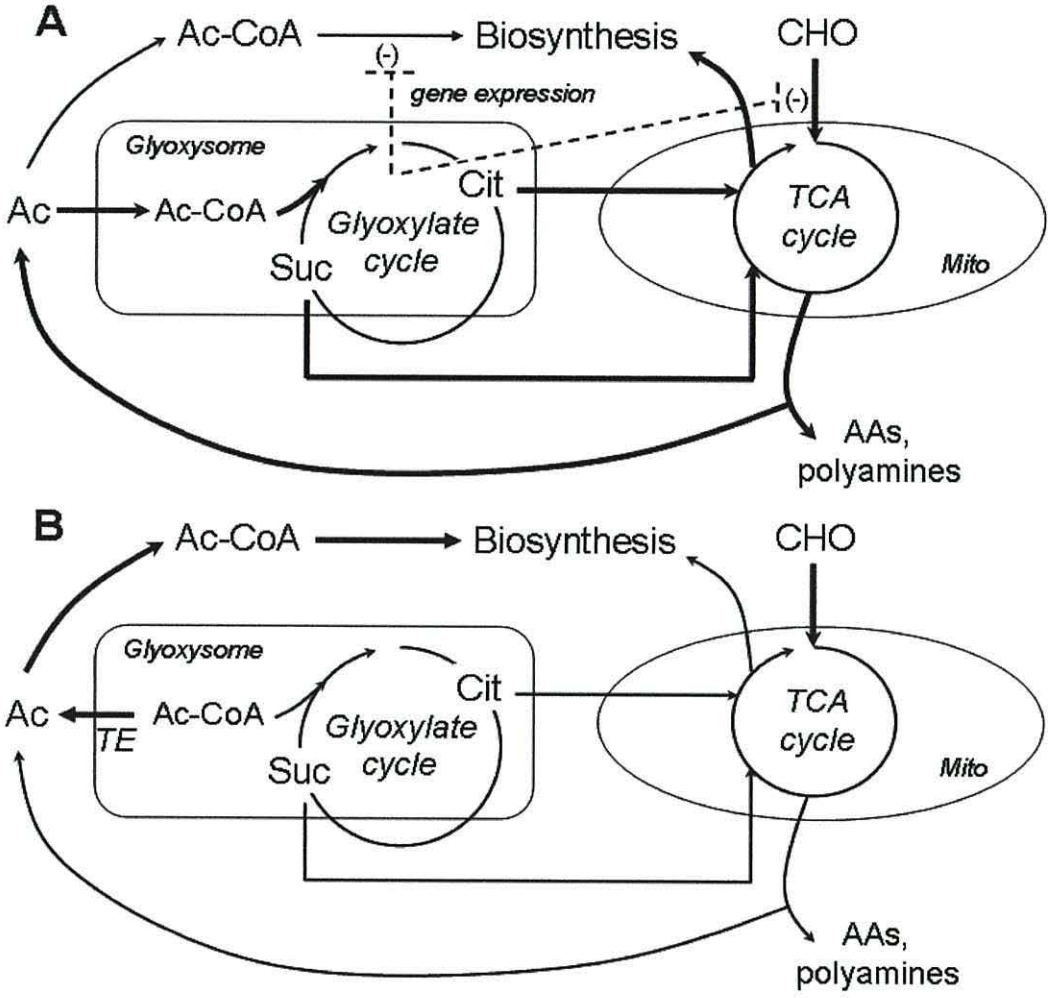


Figure 3.6: Models derived to explain the effect of the *acn1-2* mutation on metabolite levels and gene expression events. (A) Wild-type situation showing the normal function of the glyoxylate cycle. (B) *acn1-2* situation where reduced acetate import into the glyoxylate cycle might result in an accumulation of acetyl-CoA in the cytosol which may feed directly into biosynthetic processes. The possible export of acetate from the glyoxysomes through the action of peroxisomal acyl-CoA thioesterases is indicated. (CHO – carbohydrate, TCA cycle – tricarboxylic acid cycle, Cit – Citrate, AA – amino acids, Ac – acetate, AcCoA – acetyl CoA, Suc – sucrose, TE – thioesterase).

Chapter 4

Transcriptomic investigation of acetate and sugar signalling cross-talk

4.1 Introduction

A dynamic network of interactions is employed by all organisms to allow alterations in gene expression and metabolite levels in response to environmental stimuli or developmental phases. Appropriate sensing mechanisms must be established to allow these events to occur. One of the most well characterised sensing networks is that involved in sugar sensing. A number of mechanisms for sugar sensing have been suggested in plants. One is a sucrose-specific signalling pathway which probably utilises a signalling sucrose transporter; the other suggested pathways are concerned with sensing hexoses and are classified according to whether or not hexokinases are involved (Smeekens and Rook, 1997; Halford and Paul, 2003; Gonzali et al., 2006).

Sensing and signalling pathways do not function independently; rather they form part of a 'matrix' composed of multiple signalling pathways with common components or regulatory factors (Coruzzi and Zhou, 2001). Sugar signalling pathways are known to interact with a variety of other sensing mechanisms. Connections between carbon and nitrogen metabolism are indicated by the description of a number of genes known to be responsive to both carbon and nitrogen (Coruzzi and Bush, 2001; Coruzzi and Zhou, 2001). A number of genes involved in nitrogen metabolism have been shown to be more responsive to the application of exogenous glucose than nitrogen (Price et al., 2004). Sugar signalling pathways have also been shown to interact with various plant hormones including ethylene, abscisic acid and giberellic acid during germination (Zhou et al., 1998; Yuan and Diller, 2006). Interactions between carbon signalling and environmental factors such as light have also been identified (Thum et al., 2004). Further examples of sugar signalling pathways

cross-talk are reviewed in the literature (Smeekens, 2000, Rolland et al., 2002, Halford and Paul, 2003), illustrating the extent of the sugar signalling network.

4.1.2 The *acn1-2* mutant is disrupted in sugar sensing

The Arabidopsis *acn1-2* mutant was used to investigate the effects of the disruption of acetate metabolism on carbohydrate response. The mutant lacks a short-chain acyl-CoA synthetase responsible for converting acetate to acetyl-CoA for entry into the glyoxylate cycle which was identified by screening on a toxic acetate analogue, fluoroacetate (Turner et al., 2005). Several other acetate non-utilisation mutants, *acn2* and *acn3*, were identified in this screen that showed an altered phenotype when grown on sucrose. The *acn2* mutant demonstrated sensitivity to the concentration of sucrose in the growth medium and displayed different germination rates on varying concentrations of sucrose. The *acn3* mutant demonstrated bleaching of the cotyledons during the first few days of growth on 20 mM sucrose, which was overcome with the addition of acetate (Turner, 2003; Hooks et al., 2004). The *acn1* mutant has been identified as a true acetate non-utilising mutant and is disrupted in a specific aspect of acetate metabolism; the conversion of acetate to acetyl-CoA for entry into the glyoxylate cycle (Turner et al., 2005). Both the *acn2* and *acn3* mutants exhibit pleiotrophic effects when germinated on sucrose while the *acn1* mutant displays a specific response and was therefore selected for analysis in this experiment as the observed effects may be attributed directly to the disruption in acetate metabolism (Turner, 2003; Hooks et al., 2004).

4.2 Identification of sucrose-responsive genes in Col-7 and *acn1-2*

A combination of fold-change threshold for differential expression and statistical testing was used to identify genes differentially expressed in response to sucrose in *acn1-2* and Col-7. A similar number of genes were identified as sucrose-responsive in Col-7 and *acn1-2* but the behaviour of the genes was markedly

different; the majority of the genes were down-regulated in Col-7 while in *acn1-2* the majority of the genes were up-regulated. Of the 3580 genes identified as sucrose-responsive in Col-7, 1249 (35%) were induced and 2331 (65%) were repressed by sucrose. Of the 3255 sucrose-responsive genes identified for *acn1-2*, 1909 (59%) were induced and 1346 (41%) were repressed by sucrose (figure 4.1).

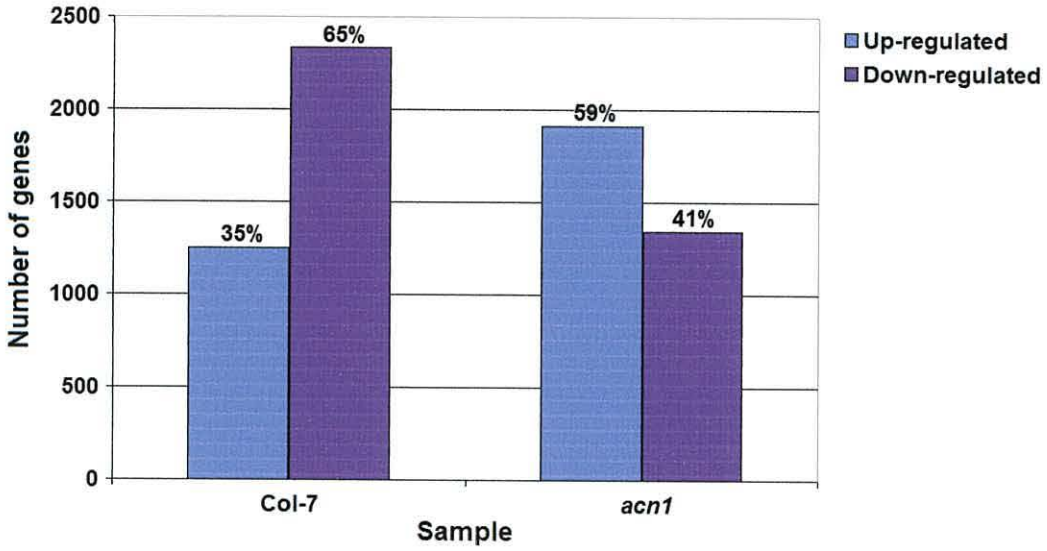


Figure 4.1: Sucrose-responsive genes identified in Col-7 and *acn1-2*. The majority of the sucrose-responsive genes in Col-7 are down-regulated while in *acn1-2* the genes are mainly up-regulated. No error bars are shown as data were summed prior to analysis.

Within a large group of genes differentially expressed in response to a stimulus it might be argued that the genes with the largest increases and decreases in expression levels are the most biologically relevant (Breitling et al., 2004). In order to determine the predominant response to sucrose in both Col-7 and *acn1-2* the genes with the ten largest increases and ten largest decreases in expression levels were identified (table 4.1)

Table 4.1: The top ten up and down regulated genes identified in Col-7 and *acn1-2* in response to sucrose. MIPS numbers are provided as gene identifiers and the corresponding fold changes in gene expression are listed (table continued on page 110).

Top ten genes up-regulated in Col-7 in response to growth on 100 mM sucrose		
Transcript Identifier	Gene Title	Fold change
At1g03495	transferase family protein	1375
At5g20150	SPX (SYG1/Pho81/XPR1) domain-containing protein	359
At1g56650	myb family transcription factor (MYB75)	348
At3g17430	phosphate translocator-related	247
At3g58060	cation efflux family protein / metal tolerance protein, putative (MTPc3)	246
At3g25980	mitotic spindle checkpoint protein, putative (MAD2)	225
At3g62860	esterase/lipase/thioesterase family protein	222
At2g34050	expressed protein	217
At4g19840	lectin-related	201
At1g63680	Mur ligase family protein	199

Top ten genes down-regulated in Col-7 in response to growth on 100 mM sucrose		
Transcript Identifier	Gene Title	Fold change
At5g64110	peroxidase, putative	322
At2g30610	BTB/POZ domain-containing protein (obsolete, replaced by At2g30600 in 2003)	312
At1g08630	L-allo-threonine aldolase-related	272
At4g19170	9-cis-epoxycarotenoid dioxygenase, putative / neoxanthin cleavage enzyme, putative / carotenoid cleavage dioxygenase, putative	267
At1g03400	2-oxoglutarate-dependent dioxygenase, putative	265
At2g20670	expressed protein	217
At3g16220	expressed protein	204
At4g37610	TAZ zinc finger family protein / BTB/POZ domain-containing protein	201
At1g29450	auxin-responsive protein, putative	185
At5g62360	invertase/pectin methylesterase inhibitor family protein	162

Top ten genes up regulated in *acn1-2* in response to growth on 100 mM sucrose		
Transcript Identifier	Gene Title	Fold change
At2g45970	cytochrome P450, putative	40
At5g51570	band 7 family protein	34
At1g74900	pentatricopeptide (PPR) repeat-containing protein	20
At4g14950	expressed protein	18
At5g37870	seven in absentia (SINA) family protein	15
At5g50890	hypothetical protein	15
At1g75660	5'-3' exoribonuclease (XRN3)	14
At4g18610	expressed protein	12
At5g09320	vacuolar sorting protein 9 domain-containing protein	12
At1g78310	VQ motif-containing protein	12
Top ten genes down regulated in *acn1-2* in response to growth on 100 mM sucrose		
Transcript Identifier	Gene Title	Fold change
At1g16540	molybdenum cofactor sulfurase (LOS5) (ABA3)	7
At5g34950	replication protein-related	5
At1g66380	myb family transcription factor (MYB114)	5
At3g05610	pectinesterase family protein	5
At1g49520	SWIB complex BAF60b domain-containing protein	5
At3g25850	DC1 domain-containing protein	5
At1g09610	expressed protein	5
At2g04870	hypothetical protein	5
At4g33050	calmodulin-binding family protein	5
At1g22340	UDP-glucoronosyl/UDP-glucosyl transferase family protein	4

The transcriptome profiles show that a differential response to sucrose exists between Col-7 and *acn1-2* on the basis of global gene expression levels. In general, the observed changes in gene expression are larger in Col-7 than in *acn1-2* in both a positive and negative direction. The largest change in gene expression in Col-7 in response to sucrose is 1375-fold up regulation of a transferase family protein (At1g03495). The remainder of the genes up regulated in Col-7 in response to sucrose showed lower increases in expression levels but remained consistently higher than the largest gene expression increases observed in *acn1-2*; fold-change increases in Col-7 were approximately 19 times those seen in *acn1-2* in response to sucrose. A similar trend is apparent in the genes which decreased in expression in response to sucrose with fold-change decreases in Col-7 approximately 47 times those seen in *acn1-2*.

The gene ontology analysis tool available from TAIR (Berardini et al., 2004) was used to examine the ontology of the sucrose-responsive genes with the largest fold-changes in expression levels. Of the genes with the largest increase

in expression levels in Col-7 three were involved in anthocyanin synthesis or metabolism, three were located to the chloroplast and two were involved in carbohydrate binding. The top ten up regulated genes in response to sucrose in *acn1-2* included three which were located to the chloroplast and two with a role in nucleic acid binding. Although the genes with the ten largest decreases in expression levels in Col-7 were not enriched in a specific ontology several of the genes had previously been associated with hormones involved in developmental regulation (At1g29450 and At1g03400). The top ten down regulated genes in *acn1-2* included two genes involved in transcriptional regulation (At1g66380 and At1g49520) and a gene previously identified as responsive to sucrose and stress (At1g16540).

4.3 Identification of sucrose-responsive genes that showed an alternate response in *acn1-2* than in Col-7

A disruption to the sucrose response in Arabidopsis seedlings was described for two acetate non-utilisation mutants, *acn1* and *acn2* (Turner, 2003; Hooks et al., 2004; Turner et al., 2005). The sucrose responsive genes identified in the previous section were used to identify a set of genes with an altered response to sucrose in *acn1-2* compared to Col-7. Expression values for each gene identified as sucrose-responsive were averaged for each of the four sample types and visualised using a 2D-SOM. A Pearson correlation distance metric was used to cluster genes as the results are based on the correlation of data points over all the variables, irrespective of scale, allowing genes that behave in a similar manner to group together (Stekel, 2003). The gene clusters produced were inspected visually in order to identify genes with the same sucrose response in both Col-7 and *acn1-2* (e.g. sucrose induces gene expression to the same extent in both genotypes, figure 4.2 A).

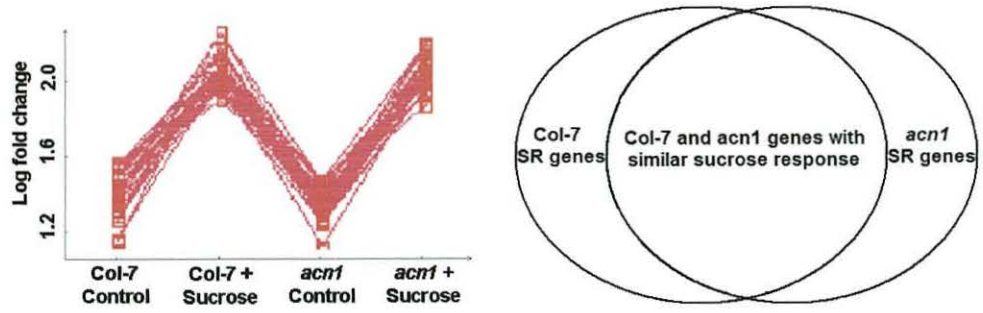


Figure 4.2: Gene expression values were averaged and visualised in a 2D-SOM. Clusters of genes with the same response to sucrose in *acn1-2* and Col-7 (A) were identified and removed from analysis. The Venn diagram (B) illustrates the dataset of genes. The sucrose-responsive genes identified from both *acn1-2* and Col-7 were pooled. Genes with similar responses to sucrose were removed (shaded area) leaving only those genes identified as sucrose responsive (SR) in Col-7 with an altered response in *acn1-2* and *visa versa*.

Clusters of genes with the same sucrose-response in *acn1-2* and Col-7 were identified and removed from further analysis. The remaining genes were re-clustered repeatedly until no more genes with an unaffected sucrose response could be identified in the clusters. A set of 342 genes was identified with an altered response to sucrose in *acn1-2* compared to Col-7 (listed in Appendix D).

4.3.1 Clustering of genes with an altered sucrose response in *acn1-2* compared to Col-7

A hierarchical clustering algorithm was used to determine the number of clusters formed by the 342 genes identified with an altered sucrose response as a result of the disruption to acetate metabolism. Approximately twelve gene clusters were formed as a result of clustering (figure 4.3).

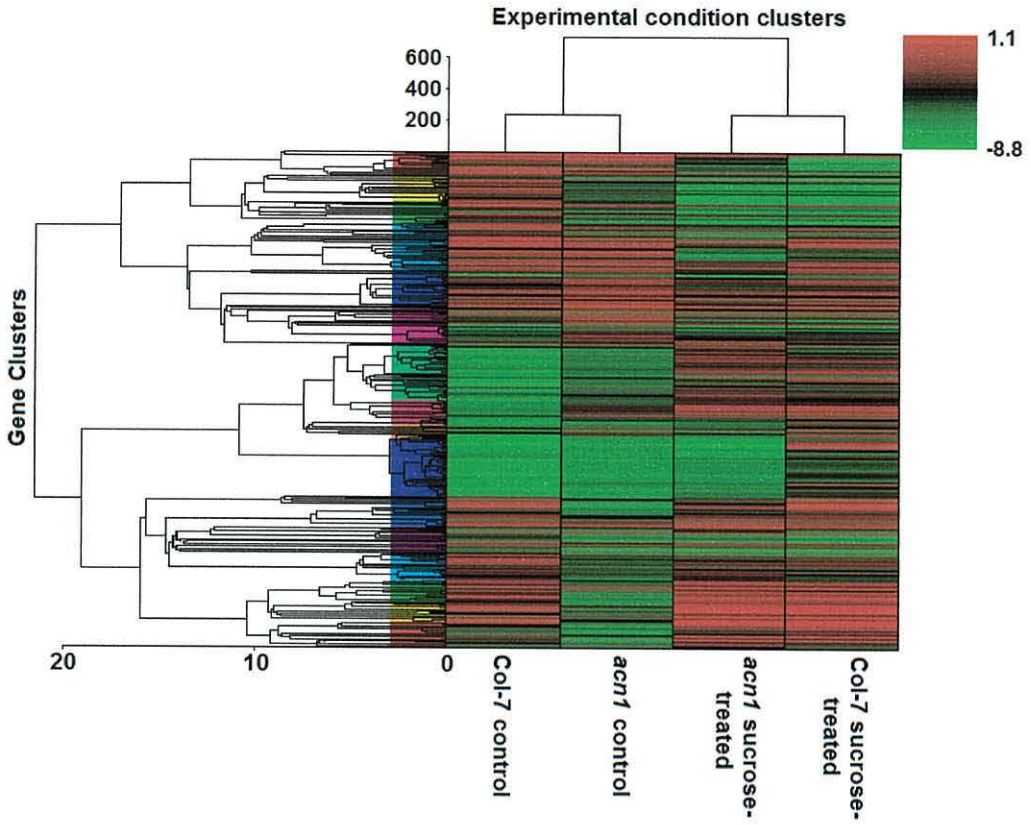


Figure 4.3: Hierarchical organisation of genes and experimental conditions based on the expression values of the 342 genes with an altered response to sucrose in *acn1-2* than in Col-7. The distance (Pearson correlation distance metric) between clusters is represented by the scale bars at the bottom and left of the clusters and gene expression intensity is described by the colour scale.

A *k*-means clustering algorithm was used to produce groups of genes for further analysis. *K*-means clustering groups data into a user-defined number of clusters, *k*, which are not organised hierarchically or by any other relationship and in which data profiles are not organised. The genes to be clustered are initially assigned to one of the *k* clusters at random and cluster centroids calculated. The gene expression profiles are examined to determine how closely each expression profile corresponds to the centroid of each of the *k* clusters. If a gene expression profile corresponds more to the centroid of a cluster other than the one in which it has been placed, it is transferred to that cluster and all of the cluster centroids are re-calculated. The process is repeated until no more gene expression profiles are moved upon calculation of the cluster centroids (Stekel, 2003). The 342 genes with an altered sucrose response were clustered using *k*-

means clustering, $k = 16$ as determined by hierarchical clustering (figure 4.4). A Pearson correlation distance metric was used so that genes with similar expression profiles grouped together.

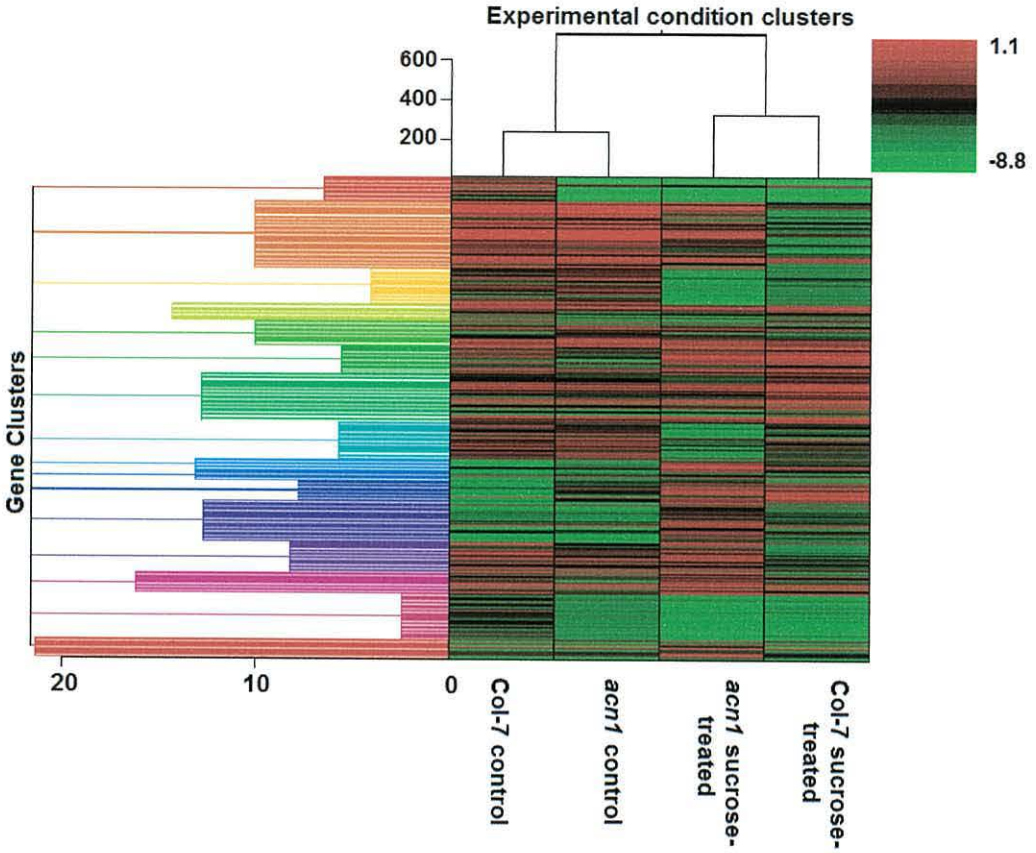
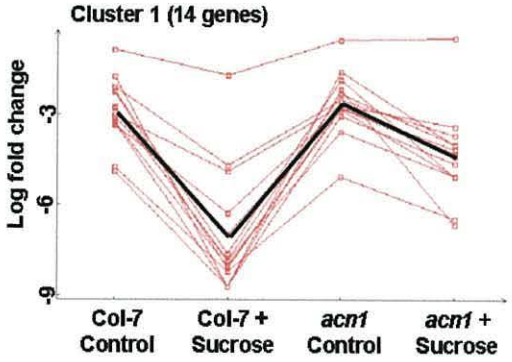


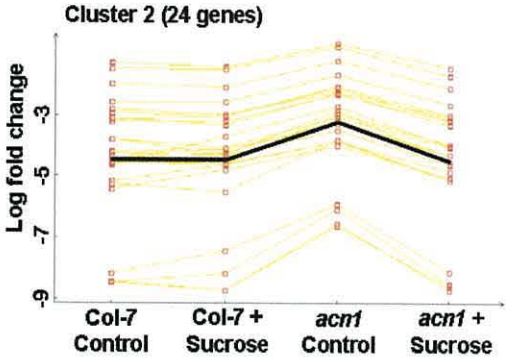
Figure 4.4: Grouping of the 342 gene expression profiles with an altered sucrose response using a *k*-means clustering algorithm ($k = 16$). The distance (Pearson correlation distance metric) between clusters is represented by the scale bars at the bottom and left of the clusters and gene expression intensity is described by the colour scale.

The expression profiles of the genes in each of the sixteen clusters were visualised as shown in figure 4.5. For ease of viewing, gene expression values were averaged and displayed across the four experimental sample types. Each cluster displays an altered response to sucrose in *acn1-2* than in Col-7; the average expression profile is represented by a thick black line running through

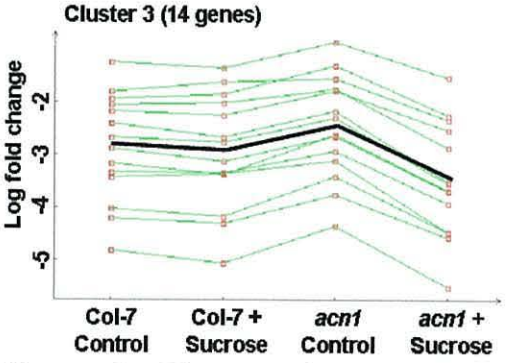
each cluster of genes. It can be seen that a wide variety of altered gene expression responses to sucrose are observed in *acn1-2* when compared to Col-7.



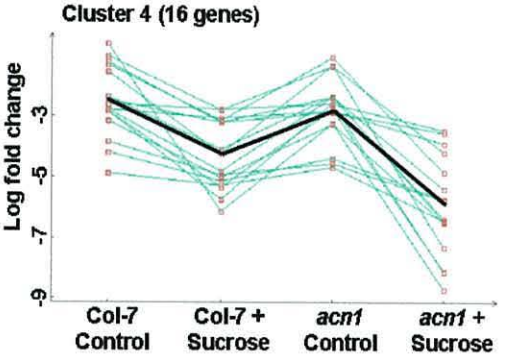
Cluster 1: The mutation represses the sucrose response



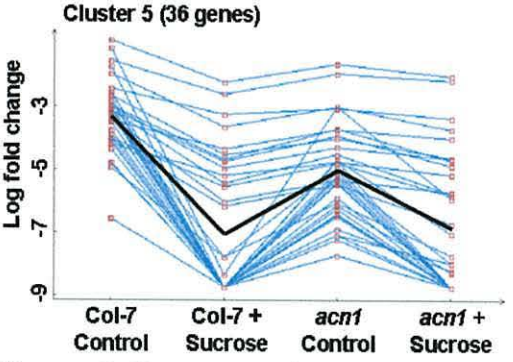
Cluster 2: The mutation increases the sensitivity to sucrose.



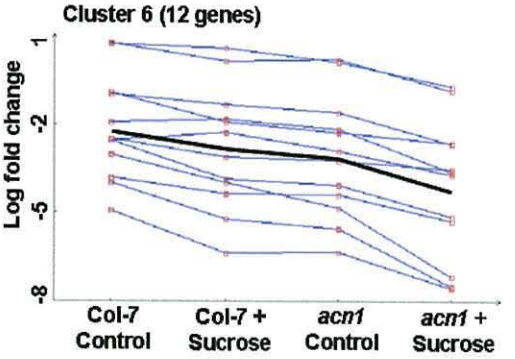
Cluster 3: The mutation alters gene expression. Sucrose represses expression.



Cluster 4: The mutation enhances the sucrose response

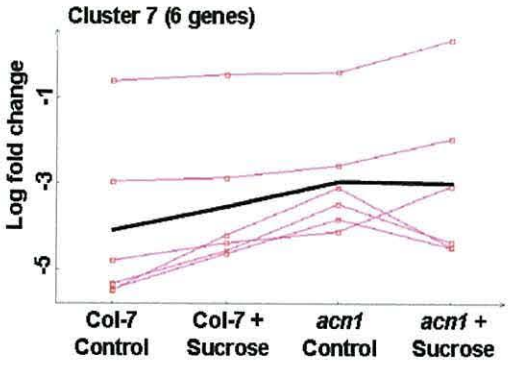


Cluster 5: Gene expression is repressed in the mutant; sucrose represses this further.

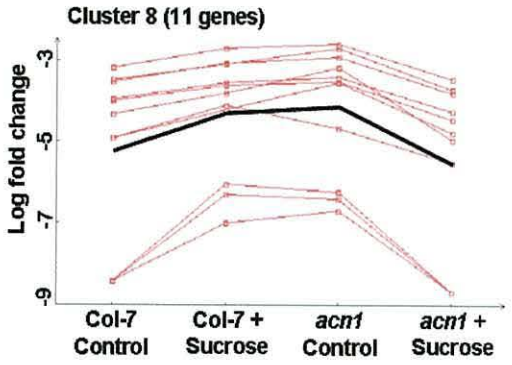


Cluster 6: Gene expression is repressed in the mutant and the sucrose response may be slightly increased.

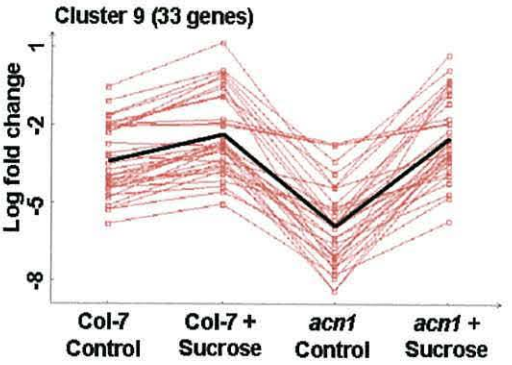
Figure 4.5: (legend on page 117)



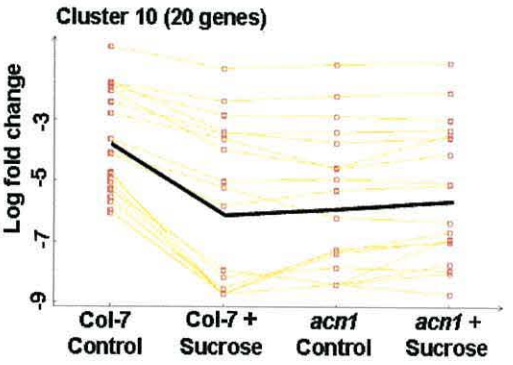
Cluster 7: The mutation de-represses gene expression and decreases the sucrose response.



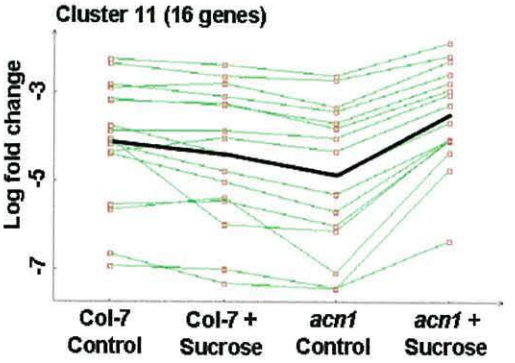
Cluster 8: The mutation de-represses gene expression and reverses the sucrose response.



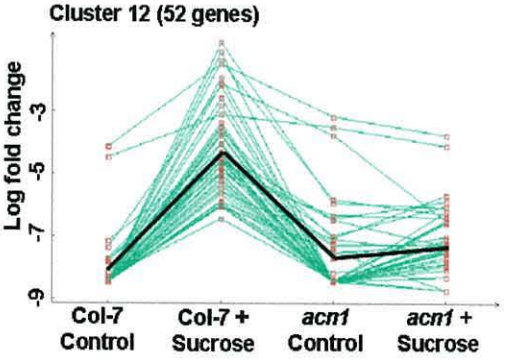
Cluster 9: The mutation represses gene expression and increases the sucrose response.



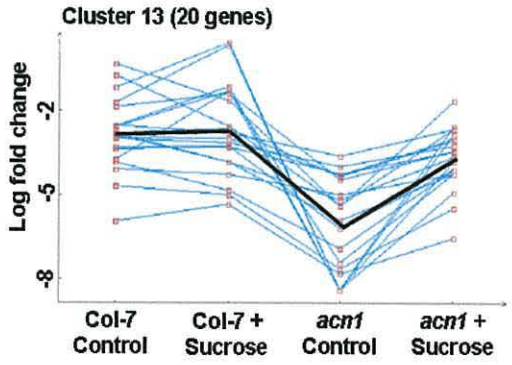
Cluster 10: The mutation represses gene expression and decreases the sucrose response.



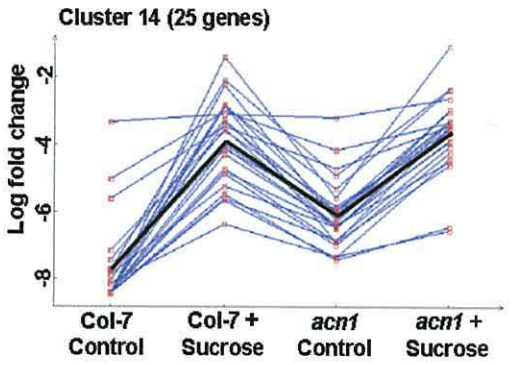
Cluster 11: The mutation represses gene expression. Sucrose restores expression levels.



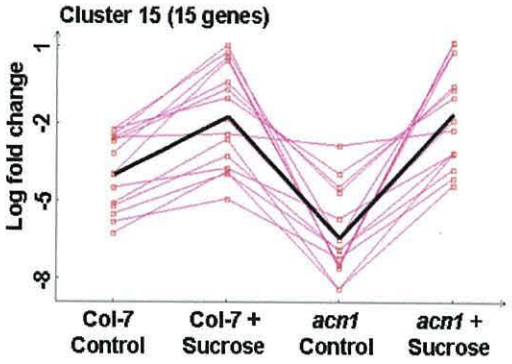
Cluster 12: The mutation represses the sucrose response.



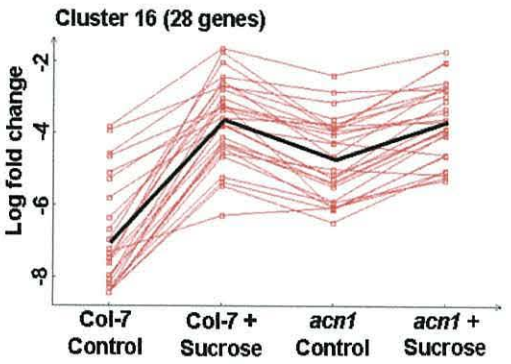
Cluster 13: The mutation represses gene expression. Sucrose acts to restore expression levels.



Cluster 14: Gene expression de-repressed in the mutant. The sucrose response is unaffected.



Cluster 15: Gene expression is decreased in the mutant and apparently enhances the sucrose response.



Cluster 16: The mutation de-represses gene expression and decreases the sucrose response.

Figure 4.5: Grouping of the 342 gene expression profiles with an altered sucrose response in *acn1-2* compared to Col-7. Gene expression values were averaged and displayed across all experimental types. The average profile for each cluster is represented by a thick line running through each chart. The $\log_2$-transformed expression values are represented by the vertical axes and a brief description of expression profiles is given.

4.4 Identification of common promoter motifs in co-expressed genes

One of the ways in which gene expression levels may be regulated at the level of transcription is through the binding of transcriptional regulators to DNA sequence elements in the promoters of genes (Schwechheimer and Bevan, 1998). A number of algorithms for the detection of motifs in the promoters of co-

expressed genes and databases containing sequences of characterised promoter motifs have been developed (Rombauts et al., 2003). Potential DNA motifs over-represented in the promoters of the co-expressed genes were compared to a background model were identified using the MotifSampler software detailed in chapter 2, section 2 (Thijs et al., 2001; 2002). The identified motifs were ranked according to log-likelihood score, consensus score and information content (Thijs et al., 2002). Potential motifs which scored well on all three counts were summarised by comparing and grouping similar motifs prior to homology searching with sequences from the online databases of plant promoter elements, PlantCARE and PLACE (Higo et al., 1999; Lescot et al., 2002). A summary of the potential DNA motifs identified in the promoter sequences of co-expressed genes is presented in table 4.2. Corresponding motifs identified in the online *cis*-acting regulatory element (CARE) databases are listed with a brief description and regions of homology are highlighted in **bold** in the sequences obtained from the databases. Sequences identified in the reverse complement orientation are highlighted in ***bold italics***. Unless otherwise stated the promoter sequences from the databases were identified in Arabidopsis.

Table 4.2: Consensus sequences of motifs found in each cluster using MotifSampler (Thijs et al., 2002). Where nucleotide ambiguity occurs, the IUPAC nucleotide symbol system has been used. Motifs were submitted to both the PlantCARE ([a]) and the PLACE ([b]) online *cis*-regulatory element databases (Higo et al., 1999; Lescot et al., 2002) (Table continued on pages 119-123).

Cluster	Motif Submitted	Sequence Found	Descriptior
1	AGATTG	**NGATT**[b]	ARR1-binding element
	ATTCTATTT	***ATAGAA***[b]	Response regulator Box II Found in plastid gene promoter (*N. tobacco*)
	GTTTA	*TAAAC*GTG[a]	G-box Light responsive element (*B. oleracea*)
	TAATTAAA	*ATTAAA*[b]	PolyA signal Found in alpha-amylase (Rice)
	TGATT	*NGATT*[b]	ARR1- binding element
		AATCACAACCATA[a]	ACA-motif

			Light responsive element
	TGTTATG	No match in database	-
2	AGAAGG	No match in database	-
	ACTGT	GGACGC**ACTGT**GA	GT-1 factor binding site
		TTAAATAT[a]	Embryo specific (*P. vulgaris*)
	CACGT		ABRE-like sequence
			Abscisic acid and drought response
		ACGTG[b]	
			ACGT motif
		ACGT[b]	ABRE element (potential)
			Abscisic acid and drought
		CACGTG[a]	response
			Box II
		ACACGTAGA[a]	Light responsive element
			G-box
		CTTC**CACGT**GGCA[a]	Light responsive element found in dicotyledenous seed-storage protein promoters
	CTGCTTCT	No match in database	-
	CTTCTT	TCATAT**CTTCTT**	Sbp-CMA1a
		CAG[a]	Involved in light responsiveness
		CTCTTAGACCGC**C**	
		TTCTTTGAAAG[a]	Jasmonate and elicitor responsive (*C. roseus*)
	GAGAGAGAG	***TCTCTCTCT***[b]	CT-rich motif (inverted GAGA)
			Enhances gene expression
		TTTC***TCTCTCTCT***C[a]	
			Py-rich region
			Enhances gene expression (*L. esculentum*)
	GGTTT	***AAACC***AATT[a]	ELI-box3
			Elicitor response element (*B. oleracea*)
	GTNTC	TGACAC**GTGGC**A[a]	G-box
			Light responsive element; has higher activity in light-grown seedlings than dark-grown
	TTCANTGAA	No match in database	-
3	AGGWTA	**AAGGATA**AGG[a]	GATA motif
			Light responsive element (*S. tuberosum*)
		TGGA**AGGTTA**GTA	Endosperm box regulation
		TGATGACATG[a]	(*Z. mays*)
	GASTCA	***TGACT***[b]	W-box
			Found in isoamylase1

			promoter (Barley)
		TGACY[b]	
			W-box Wounding induced activation of a transcriptional repressor
		TGAGTC*A*[a]	(*N. tobacco*)
			GCN4 motif Located to endosperm (*O. sativa*)
	WATTTCCT	**AAATTTCCT**[a]	WUN-motif Wound responsive (*B. oleracea*)
	TTAGTTAA	No match in database	-
4	AAACTGA	No match in database	-
	ATTG	CAATW**ATTG**[a]	HD-Zip 1 and 2 Regulates development
		ATCTTATGTC**ATTG**	Ocs element
		ATGACGACCTCC[a]	Auxin, salicylic acid and oxidative stress response
	TATATAAA	**TATATAA**[b]	TATA-box Core promoter element
	TCAGTTTT	No match in database	-
5	AGRGT	AG**AGAGT**[a]	GAG-motif Light responsive element
	CCTTTGATT	**NGATT**[b]	ARR1-binding element Response regulator
	GAGNAAGAG	No match in database	-
	GATTTG	TG***CAAATC***T[a]	CAAT-box Common promoter element (*B. juncea*)
	TTANGGTT	No match in database	-
6	AAGCT	TTCTTC**AAGCT**TCA	Heat shock element
		AGACAATCCTAGA	Heat stress and abscisic acid response
		AATTAC[a]	(*H. annuus*)
	AATCGCA	***NGATT***[b]	ARR1-binding element Response regulator
	AGATGCC	No match in database	-
	CGATT	**NGATT**[b]	ARR1-binding element Response regulator
	GNAAGC	***GCTTAC***CTACCA[a]	AC-I element Regulation of xylem and phloem expression
	TCGNNTCG	No match in database	-
	TGCGAT	TGACGAA**TGCGAT**	As-1-box
		GACC[a]	Various stress responses (*N. tabacum*)
	TTTCCT	AAA**TTTCCT**[a]	WUN-motif Wound response (*B. oleracea*)

7	AACGGNTAA	**AACGG**[b]	Myb core Reporter gene activation
	AGGCA	GAA**AGGCA**GAC[a]	CAG-motif Light responsive element
	AGTGCC	No match in database	-
	CGYTA	No match in database	-
	CTAAC	GTCCAT**CTAAC**CT A CCAC[a]	Box I Light responsive element
	TACTTA	**TTACTTAA**[a]	Chs-CMA2a Light responsive element (*D. carota*)
	TAGCGA	No match in database	-
	TTAATA	TGTGTGG**TTAATA** TG[a]	GT1-motif Light responsive element
	TTANCCNTT	No match in database	-
8	AAKGCA	GA**AAGGCA**GAC[a]	CAG-motif Light responsive element
	AATTCAA	No match in database	-
	GAAGA	CAAAT**GAARA**[a]	Gap box Light responsive element
	GATANATGG	No match in database	-
	GATGGCATC	No match in database	-
	TCGACG	**CGACG**[b]	CGACG element Found in amylase gene promoter (*O. sativa*)
		AGAT**CGACG**[a]	NON box Meristem specific activity
	TTATGG	No match in database	-
9	ATTCGGANTC	No match in database	-
	CGTTNTT	TTTCTTG**CGTTTTT** T GGCATAT[a]	Unnamed_7
	GAAGAAGA	No match in database	-
	GAGWTTC	No match in database	-
	TATATA	**TATATA**[a]	TATA box Core promoter element
	TCGANTCG	No match in database	-
	TTCTCCTT	No match in database	-
10	AGATANNA	**AAGATAAGA**TT[a]	GATA motif Light responsive element Expressed in seedlings
	TGNATC	No match in database	-
	TTWGGWTT	No match in database	-
11	ACTGA	AAGTCATA**ACTGA**[a]	Unnamed_1 Seed specific (*O. sativa*)
		TAAGTCATA**ACTG** ATGA[a]	GTCA direct repeat (*O. sativa*)
	CTWAR	No match in database	-
	KAARCC	No match in database	-
	TCTCT	AGAGAGT[a]	GAG motif

			Light responsive element
12	CTTTCTC	No match in database	-
13	ACACGTGT	**ACGTG**[b]	ABRE-like motif Dehydration responsive
		CACGTG[b]	CACGTG motif Transcriptional repressor Embryogenesis related
		A**CACNNG**[b]	b-Zip transcription factor ABA responsive, embryo specific
		CANNTG[b]	MYC recognition site Dehydration, drougt and stress response. Seed specific.
		TGA**CACGTG**GC[a]	G-box Light responsive element; has higher activity in light-grown seedlings than dark-grown
	ACCNTWAA	No match in database	-
	ATGCATT	CTATAA**ATGCATT**TC[a]	HMG-TATA region Confers high levels of transcription
	ATTNAT	CA**ATWAT**TG[a]	HD-Zip1 Control of development
	CAAGAAGA	No match in database	-
	CMKACMC	No match in database	-
	CTTGT	No match in database	-
	GRNTC	CGC**GGATC**[a]	OCT element Meristem specific activation
14	AAATTTAATT	***ATTAAA***[b]	Poly-A signal Found in rice alpha amylase
	CTTTCAC	No match in database	-
	TNACTG	**TAACTG**[a]	MBS MYB binding site Drought response
15	AAGGNAAG	No match in database	-
	GGTAARWGT	No match in database	-
	GTTGATT	**NGATT**[b]	ARR1-binding element Response regulator
	TGATNGTTG	No match in database	-
16	AATTGA	AGC***TCAATT***TCA[a]	CAAT-box Common *cis*-acting element (*D. carota*)
	GATTCA	No match in database	-
	GTGWGA	***TCTCA***CCAACCCCA[a]	4cl-CMA2b Light responsive element (*P. crispum*)
		TCTCACCTACC[a]	L-box Light responsive element (*P. crispum*)

	TCTCACCAACC[a]	AC-I element Regulation of xylem and phloem expression (*P. tremuloides*)
GTTTAAG	No match in database	-
TTGYCGT	No match in database	-

A number of potential DNA motifs identified in the promoters of co-expressed genes showed homology to previously identified motifs in the PlantCARE and PLACE databases. The majority of the identified motifs were involved in light and stress response. A number of motifs responsive to the transcriptional activator ARR1 were also identified (Sakai et al., 2000).

4.5 Discussion

4.5.1 Identification of sucrose-responsive genes in Col-7 and *acn1-2*

A combination of fold-change threshold in gene expression levels and a t-test ($p < 0.05$) was used to identify genes differentially expressed in response to germination on sucrose in both Col-7 and *acn1-2*. Sucrose is well established as a regulator of gene expression and unsurprisingly, a large number of genes were observed to change expression in response to sucrose in both Col-7 and *acn1-2* (Koch, 1996). Approximately 13% of the genes found on the Affymetrix ATH1 microarrays were identified as sucrose responsive; a group of 3580 sucrose responsive genes was identified for Col-7 (1.3-fold change in expression) and 3255 genes identified for *acn1-2* (1.4-fold change in expression). As no osmotic control was made, it should be noted that the differentially expressed gene set identified in response to growth on high concentrations of sucrose in this experiment contains genes responsive to osmotic changes as well as sucrose. From this, all genes affected by the presence of high levels of sucrose were identified. The growth of seedlings with an osmotic control, such as mannitol which is not metabolised, would provide further information on which genes are specifically responsive to the presence of sucrose.

The behaviour of the sucrose-responsive genes differed between Col-7 and *acn1-2*; of the 3580 sucrose-responsive genes in Col-7 35% were induced and 65% were repressed by sucrose. This profile was reversed in *acn1-2* where the majority of sucrose-responsive genes were induced (59%) and 41% were repressed by germination on sucrose. As the Col-7 and *acn1-2* seedlings were harvested at the same principal growth stage, 1.0 (Boyes et al., 2001) suggests that a developmental response is not primarily responsible for the observed differences in the transcriptome profiles. A potential overlap between acetate and sugar signalling has been suggested by the co-ordinate regulation of seven photosynthetic genes in a maize protoplast system by sucrose, glucose and acetate (Sheen, 1990). The altered phenotypes of the related *acn2* and *acn3* Arabidopsis mutants indicate that a disruption in acetate metabolism influences carbohydrate perception in germinating seedlings (Hooks et al., 2004). This suggests that differences in the observed changes of the response to sucrose in Col-7 and *acn1-2* is a result of a specific disruption to acetate metabolism.

4.5.2 Genes differentially expressed in response to growth on sucrose in Col-7 and *acn1-2*.

The top ten up and down regulated genes in Col-7 displayed greater fold-changes in expression levels than those in *acn1-2*; fold-changes were approximately 19 times higher for the top ten up regulated and 47 times higher for the top ten down regulated genes in Col-7 compared to *acn1-2*. A regulatory role for acetate in the transition from heterotrophic to autotrophic metabolism in seedlings has been proposed (Sheen, 1990). In this regulatory capacity, acetate interactions with other factors such as sucrose might control developmental progression. The decreased changes in gene expression levels in *acn1-2* indicate that the disruption to acetate metabolism alters the perception or the signalling mechanism of sucrose. The lower levels of acetate observed in *acn1-2* suggest a possible involvement of acetate in sugar sensing or signalling during germination.

The gene ontology analysis tool available at TAIR was used to determine the functional characterisation of the genes with the largest increases and decreases in expression levels (Berardini et al., 2004). A common response to growth on

high concentrations of sucrose is an increase in anthocyanin content and induction of genes involved in anthocyanin biosynthesis in Arabidopsis (Solfanelli et al., 2006; Luo et al., 2007). The largest change in gene expression was observed in Col-7 (1375-fold increase) for a gene encoding an enzyme with anthocyanidin 3-O-glucoside coumaroylCoA transferase activity previously shown to be significantly induced in response to growth on sucrose (Luo et al., 2007). A further two genes were identified in the Col-7 up-regulated genes with a role in anthocyanin biosynthesis and metabolism. The sucrose responsive genes exhibiting the largest increases in expression also included two genes related to carbohydrate binding and signalling. The most up-regulated genes in response to sucrose in *acn1-2* did not include genes associated with anthocyanin biosynthesis, suggesting that response to sucrose might be altered in *acn1-2*.

The genes with the largest decreases in gene expression in Col-7 and *acn1-2* both showed associations with hormonal regulators of development. Germination of Col-7 on sucrose repressed genes associated with positive regulators of seed germination and development; an auxin-responsive gene and an ethylene biosynthetic gene showed decreases in expression levels. Conversely, the down-regulated genes in *acn1-2* included a gene associated with abscisic acid biosynthesis, a negative regulator of germination. This suggests a possible regulatory role for acetate levels, in conjunction with other factors such as sucrose to regulate the progression of development.

4.5.3 Identification of genes that show an altered response to sucrose in Col-7 than in *acn1-2*

A set of sucrose responsive genes was identified in Col-7 using a combination of fold-change threshold and a t-test. These genes were examined to identify a subset of 342 genes with an altered response to sucrose as a result of disrupting acetate metabolism. Hierarchical and *k*-means clustering were used to produce twelve clusters of genes demonstrating the variation in the responses of gene expression over the four sample types (figure 4.5). A number of gene expression patterns were observed in the data; a predominant feature was that a disruption to gene expression levels by the mutation was often restored to wild-type levels by

the addition of sucrose (as seen in clusters 4, 5, 10, 13, 14, 15 and 16). This suggests that carbohydrate signalling has the potency to overrule disruptions to acetate signalling in seedlings. A number of the clusters present themselves for further investigation. Cluster 1 showed the expression levels of a set of genes which were repressed in response to sucrose in the wild-type. The mutation altered the response to sucrose by repressing the extent of gene expression changes. In cluster 4, sucrose repressed wild-type genes to a small extent but the mutation enhanced the response to sucrose, repressing gene expression further. The genes in cluster 12 are induced in response to sucrose in the wild-type seedlings but show little or no response to sucrose in the mutant. Further investigation of the genes in these clusters might provide further information on interactions between organic acid and sugar signalling pathways.

The variation in gene expression responses suggests that the disruption to acetate metabolism may have an effect on a number of different signalling pathways. The promoter sequences of each group of co-expressed genes were analysed for the presence of over-represented DNA motifs (table 4.3). Unsurprisingly, a large proportion of the identified motifs were involved with carbohydrate signalling, or seed specific activities. Motifs identified as responsive to the transcriptional activator ARR1 were identified in a number of clusters of genes, indicating an involvement of ARR1 in the sucrose-response. The alteration in expression levels of sucrose-responsive genes in *acn1-2* and the presence of a large number of elements identified as responsive to light, development, hormones and stress suggests that organic acid signalling mechanisms cross-talk with sugar, stress and/or developmental signalling pathways. A number of the potential DNA motifs identified showed no homology to promoter elements in the databases. These motifs might be involved in regulation of the transcriptional response to acetate or sucrose signalling. The absence of identified promoter motifs in these co-expressed genes suggests that if the genes are regulated at the transcriptional level then the motifs identified are strong candidates for the binding of regulatory factors.

Chapter 5

Metabolite and transcript profiling during germination and seedling development

5.1 Introduction

The transition from imbibed seed to photosynthetically competent seedling requires a large-scale reorganisation of metabolism. A number of experiments have described the metabolic and transcriptomic switches that occur during seed maturation and germination (Girke et al., 2000; Fait et al., 2006). The metabolic regulation of gene expression is well described for a range of metabolites including amino acids, sugars and organic acids (Monroy and Schwartzbach, 1984; Kindle, 1985; Sheen, 1990; Grierson et al., 1994; Koch, 1996; Tischner, 2000; Corruzzi and Bush, 2001; Gibson, 2005). It is possible that the changes in metabolism may act to regulate transcriptional changes during development. The production of metabolite and transcript profiles during germination and seedling establishment may provide further information about changes in metabolic content and gene expression during development. Metabolite and transcriptome profiles were produced using ^{1}H-NMR and microarrays respectively for distinct stages from seed imbibition to the establishment of photosynthetically competent seedlings (figure 5.1). The profiles produced were used to characterise the metabolic and transcriptomic states at each time point sampled.

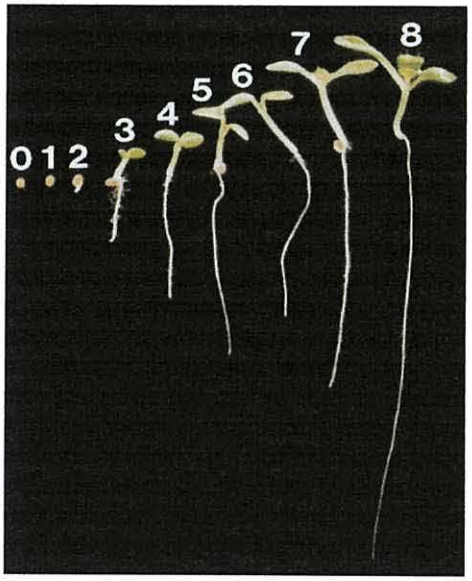


Figure 5.1: Growth stages of Arabidopsis (ecotype Col-0) seeds and seedlings harvested for metabolic and transcriptome profiling. Seeds were imbibed in the cold and the dark for four days before transfer to continuous light. Samples for the first time point (0) were harvested one hour after transfer to light and subsequent samples were taken at 24 hour time intervals until 8 days old. Photograph reproduced from Rylott et al. (2001).

Seeds were imbibed in the dark at 4 °C for four days prior to transfer to continuous light at 20 °C. Growth under continuous light minimises the effects of circadian related processes on the results allowing developmental effects to be observed more clearly in discrete steps from imbibition to seedling establishment. Arabidopsis ecotype Col-0 was used in these experiments as it demonstrates a non-deep physiological dormancy which is removed by a sufficient after-ripening period of approximately 3 weeks (Derkx and Karssen, 1993). Samples for the first time point (day 0) were harvested one hour after transfer to light conditions. During vernalisation the metabolites accumulated during seed desiccation are consumed, preparing the seed for germination (Fait et al., 2006). Although few physiological changes are apparent in days 0 and 1 this period, prior to the emergence of the radicle, is associated with a burst of metabolic activity; a number of processes including protein and mitochondria synthesis and the repair of membranes damaged during desiccation are initiated (Bewley, 1997; Fait et al., 2006). The visible signs of termination of the germination process are apparent by day 2 with the penetration of the seed coat

by the radicle which continues to expand as the first root throughout development (Bewley and Black, 1994). Post-germinative growth involves mobilisation of seed storage reserves, elongation of the radicle, cell division and the synthesis of DNA, proteins and mitochondria (Bewley, 1997). It has been demonstrated that the transition of cotyledons from a storage tissue to a photosynthetic tissue occurs between 48 and 60 hours after imbibtion in Arabidopsis (Mansfield and Briarty, 1996); the visible signs of which (emergence, unfurling, enlargement and greening of cotyledons) are not apparent until day 4 in this time series. This is most likely due to the differences in growth conditions between 20 °C used in this experiment and 25 °C used by Mansfield and Briarty (1996), illustrating the dramatic effects of temperature fluctuations on development. The shoot apical meristem is the region of shoot located between the cotyledons and becomes active during photomorphogenesis, producing the first true leaves and allowing postembryonic development to proceed (Howell, 1998). Under these growth conditions, 3 and 4 day-old seedlings were approximately at Principal Growth Stage 0.7 and 1.0, respectively (Boyes *et al.*, 2001). Since Boyes *et al.* (2001) do not strictly specify a developmental stage for each time point selected in this experiment, seedling development is referred to by day.

5.2 Metabolite profiling throughout development

A principal component analysis (PCA) plot was produced in order to investigate groupings of sampled time points based on measured metabolite levels at each time point (figure 5.2). It was seen that a cluster was formed by samples from days 2 to 8. A general curvature in the data was observed, indicating that a gradual and directional change in metabolite levels was occurring from day 2 to day 8. It should be noted that PCA plots such as these are open to interpretation by individual users. The clustering observed in this plot is not particularly clear and so relationships will be described in more detail in the following chapter. In order to identify the metabolites with a significant difference in measured levels between days 1 and 2, i.e. those which contributed most to the observed

groupings a t-test was applied ($p > 0.05$). A significant difference was observed for the levels of fructose, proline, glutamine and the unknown peaks D5.69 and Q5.18 (data not shown).

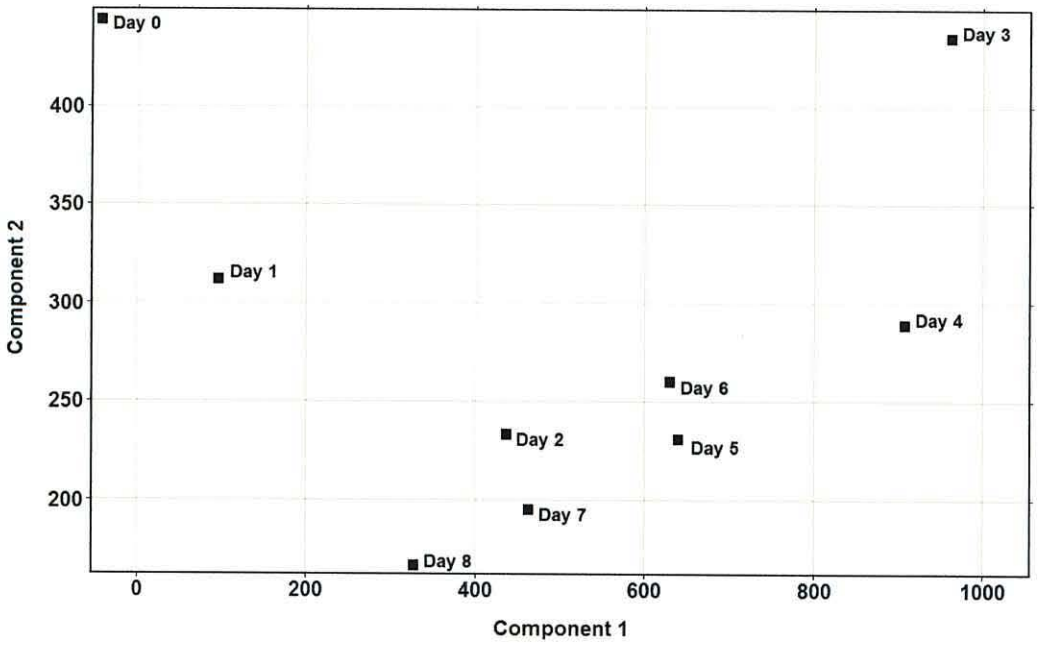
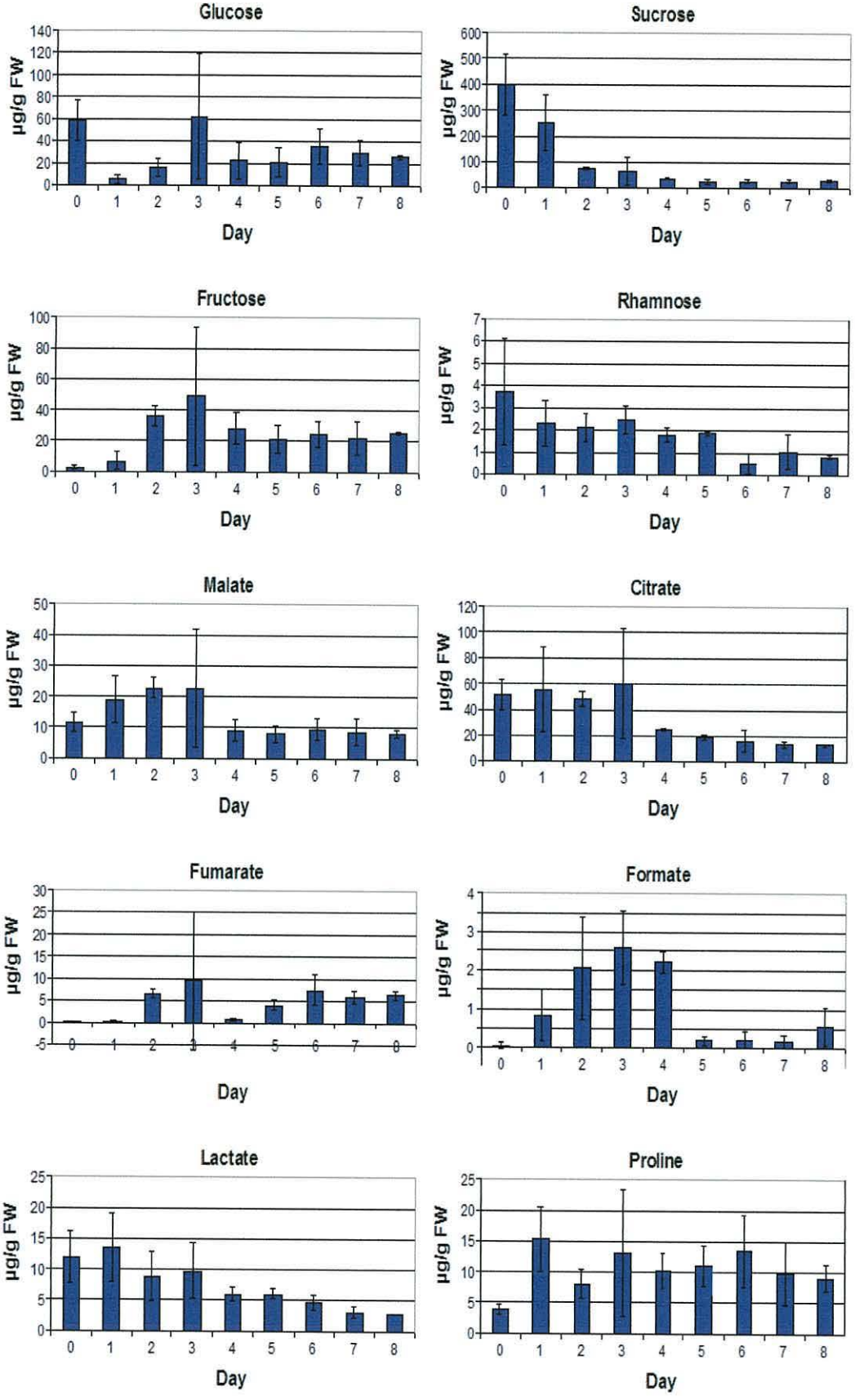


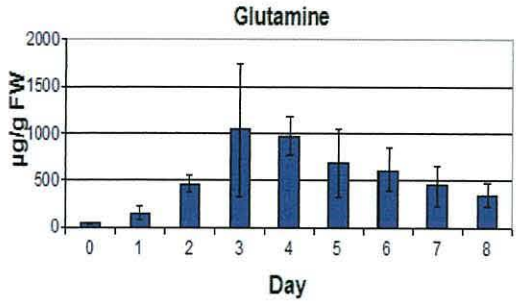
Figure 5.2: Principal component analysis of the time points sampled during germination and seedling establishment based on metabolite levels. Principal component 1 and principal component 2 describe 89.8% and 7.1% of the variance, respectively. It can be seen that samples from days 2 to 8 form a grouping which excludes samples from days 0. Samples from day 1 also tend to be excluded from the cluster.

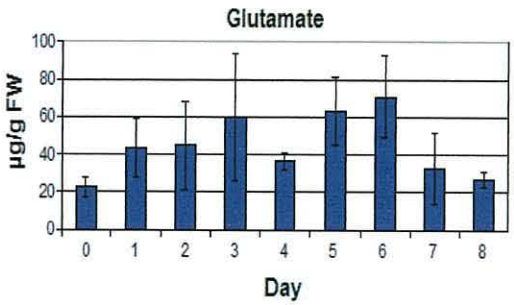
5.2.1 Variation in metabolite levels over time

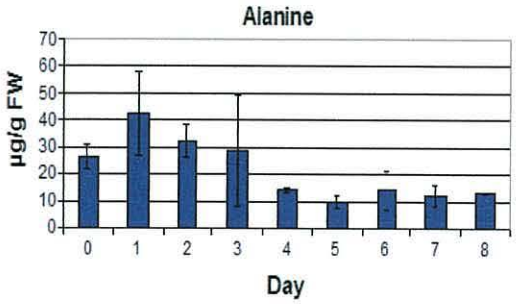
A total of 27 peaks were identified from the ^{1}H-NMR spectra corresponding to a variety of known and unknown metabolites including soluble carbohydrates, amino acids and organic acids. A number of metabolites decreased throughout the time period while others showed an increase. Biphasic profiles (i.e. profiles which increased then decreased over time or *visa versa*) were observed for a number of metabolite profiles which increased to a certain point then decreased again. The levels of some metabolites showed no overall change during the developmental phase (figure 5.3). Metabolites with an observed alteration in levels are more likely to be involved in signalling during development as a

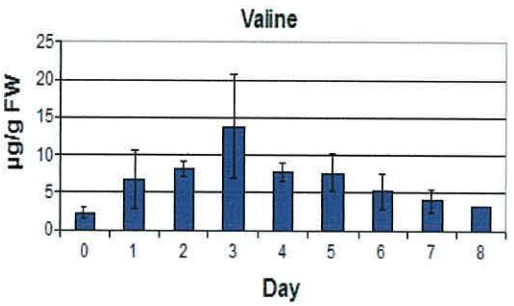
change is effected, while the metabolites with no overall change in level are probably relatively unimportant with respect to signalling during germination and seedling establishment.

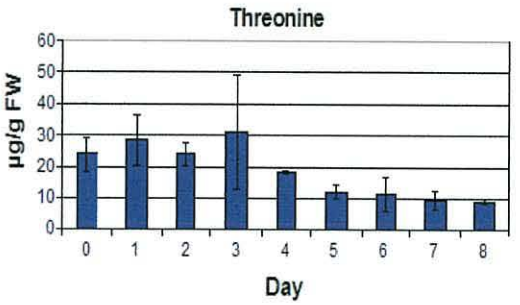

Glucose
µg/g FW
Day
Sucrose
µg/g FW
Day
Fructose
µg/g FW
Day
Rhamnose
µg/g FW
Day
Malate
µg/g FW
Day
Citrate
µg/g FW
Day
Fumarate
µg/g FW
Day
Formate
µg/g FW
Day
Lactate
µg/g FW
Day
Proline
µg/g FW
Day

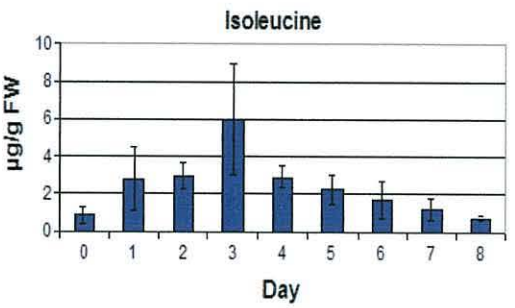

Glutamine
µg/g FW
2000
1500
1000
500
0
0 1 2 3 4 5 6 7 8
Day

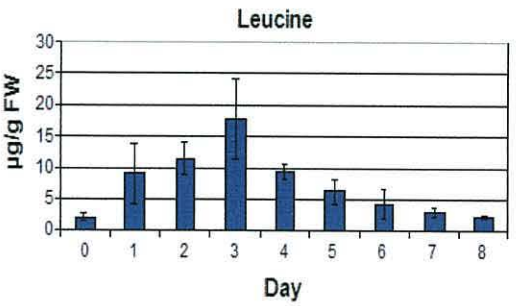

Glutamate
µg/g FW
100
80
60
40
20
0
0 1 2 3 4 5 6 7 8
Day

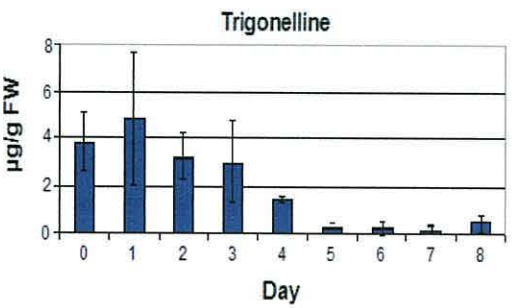

Alanine
µg/g FW
70
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8
Day

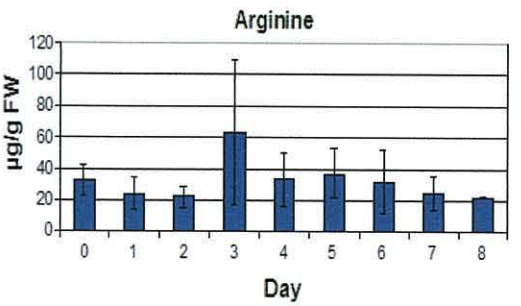

Valine
µg/g FW
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8
Day

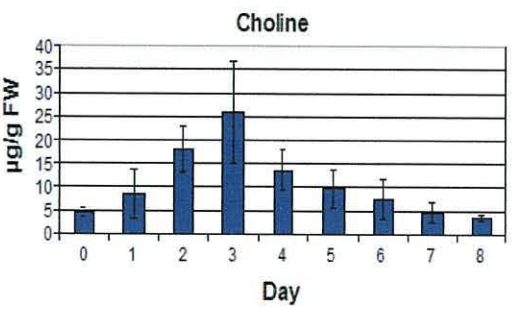

Threonine
µg/g FW
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8
Day


Isoleucine
µg/g FW
10
8
6
4
2
0
0 1 2 3 4 5 6 7 8
Day


Leucine
µg/g FW
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8
Day


Trigonelline
µg/g FW
8
6
4
2
0
0 1 2 3 4 5 6 7 8
Day


Arginine
µg/g FW
120
100
80
60
40
20
0
0 1 2 3 4 5 6 7 8
Day


Choline
µg/g FW
40
35
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8
Day

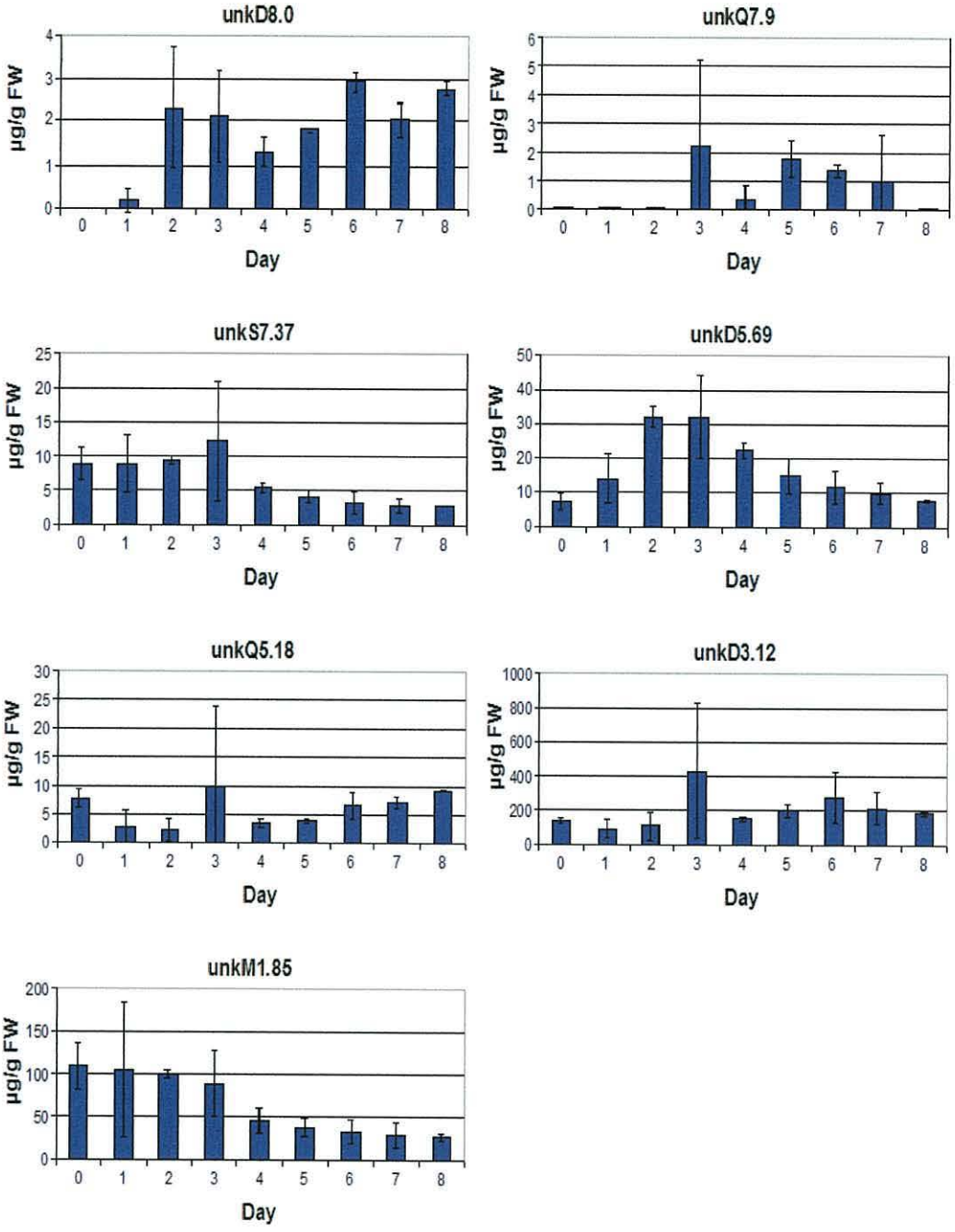


Figure 5.3: The varying concentrations of metabolite levels throughout germination. The letters and numbers of unknown peaks represent the type of signal and the chemical shift relative to TMS: D, doublet; M, multiplet; Q, quintuplet; S, singlet.

The metabolites measured in this experiment were clustered in order to identify the predominant behavioural profiles present in the dataset using GeneSight data analysis software (BioDiscovery, Inc.). The approximate number

of clusters formed by the dataset was determined using hierarchical clustering (figure 5.4).

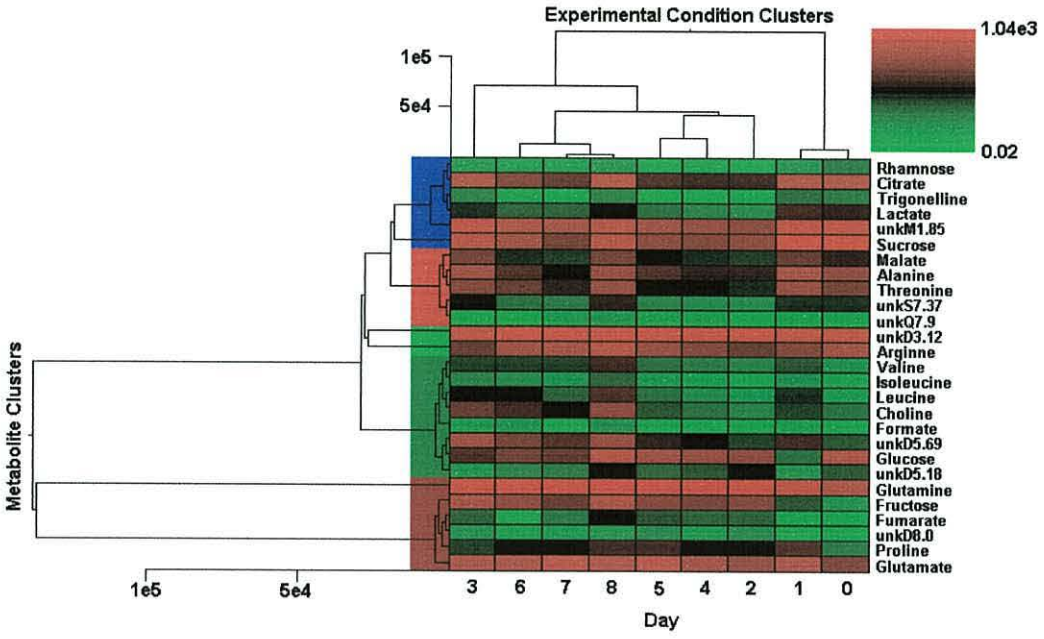


Figure 5.4: Hierarchical clustering to identify the number of clusters formed by the metabolites measured in this experiment. The horizontal clusters show the grouping of metabolites into approximately five clusters (denoted by coloured blocks). The horizontal clusters demonstrate the grouping of time points based on the metabolite levels measured for each day. It can be seen that days 0 and 1 group together while profiles from days 3 to 8 form a second grouping.

Hierarchical clustering of the 27 metabolites measured over time indicates that five groupings are formed throughout development. The sample time points are seen to form two main groupings with days 0 and 1 clustering together and a second group formed from samples from days 3 to 8. A *k*-means clustering algorithm was used to produce groups of genes for further analysis using the information obtained from hierarchical clustering ($k = 5$). A Pearson correlation distance metric was used so that genes with similar expression profiles grouped together (figure 5.5).

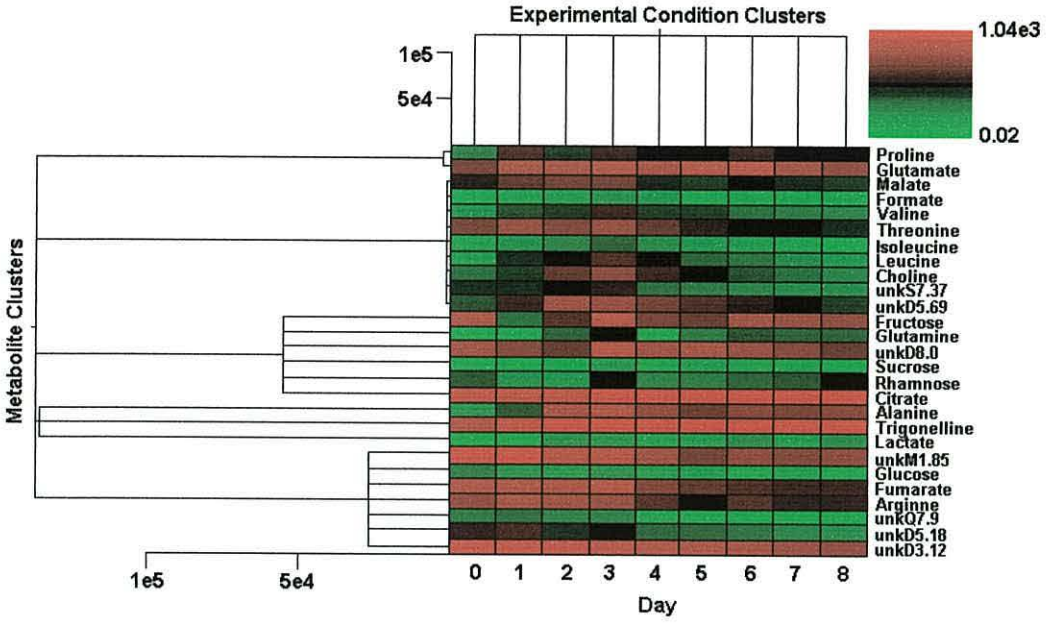


Figure 5.5: *K* – means clustering ($k = 5$) of the 27 metabolites measured during development was used to produce groups of metabolites for further analysis.

The metabolites assigned to each of the five clusters identified through a combination of hierarchical and *k*-means clustering were visualised (figure 5.6). The metabolite levels are shown for each time point sampled and the representative profile for each cluster is displayed as a thick black line running through each cluster.

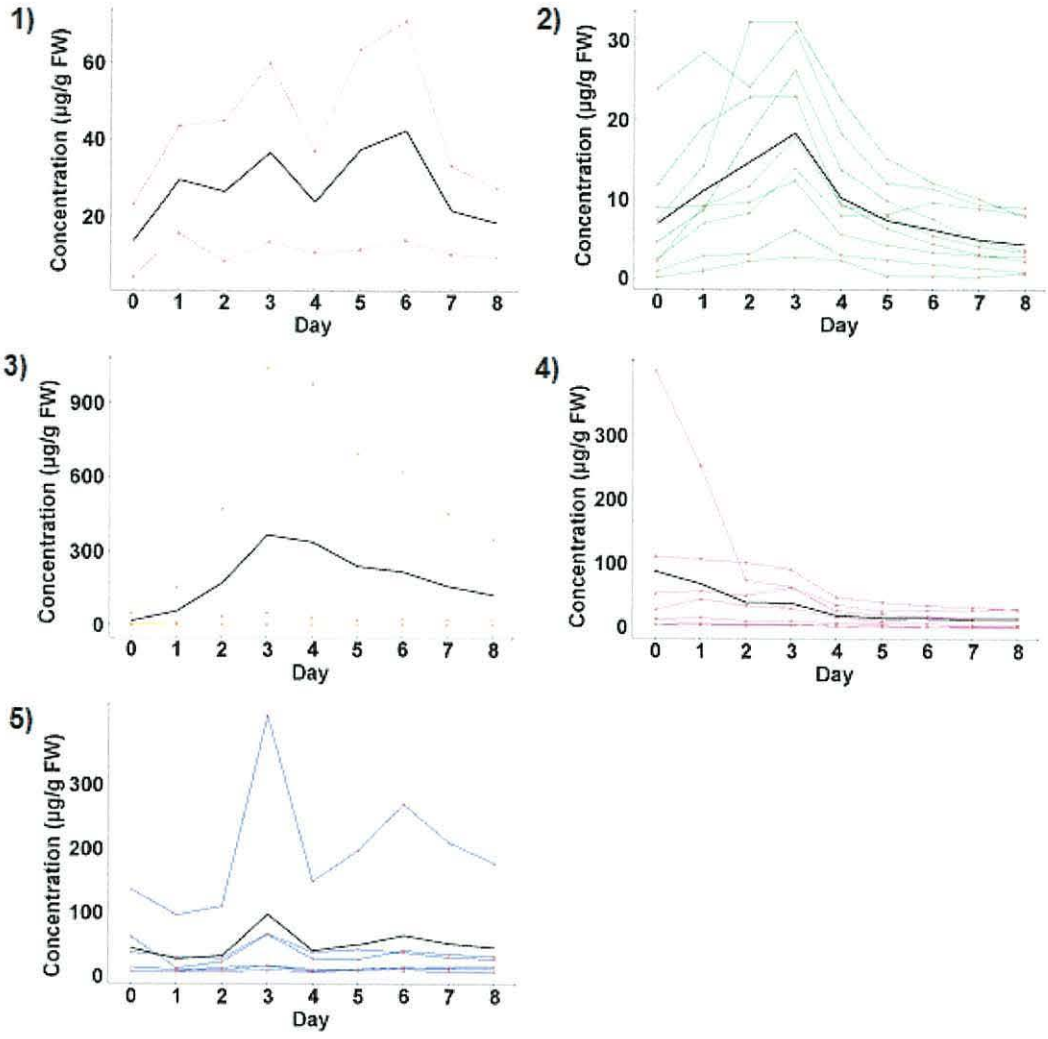


Figure 5.6: The profiles of the metabolites present in each of the five clusters are displayed for each of the time points sampled. The average profile for each cluster is presented as a black line running through each cluster: 1) proline and glutamate levels fluctuate but show little overall change; 2) the levels of malate, formate, valine, threonine, isoleucine, leucine, choline, unkS7.37 and unkD5.69 gradually increase until day 3 and then fall sharply; 3) fructose, glutamine and unkD8.0 increase sharply until day 3 before gradually decreasing; 4) the levels of sucrose, rhamnose, citrate, alanine, trigonelline, lactate and unkM1.85 decrease over time; 5) glucose, fumarate, arginine, unkQ7.9, unkQ5.18 and unkD3.12 levels show a peak around day 3 and a smaller peak around day 6.

The five primary patterns of variation in the measured metabolites throughout development are illustrated in figure 5.6. It should be noted that the average profile for each cluster (represented by a thick dark line running through the plot) does not represent the behaviour of a specific metabolite, but describes the *trend* of the profiles in that plot. Proline and glutamate levels (cluster 1) increase post-

imbibition and fluctuate throughout development with no overall change. The levels of malate, formate, valine, threonine, isoleucine, leucine, choline and the unidentified peaks unkS7.37 and unkD5.69 (cluster 2) steadily increase until day 3 and then decrease sharply. Fructose, glutamine and the unidentified peak unkD8.0 (cluster 3) increase sharply until day 3 then show a gradual decline. Sucrose, rhamnose, citrate, alanine, trigonelline, lactate and the unidentified peak unkM1.85 (cluster 4) decrease throughout development. Glucose, fumarate, arginine and the unidentified peaks unkQ7.9, unkQ5.18 and unkD3.12 (cluster 5) peak around day3 and show another, smaller peak around day 6.

5.3 Gene expression profiling throughout development

In order to identify genetic switches during germination and seedling development, PCA was used to group the time points sampled on the basis of transcriptome profiles (figure 5.7). The PCA plot showed that when samples were grouped on the basis of gene expression levels the data may be interpreted that samples from days 3 to 8 formed a cluster which excluded samples produced from days 0, 1 and 2, although these groupings are not very clear and are open to user-interpretation. Samples from days 0, 1 and 2 did not appear to group together to a great extent. Relationships in this dataset are examined in more detail in chapter 6. Microarrays were also produced from RNA samples pooled from days 0, 2, 4, 6 and 8 throughout the time course. As might be expected, these samples group away from the majority of individual time points.

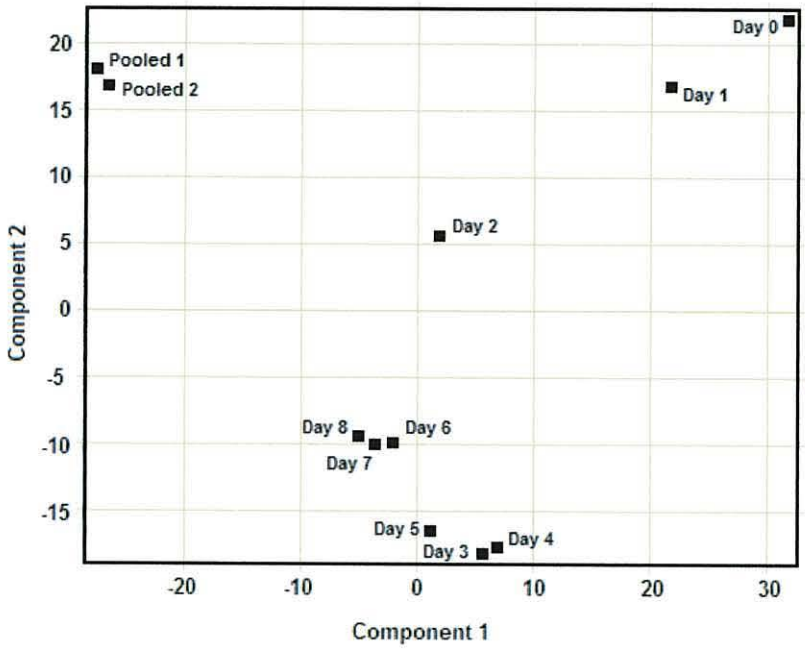


Figure 5.7: Principal component analysis of the time points sampled during germination and seedling establishment based on gene expression values. Principal component 1 and principal component 2 describe 50.1% and 23.7% of the variance, respectively. It can be seen that samples from days 3 to 8 form a grouping which excludes samples from days 0, 1 and 2. Due to the large computational power required to process the dataset, replicated gene expression values were combined prior to analysis.

In order to identify gene expression events which contribute to the observed groupings of samples over time, differentially expressed genes were examined between days 0, 1, 2 and 3 sequentially. A 2-fold increase or decrease in gene expression was set as the threshold for differential expression. The main focus of the investigation was to identify the genes most relevant to the observed grouping of gene expression profiles. Consequently the top thirty most up and down regulated genes were examined between sequential time points as these contributed most to the observed grouping of profiles. The top thirty differentially up and down regulated genes for day 0 vs. day 1, day 1 vs. day 2 and day 2 vs. day3 are presented in tables 5.1, 5.2 and 5.3 respectively. The MIPS gene identifier and gene title were obtained from the TAIR online database (available at www.arabidopsis.org) and are provided for each gene along with the corresponding fold change in expression.

Table 5.1: The top thirty up and down regulated genes identified between transcriptome profiles produced on days 0 and 1. Each gene is listed with the relevant MIPS identifier and gene title. The corresponding fold-changes in expression are listed (table continues on page 141).

Genes down regulated in day 1 compared to day 0		
Gene ID	**Gene Name**	**Fold Change**
At3g15670	late embryogenesis abundant protein, putative	118.7
At4g27160	2S seed storage protein 3	63.0
At5g54450	expressed protein	53.3
At2g02120	plant defensin-fusion protein, putative	46.1
At4g28520	12S seed storage protein, putative / cruciferin, putative	42.4
At5g44120	12S seed storage protein (CRA1)	40.2
At2g40170	Em-like protein GEA6: Small hydrophilic plant seed protein (group 1 late embryogenesis abundant protein family)	34.1
At3g53230	cell division cycle protein 48, putative	26.0
At2g26020	plant defensin-fusion protein, putative	24.6
At5g54740	protease inhibitor/seed storage/lipid transfer protein family protein, similar to 2S seed storage proteins	23.7
At4g27170	2S seed storage protein 4	19.8
At1g65980	peroxiredoxin type 2, putative,	19.0
At1g20790	F-box family protein	18.4
At1g04560	AWPM-19-like membrane family protein; similar to late embryogenesis abundant protein	18.1
At1g14930	major latex protein-related / MLP-related: Pathogenesis-related protein Bet v I family	17.7
At1g27400	60S ribosomal protein	15.9
At5g06360	ribosomal protein S8e family protein	15.8
At5g62490	ABA-responsive protein (HVA22b)	15.4
At1g11760	expressed protein, weak similarity to 4'-phosphopantetheinyl transferase superfamily	15.3
At2g31980	cysteine proteinase inhibitor-related	14.5
At5g04340	zinc finger (C2H2 type) family protein,	14.3
At5g65660	hydroxyproline-rich glycoprotein family protein	13.5
At3g51880	high mobility group protein alpha (HMGalpha)	13.3
At2g23580	hydrolase, alpha/beta fold family protein, similar to ethylene-induced esterase	12.4
At2g33070	jacalin lectin family protein, similar to myrosinase-binding protein homolog	12.4
At1g52690	late embryogenesis abundant protein, putative	12.3
At3g12490	cysteine protease inhibitor, putative / cystatin, putative	12.3
At1g70840	Bet v I allergen family protein: Pathogenesis-related protein Bet v I family	11.9
At4g25140	glycine-rich protein / oleosin	11.8
At5g25900	ent-kaurene oxidase, putative (GA3) / cytochrome P450	11.5

Genes up regulated in day 1 compared to day 0		
Gene ID	**Gene Name**	**Fold Change**
At2g18020	60S ribosomal protein L8	47.3
At3g24480	leucine-rich repeat family protein / extensin family protein	45.4
At5g13430	ubiquinol-cytochrome C reductase iron-sulfur subunit, putative	27.6
At5g60120	AP2 domain-containing transcription factor, putative	27.4
At1g54010	myrosinase-associated protein, putative	26.0
At2g24980	proline-rich extensin-like family protein,	23.0
At3g22070	proline-rich family protein,	17.5
At4g39820	expressed protein	17.1
At5g60840	expressed protein	15.8
At3g42960	alcohol dehydrogenase	15.4
At2g21160	translocon-associated protein alpha family protein	15.3
At5g35160	endomembrane protein 70, putative	15.0
At5g09750	basic helix-loop-helix family protein	14.4
At4g26060	expressed protein	14.3
At3g58570	DEAD box RNA helicase, putative	13.6
At4g11010	nucleoside diphosphate kinase 3, mitochondrial	13.6
At5g05290	expansin, putative	12.8
At5g05090	myb family transcription factor	12.5
At5g13650	elongation factor family protein	12.3
At2g15830	expressed protein	12.0
At2g38540	nonspecific lipid transfer protein 1	11.8
At3g18490	aspartyl protease family protein	11.4
At5g46180	ornithine aminotransferase, putative	11.3
At4g30190	ATPase 2, plasma membrane-type, putative / proton pump 2	11.0
At1g09210	calreticulin 2	10.7
At2g30470	B3-domain family protein	10.7
At1g71880	sucrose transporter / sucrose-proton symporter	10.5
At4g08400	proline-rich extensin-like family protein	10.4
At5g66920	multi-copper oxidase type I family protein	10.1
At4g20110	vacuolar sorting receptor, putative	10.1

Of the nineteen down-regulated genes with an assigned identity in the TAIR database in day 1 compared to day 0, six showed an involvement in seed storage (At4g27160, At4g28520, At5g44120, At4g27170, At4g25140 and At5g54740). Four late embryogenesis abundant related proteins (At3g15670, At2g40170, At1g04560 and At1g52690) were identified, and three genes associated with known regulators of seedling development were found; light (At5g25900), ethylene (At2g23580) and abscisic acid (At5g62490). Four genes involved with defence were down regulated over time; two involved in pathogenesis response (At1g14930 and At1g70840) and two defensin-fusion proteins (At2g02120 and At2g26020). Genes involved in general metabolism such as ribosomal proteins

(At1g27499 and At5g06360) and transcriptional regulators (At1g20790 and At5g04340) were also listed.

Sixteen of the top thirty up regulated genes in day 1 compared to day 0 had an assigned identity in the TAIR database. Genes with a role in cell wall formation are identified; three extensin-related proteins (At3g24480, At2g24980 and At4g08400), one elongation factor family protein (At5g13650) and a putatuive expansin protein (At5g05290) were identified. Genes involved in lipid metabolism were identified; a myrosinase-associated protein and a lipid transfer protein were up regulated in day 1 compared to day 0. Three transcription factors (At5g60120, At5g09750 and At5g05090) and a ribosomal gene (At2g18020) were also up regulated in day 1. The proton-pump 2 gene (At4g30190) has been associated with the active transport of nutrients by proton symport. Two genes involved with oxidative stress response (At4g11010, At1g09210) and two genes associated with sucrose response (At2g30470, At1g71880) were identified.

Table 5.2: The top thirty up and down regulated genes identified from a comparison between transcriptome profiles produced on days 1 and 2. Genes are identified by MIPS numbers and gene titles. The corresponding fold-changes in expression are listed (table continues on page 144).

Genes down regulated in day 2 compared to day 1		
Gene ID	**Gene Name**	**Fold Change**
At3g13920	eukaryotic translation initiation factor 4A-1 / DEAD/DEAH box helicase	45.2
At5g17950	hypothetical protein	28.6
At3g61870	expressed protein, hypothetical protein	25.1
At3g05165	sugar transporter, putative, similar to sugar-porter family proteins 1 and 2	20.3
At1g55370	expressed protein	15.3
At5g27600	AMP-binding protein, putative	15.1
At2g02040	peptide transporter (PTR2-B) / oligopeptide transporter 1-1, putative	13.6
At4g13940	adenosylhomocysteinase / S-adenosyl-L-homocysteine hydrolase / AdoHcyase	13.0
At5g54680	basic helix-loop-helix family protein	12.5
At3g58140	phenylalanyl-tRNA synthetase class IIc family protein	12.3
At5g67510	60S ribosomal protein L26	9.3
At4g24770	31 kDa ribonucleoprotein	9.2
At1g02305	cathepsin B-like cysteine protease, putative	9.1
At1g26770	expansin, putative (EXP10)	8.8
At1g64510	ribosomal protein S6 family protein	7.5
At1g33240	trihelix DNA-binding protein, putative	7.4
At1g63770	peptidase M1 family protein	7.4
At2g18020	60S ribosomal protein L8	7.0
At2g17870	cold-shock DNA-binding family protein	6.9
At5g54430	universal stress protein family protein	6.9
At1g01620	plasma membrane intrinsic protein 1C (PIP1C) / aquaporin PIP1.3 (PIP1.3) / transmembrane protein B (TMPB)	6.6
At3g04840	40S ribosomal protein S3A	6.5
At4g03210	xyloglucan:xyloglucosyl transferase, putative / xyloglucan endotransglycosylase, putative	6.3
At5g27770	60S ribosomal protein L22	6.3
At1g06040	zinc finger (B-box type) family protein / salt-tolerance protein	6.2
At5g06320	harpin-induced family protein / HIN1 family protein / harpin-responsive family protein / NDR1/HIN1-like protein 3	6.2
At5g08680	ATP synthase beta chain	6.2
At1g24793	UDP-3-0-acyl N-acetylglucosamine deacetylase family protein / F-box protein-related	6.1
At1g02280	GTP-binding protein	6.1
At3g17810	dihydroorotate dehydrogenase family protein / dihydroorotate oxidase family protein	6.1

Genes up regulated in day 2 compared to day 1

Gene ID	Gene Name	Fold Change
At4g23650	calcium-dependent protein kinase, putative	73.2
At4g36880	cysteine proteinase, putative	44.1
At1g33790	jacalin lectin family protein, similar to myrosinase binding protein homolog	38.3
At1g27400	60S ribosomal protein L17	33.0
At5g64100	peroxidase, putative	31.7
At1g31240	expressed protein	29.4
At3g16780	60S ribosomal protein L19	27.5
At5g46110	phosphate/triose-phosphate translocator, putative	23.7
At1g28510	expressed protein	20.8
At2g18540	cupin family protein	20.3
At4g24190	shepherd protein (SHD) / clavata formation protein, putative,	19.4
At3g12630	zinc finger (AN1-like) family protein	19.2
At5g47700	60S acidic ribosomal protein P1	18.1
At2g43150	proline-rich extensin-like family protein	17.2
At2g40590	40S ribosomal protein S26 (RPS26B)	17.1
At2g45240	methionyl aminopeptidase, putative / methionine aminopeptidase, putative / peptidase M, putative	16.8
At4g37120	expressed protein	16.5
At2g27760	tRNA isopentenyltransferase 2 / IPP transferase 2 (IPT2)	16.3
At4g33680	aminotransferase class I and II family protein	15.6
At3g59440	calcium-binding protein, putative	15.4
At1g74060	60S ribosomal protein L6	15.3
At4g02700	sulfate transporter	15.2
At5g23740	40S ribosomal protein S11	15.1
At1g78420	expressed protein	15.1
At1g73840	hydroxyproline-rich glycoprotein family protein	15.1
At5g20620	polyubiquitin (UBQ4)	15.0
At3g54110	plant uncoupling mitochondrial protein	14.6
At1g09310	expressed protein	14.2
At4g40030	histone H3.2	13.7
At2g17710	expressed protein	13.6

Of the 17 top thirty down regulated genes in day 2 compared to day 1 with an assigned TAIR identity, five encode ribosomal proteins (At5g67510, At1g64510, At2g18020, At3g04840 and At5g27770). Five genes involved in the regulation of translation and transcription were identified (At3g13920, At5g54680, At1g33240, At2g17870 and At1g06040). Two genes involved in the regulation of embryonic development ending in seed dormancy were identified (At4g13940 and At2g18020). Genes involved in the transport and metabolism of carbohydrates were identified (At3g05165 and At4g03210). Interestingly, a number of genes previously identified as associated with chloroplasts were down

regulated in day 2 compared to day 1 (At3g58140, At4g24770, At1g64510 and At1g02280).

Fourteen of the top thirty up regulated genes in day 2 compared to day 1 had an assigned identity in the TAIR database. Six genes were identified that encoded ribosomal proteins (At1g27400, At3g16780, At5g47700, At2g40590, At1g74060 and At5g23740). A number of genes associated with mitochondria or chloroplasts were up regulated in day 2 (At5g64100, At4g24190, At3g54110 and At4g40030). A developmental-associated gene (At2g45240) and a gene involved in cytokinin biosynthesis (At2g27760) were also identified.

Table 5.3: The top thirty up and down regulated genes when the transcriptome profile produced on day 3 was compared to the profile produced on day 2. MIPS numbers are provided with gene names as gene identifiers. The corresponding fold-changes in expression are listed (table continues on page 147).

Genes down regulated in day 3 compared to day 2		
Gene ID	**Gene Name**	**Fold Change**
At4g16240	expressed protein	201.2
At5g20620	polyubiquitin (UBQ4)	123.7
At2g44380	DC1 domain-containing protein,	121.5
At5g05620	tubulin gamma-2 chain / gamma-2 tubulin	97.4
At3g20850	proline-rich family protein	88.3
At1g04480	60S ribosomal protein L23	85.2
At1g48410	argonaute protein	84.1
At1g31130	expressed protein	70.0
At3g09260	glycosyl hydrolase family 1 protein	69.1
At4g15755	C2 domain-containing protein	66.7
At4g11420	eukaryotic translation initiation factor 3 subunit 10 / eIF-3 theta / eIF3a	65.0
At3g53470	expressed protein, ribosomal protein S25	62.6
At2g22250	aminotransferase class I and II family protein	59.4
At3g61650	Required for centrosomal and noncentrosomal microtubule nucleation	56.5
At5g53750	similar to CBS domain-containing protein	53.6
At1g73470	expressed protein	52.9
At2g47570	60S ribosomal protein L18	49.1
At3g52150	RNA recognition motif-containing protein	47.3
At2g37570	expressed protein	44.4
At1g30970	zinc finger (C2H2 type) family protein	42.7
At4g36880	cysteine proteinase, putative	42.1
At3g52730	ubiquinol-cytochrome C reductase-like family protein,	40.0
At1g13440	glyceraldehyde 3-phosphate dehydrogenase, putative / NAD-dependent glyceraldehyde-3-phosphate dehydrogenase, putative	39.9
At3g27570	expressed protein	39.8
At1g31240	expressed protein	38.8
At1g14650	Suppressor-of-White-APricot/surp domain-containing protein / ubiquitin family protein	37.3
At1g36580	2,4-dienoyl-CoA reductase-related	36.1
At2g41650	expressed protein	34.5
At1g68380	expressed protein	34.3
At5g50810	mitochondrial import inner membrane translocase	34.1

Genes up regulated in day 3 compared to day 2		
Gene ID	**Gene Name**	**Fold Change**
At3g12780	phosphoglycerate kinase, putative	29.6
At4g05180	oxygen-evolving enhancer protein 3, putative	28.0
At2g34430	chlorophyll A-B binding protein / LHCII type I	24.7
At3g27690	chlorophyll A-B binding protein (LHCB2:4)	23.2
At3g16240	delta tonoplast integral protein	21.1
At5g17170	rubredoxin family protein	19.2
At3g08940	chlorophyll A-B binding protein (LHCB4.2)	16.7
At5g63790	no apical meristem family protein	16.7
At4g25100	superoxide dismutase (Fe) / iron superoxide dismutase	16.5
At1g31580	expressed protein,	16.1
At5g66570	Encodes a protein which is an extrinsic subunit of photosystem II and which has been proposed to play a central role in stabilization of the catalytic manganese cluster	16.0
At5g54270	chlorophyll A-B binding protein / LHCII type III (LHCB3),	15.8
At5g63570	glutamate-1-semialdehyde 2,1-aminomutase 1 / glutamate-1-semialdehyde aminotransferase 1	15.0
At3g47470	chlorophyll A-B binding protein 4 / LHCI type III CAB-4	14.2
At4g24770	31 kDa ribonucleoprotein, putative	14.0
At1g12900	glyceraldehyde 3-phosphate dehydrogenase, putative / NADP-dependent glyceraldehydephosphate dehydrogenase, putative,	13.7
At5g09660	microbody NAD-dependent malate dehydrogenase	13.6
At5g40890	chloride channel protein (CLC-a)	13.5
At1g19150	chlorophyll A-B binding protein, putative / LHCI type II, putative,	12.8
At3g05880	hydrophobic protein (RCI2A) / low temperature and salt responsive protein (LTI6A)	12.2
At1g01620	plasma membrane intrinsic protein 1C / aquaporin PIP1.3 / transmembrane protein B	11.9
At3g50820	Encodes a protein which is an extrinsic subunit of photosystem II and which has been proposed to play a central role in stabilization of the catalytic manganese cluster	11.6
At3g55800	Encodes the chloroplast enzyme sedoheptulose-1,7-bisphosphatase	11.2
At1g09340	expressed protein	11.0
At2g20260	photosystem I reaction center subunit IV, putative / PSI-E, putative	10.9
At1g51400	photosystem II 5 kD protein	10.9
At5g20630	germin-like protein (GER3)	10.3
At3g54890	chlorophyll A-B binding protein / LHCI type I (CAB),	9.8
At5g32440	expressed protein	9.6
At1g67700	expressed protein	9.6

Twelve of the top thirty genes down regulated in day 3 compared to day 2 had an assigned identity in the TAIR database. Three genes encoding ribosomal proteins were identified (At1g04480, At3g53470 and At2g47570). Three genes involved in the regulation of transcription and translation were identified

(At4g11420, At1g30970 and At1g14650). Two genes required for centrosomal and non-centrosomal microtubule nucleation were down regulated in day 3 (At5g05620 and At3g61650). One of the most abundant proteins in Arabidopsis seedlings was down regulated in day 3 (At3g09260). Genes associated with developmental processes or known developmental regulators were also identified (At1g48410, At3g33350, At3g52730 and At1g13440).

Of the top thirty up regulated genes in day 3 compared to day 2, nineteen of these showed an involvement in photosynthesis; seven chlorophyll A-B binding proteins (At2g34430, At3g27690, At3g08940, At5g54270, At3g47470, At1g19150 and At3g54890), four photosystem I and II associated proteins (At5g66570, At3g50820, At2g20260 and At1g51400) and eight chloroplast or mitochondria associated proteins were up regulated in day 3 (At3g12780, At4g05180, At5g17170, At4g25100, At5g63570, At4g24770, At1g12900 and At3g55800). Two genes involved in the mobilisation of nutrient reservoirs were up regulated (At5g09660 and At5g20630). Three genes with a role in establishing water channels (At3g16240 and At1g01620) and chloride channels (At5g40890) were up regulated in day 3.

5.4 Discussion

5.4.1 Clustering of samples by day differs according to metabolic or transcriptomic content

Principal component plots were used to identify differences between the groupings of sampled time points on the basis of both metabolite and mgene expression levels (figures 5.2 and 5.7). The PCA plot produced based on metabolite levels was organised so that days 0 and 1 were relatively isolated from the remainder of the time points. When gene expression levels were used as the basis for PCA, days 0 and 1 appeared to group more closely to day 2. Although tentative groupings were observed from this analysis, it must be noted that the interpretation of PCA plots is open to user-bias. The PCA analysis raises the possibility that an alteration in metabolic content, at least for the metabolites measured in this experiment, occurs prior to transcriptomic changes. This presents the possibility that the seedling metabolic state has the potential to influence gene expression events either through direct regulation of gene expression or through indirectly influencing signalling pathways. In order for a metabolite to induce changes in gene expression levels, a degree of variation in metabolite levels is required. The observed variation in a number of metabolite levels indicates that this is a possibility. Interactions between gene expression and metabolite levels are investigated further in the next chapter using a spring-embedding algorithm.

5.4.2 Metabolic switches during germination and development

A series of metabolic switches have been described for seed development and germination (Fait et al., 2006). Expanding on this work, characterisation of metabolite levels was performed throughout subsequent developmental stages in order to identify further metabolic switches. A decrease in amino acids, sugars and organic acids during seed maturation implied the incorporation of these metabolites into storage compounds. An accumulation of distinct sugars, fatty

acids and free amino acids was observed during the transition to seed desiccation, suggesting the establishment of a metabolic state which supports the rapid metabolic recovery during seed imbibiton, prior to the consumption of storage reserves. The metabolites accumulated during desiccation were consumed during vernilisation and a burst of metabolic activity was observed during post-imbibition germination (Fait et al., 2006). The uptake of water by the quiescent seed corresponds to a resumption of metabolic activity within the seed. The metabolic apparatus required for this burst of respiratory activity is present in the quiescent seed; the water intake is sufficient to allow the initiation of metabolic activity within the seed (Bewley, 1997). Accordingly, the majority of metabolites measured in this experiment demonstrated an increase in levels in day 1 compared to day 0.

The accumulation of proteins during the mid to late stages of seed maturation and subsequent compartmentalisation in membrane-bound storage bodies is well-characterised (Müntz, 1998). The absence of protein turnover and degradation during this time suggests that during this phase, proteins are protected against proteolytic attack (Madison et al., 1981). The decrease in a number of protease inhibitor gene transcripts in day 1 compared to day 0 indicates that these genes regulate the mobilisation of storage reserves when seeds are exposed to moisture and light. The release of amino acids via the protease-mediated breakdown of storage compounds provides a means for *de novo* protein biosynthesis during germination (Vierstra, 1993; Schaller, 2003). Clustering of metabolite levels indicated that for clusters 2, 3 and 5 a peak in levels was observed around day 3 in the time series, after the termination of germination and prior to the emergence of the cotyledons (figure 5.6). This might indicate that the metabolites measured in these groups are released from storage compounds for use by the seed during germination, after which point normal seedling metabolism is established.

A potential role for the oxidative pentose phosphate pathway (OPPP) in seed germination has been suggested (Botha et al., 1992). NADPH is oxidised during proline biosynthesis, resulting in the maintenance of a high NADP+/NADPH ratio within the cell. The dependence of the OPPP on the availability of $NADP^+$ has suggested that the OPPP may be supported by proline synthesis (Hare and Cress, 1997). The suggested supporting role of proline biosynthesis in OPPP activity, when taken in conjunction with the potential role in seed germination

for the OPPP has led to the suggestion that proline levels within the seed are indirectly able to stimulate germination (Hare et al., 2003). Measurement of endogenous proline levels in germinating Arabidopsis seeds was seen to increase four-fold prior to radicle emergence (Hare et al., 2003); in support of these data, the measured levels of proline in day 1 show a four-fold increase when compared to day 0 (data not shown). K-means clustering of the metabolite profiles shows that the profile produced for glutamate correlates with that of proline (figure 5.10), highlighting the potential for the levels of the amino acids to be interlinked throughout development. It has been suggested that glutamate, as a precursor for proline synthesis, might induce the production of proline and indirectly stimulate germination through the action of pentose pathways (Horii et al., 2007).

Three of the unknown peaks showed an increase in concentration (unkD5.69, unkD8.0 and unkS7.37); three showed a decrease (unkD3.12, unkM1.85 and unkQ5.18) and one showed no change in concentration (unkQ7.9) between days 0 and 1. The unknown peaks with an alteration in levels may well have the potential to act as signalling molecules with a role in the regulation of gene expression during germination.

5.4.3 Genetic switches during germination and seedling establishment

In order to investigate the gene expression events which contributed to the observed grouping of samples (figure 5.8), differentially expressed genes were identified between sequential time points. Of particular interest were the genes responsible for the placement of samples from days 0, 1 and 2 relatively distant from other samples and also the genes surrounding the break point observed in samples over development between days 2 and 3.

5.4.3.1 Genes differentially expressed between day 0 and day 1

Examination of the genes showing the largest increases and decreases between days 0 and 1 showed that a variety of genes contribute to the separation of the two time points. A number of interesting points may be noted from these

changes. A number of genes encoding late embryogenesis abundant (LEA) proteins were down regulated in day 1 compared to day 0 (table 5.1). Correlations between drought stress and LEA transcript levels, coupled with the abundance of LEA transcripts during late embryogenesis has led to a proposed role in the tolerance of seed desiccation (reviewed in Wise and Tunnacliffe, 2004; Tunnacliffe and Wise, 2007). Seeds in this experiment were kept in the cold and dark for a period of four days on agar plates during which time tissues were exposed to moisture, becoming re-hydrated. The rapid decline of LEA gene transcripts within 24 hours post-imbibition (corresponding to day 1 in this experiment) has been described for *Ricinus communis* (Han et al., 1996). A similar observation was made in this experiment where LEA gene transcript levels were maintained at measurable levels until day 1 after which point a number of LEA genes were amongst the most down-regulated in day 1 compared to day 0, indicating that the re-hydration of seed tissue alone is not sufficient to decrease LEA transcript levels. A comparative transcriptome analysis of *acn1* and Col-7 demonstrated that a number of LEA genes expressed in Col-7 seedlings were amongst the most repressed genes in *acn1-2* (chapter 3) in seedlings harvested at principal growth stage 1.0 where cotyledons are fully emerged (Boyes et al., 2001). The maintenance of LEA transcript levels throughout imbibition and post-germination raises the possibility that the role of these genes may be further expanded to involve a developmental stage further than previously described. A protective role has been proposed for LEA proteins based on the correlation of transcript levels with dehydration stress (reviewed in Wise and Tunnacliffe, 2004; Tunnacliffe and Wise, 2007), osmotic stress (Espelund and Jakobsen, 1992; Naot et al., 1995; Xu et al., 1996; Chourey et al., 2003) and cold stress (Sutton et al., 1992; Welin et al., 1994, 1995; Tsuda et al., 2000; Dong et al., 2002; Oliver et al., 2004). It is possible that this proposed role extends beyond the tolerance to seed desiccation to protect the seed throughout germination until the radicle and cotyledons have emerged. Certainly, an alternative regulator to the availability of water for LEA transcript abundance is possible as transcript levels are maintained following re-hydration of the seed tissue.

During mid to late stages of the seed maturation process seed storage genes are expressed so that storage reserves are synthesised for use during germination.

The two main seed storage proteins in Arabidopsis are 12S globulin and 2S albumin proteins referred to as cruciferin and arabin respectively (Heath et al., 1986). A number of genes encoding both 12S and 2S seed storage proteins, along with two genes encoding proteins with a role in seed storage were down regulated in day 1 compared to day 0 (table 5.1). These changes may be related to observed changes in metabolism. An increase in free amino acid content and sugars coincided with the large decrease in seed storage genes. The increase in free amino acids for *de novo* protein synthesis would be expected as storage reserves are depleted (reference) and an increase in sugar content has been shown to repress genes involved in storage (reference), demonstrating the complexity of gene-metabolite networks during development.

The turgor pressure-driven emergence of the radicle from the seed is generally accepted as the visible sign that the germination process has been completed (Bewley, 1997; Cosgrove, 1997; Koornneef et al., 2002; Finch-Savage and Leubner-Metzger, 2006). A number of enzymes with a suggested role in the modification of cell-walls are known to be induced prior to radicle emergence (Voigt and Bewley, 1996; Dahal et al., 1997). The expression of Arabidopsis extension-like (AtEPR1) gene transcripts has been shown in the endosperm of germinating seeds where it has been suggested that AtEPR1 may modify the cell-wall structure either to facilitate or as an indirect effect of radicle protrusion (Dubreucq et al., 2001). Expansins are able to loosen plant cell walls by disrupting hydrogen bonding (Cosgrove, 2000; Lee et al., 2001). A role in the control of seed germination is suggested by the expression of expansins in tomato seeds (Chen and Bradford 2000; Chen et al., 2001) and the induction of a large proportion of expansin gene family members by gibberellic acid, which increases just before radicle emergence (Ogawa et al., 2003). Lending further support to the hypothesised role of cell-wall modifying proteins in radicle emergence an increase in gene transcripts encoding extensin-related, elongation factor, hydroxyproline-rich and expansin proteins was observed in day 1 compared to day 0 (table 5.1). It is difficult to discern whether the up-regulation of these genes facilitates the protrusion of the radicle, or occurs as an indirect result of radicle emergence (Dubreucq et al., 2001). The transcriptome results in this experiment indicate that cell-wall modifications may occur as early as within 24 hours post-imbibition (day 1). Although radicle emergence is not observed

until day 2 in this experiment it is not possible to determine the order of events. Further examination of cell-wall gene expression levels at more regular intervals and comparison of radicle state may help elucidate the steps in this process further. A conclusion that may be drawn from the gene expression levels observed in this experiment is that some of the events in the preparation for radicle emergence are initiated within the first 24 hours post-imbibition.

Germinating seeds undergo a period of anoxic stress post-imbibiton and prior to the emergence of the radicle (Crawford, 1977). The density of material within the seed and the restrictions of the seed coat result in poor gaseous diffusion and consequently low oxygen concentrations within the seed (Bewley, 1997). A number of studies have suggested possible intersections between pathways of abiotic and biotic stresses (Pastori and Foyer, 2002; Mahalingham et al., 2003; Swindell, 2006). Investigation into gene expression responses to nine abiotic stresses showed that a number of genes previously identified as responsive to parasite and pathogen attack were up regulated in response to abiotic stress (Swindell, 2006). A number of genes involved in defence or pathogen response were amongst the most down regulated in day 1 compared to day 0 (table 5.1). It is possible that these genes were induced in response to the anoxic conditions within the seed immediately post-imbibition, illustrating the potentially large cross-talk between abiotic and biotic signalling pathways. The cross-talk and convergence of signalling pathways in response to distinct stimuli such as biotic versus abiotic stress serves to highlight the potential ambiguity in gene nomenclature and the necessity to describe and understand signalling networks.

The mobilisation of storage reserves in *Brassica napus* has been shown to peak around four days post-germination (Hutton and Stumpf, 1969). A number of genes involved in lipid metabolism, nutrient transport and sucrose response were amongst the most up-regulated genes in day 1 compared to day 0 (table 5.1). The observed increase in genes involved in storage mobilisation corresponds to an increase in amino acid and sugars levels immediately post-imbibition, demonstrating that in Arabidopsis seeds, the mobilisation of storage reserves occurs early on in germination.

5.4.3.2 Genes differentially expressed between day 1 and day 2

The genes with the largest increases and decreases in gene expression between days 1 and 2 were identified in order to investigate the contribution of these genes to the sample groupings. A re-organisation of genes related to protein synthesis was observed between these days. A number of genes involved in translation, transcription and ribosomal synthesis were repressed in day 2 compared to day 1 and a different set of genes encoding ribosomal proteins were induced in day 2 (table 5.2). Protein synthesis immediately post-imbibition is dependent on the presence of extant ribosomes in the dry seed but within the first 24 hours of imbibition the synthesis of new ribosomes is apparent (Dommes and Van der Walle, 1990). Beltrán-Peña et al. (1995) identified the presence of mature mRNA transcripts encoding the ribosomal proteins S_4 and S_6 stored in maize embryos and determined that other genes encoding ribosomal proteins were transcribed as development progressed. Examination of the gene transcripts in this experiment provides an indication of ribosomal protein synthesis in Arabidopsis seeds. Gene transcripts encoding 60S ribosomal proteins L_{26}, L_8, L_{22} and the 40S ribosomal protein S_{3A} are either present in the imbibed seed, or are transcribed immediately post-germination while gene transcripts encoding 60S ribosomal proteins L_{17}, L_{19}, L_6, P_{17} and the 40S ribosomal proteins S_{26}, S_{11} are transcribed later in the germination process.

Tissues of the quiescent seed are known to contain mitochondria, the development of which may be classified according to the main storage compound of the seed. Seeds with starch as the main storage compound repair existing mitochondria, which are damaged on desiccation and re-hydration, while oil-storing seeds produce new mitochondria (Morohashi, 1986). A number of genes associated with photosynthetic activity were repressed in day 2 compared to day 1 while a separate set of genes associated with the mitochondria were induced (table 5.2). It is unclear why photosynthetic genes are amongst the most down-regulated genes in day 2 compared to day 1. One possible explanation is that these genes are involved in the production of the mitochondria found in tissues of the quiescent seed. Arabidopsis, as an oil-seed, undergoes a period of *de novo* mitochondrial synthesis to replace the damaged extant mitochondria. It

is possible that the set of mitochondrial-associated genes induced in day 2 might have a role in this process.

The thirty most down-regulated genes in day 2 compared to day 1 included transcripts encoding proteins typically associated with the later stages of embryonic development and seed dormancy. The persistence of these signals in post-embrogenesis developmental stages suggests that the assigned roles for these genes might be expanded to include a role in germination or seedling establishment. It is known that mature seeds may become imbibed and then enter a phase of secondary dormancy for extended periods of time until germination conditions become favourable once more (Bewley and Black, 1994). This suggests a potential role for the presence of gene transcripts associated with embryonic development identified in days 1 and 2 (table 5.1 and 5.2) and at principal growth stage 1.0 (chapter 3) in the regulation of dormancy. It is possible that the presence of these gene transcripts may allow the seed to return to dormancy if required. The decrease in transcript levels over time corresponds to the development of the seed to the point where radicle emergence has occurred and the seedling is committed to a programme of photosynthetic establishment.

5.4.3.3 Genes differentially expressed between day 2 and day 3

The grouping of the transcriptome profiles showed that a cluster of samples was formed by profiles from days 3 to 8, excluding samples from days 0, 1 and 2. This suggests that a series of changes in gene expression levels occurs until day 3, when levels become more stable and similar to each other. In order to identify the genes primarily responsible for the observed grouping, the top thirty differentially expressed genes were identified between days 2 and 3 (table 5.3). A highly notable feature of the up-regulated genes is the increase in expression in a large number genes involved in photosynthesis; of the twenty six genes with an assigned identity in the TAIR database, nineteen showed a role in chlorophyll binding, photosystems I and II, mitochodria or chloroplasts (table 5.3). Developmental transitions from one stage to another, such as the transition from heterotrophic metabolism seen in seeds to photosynthetically competent seedlings, involves the down regulation of genes involved with the previous

phase and induction of genes associated with the next developmental phase (Kermode, 1990). The induction of a large number of photosynthetic genes between days 2 and 3 indicates the possibility that a developmental transition is occurring around this time.

Genes with a previously described role in development were identified amongst the genes with the largest decreases in gene expression in day 3 compared to day 2. The role of microtubules in the co-ordination of cell division and expansion has been described (Wasteneys, 2002). The *TUBG1* and *TUBG2* γ-tubulin isoforms have been identified as essential in microtubule nucleation (Liu et al., 1994). Mutations in either *TUBG1* or *TUBG2* do not produce a detectable phenotypic change however, although a combination of mutations in both genes results in normal development throughout seed formation and early seedling development, severe developmental defects are observed thereafter ultimately leading to seedling death after three weeks (Pastuglia et al., 2006). The identification of *TUBG1* and *TUBG2* in the gene expression profiles indicates their importance in microtubule organisation during development. It is unclear why the expression of these genes decreases dramatically between days 2 and 3 (97-fold and 57-fold decrease, respectively). The results suggest that the genes may play a more important role in the events leading up to the emergence of cotyledons than in seedling establishment. This provides an interesting focus for further investigation into the role of these genes in development.

A gene (At5g09660) involved in the mobilisation of nutrient resources was up regulated in day 3 compared to day 2 (table 5.3). At5g09660 has recently been identified as a peroxisomal NAD+-malate dehydrogenase PMDH1 required for β -oxidation of fatty acids in post-germinative growth of Arabidopsis seedlings (Pracharoenwattana et al., 2007). The induction of PMDH1 between one and two days after germination shows similarities with other genes involved in β-oxidation (Rylott et al., 2001; Fulda et al., 2002).

5.4.4 Model of development based on gene expression and metabolite profiles

The information obtained from metabolite and transcriptome profiles of Col-0 was used to develop an overview of gene expression and metabolite changes during seed development and germination (figure 5.8). The aim of this experiment was to investigate profiles produced throughout germination and seedling development. Extensive research has been carried out on the developmental period encompassing seed development, maturation, germination and seedling establishment. A descriptive model was produced to provide an overview of the results identified in this experiment in the context of profiling experiments produced at other developmental stages. A number of metabolic and gene expression changes may be observed. Storage compounds are synthesised and accumulated during seed development and maturation, while the expression of a number of genes associated with these processes increases. Conversely, the expression of genes involved in metabolism of reserves decreases during the maturation phase. At the onset of seed germination, a number of storage compounds are released and metabolised, which is associated with the expression of genes involved in these processes. As storage reserves become depleted over time, seedlings become established with the synthesis of photosynthetic apparatus and the dismantling of machinery associated with reserve mobilisation. Phytohormones play an important role during these developmental phases and the major events in levels are indicated briefly.

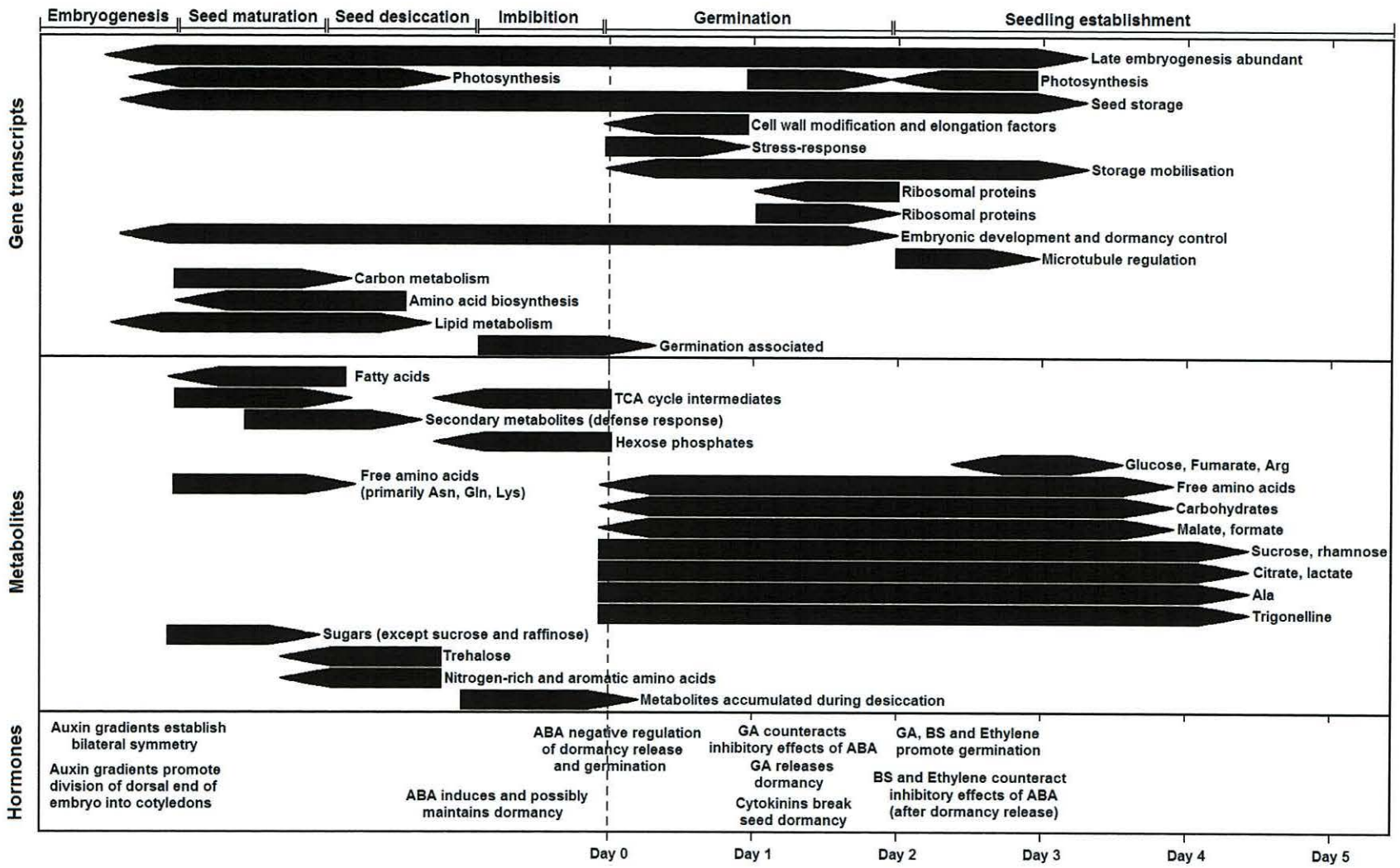


Figure 5.8: An overview of metabolite, gene expression and phytohormone events and roles during seed development and germination derived from a number of published reports.

Chapter 6

Integration of transcript and metabolite data

6.1 Introduction

Seed germination is regulated, in part, by a balance of the plant hormones gibberellin and abscisic acid which directs the gene expression events required for the release of seed dormancy and the emergence of the radicle (Bewley and Black, 1994). The existence of Arabidopsis mutants, such as the Cape Verde Island mutant, which exhibit a seed dormancy not relieved by the application of exogenous giberellin indicates that other factors are required for germination to proceed (Ali-Rachedi et al., 2004; Penfield et al., 2006). The accumulation of storage material during seed maturation is known to affect the efficiency of reserve mobilisation and subsequent seedling establishment and is influenced by carbon and nitrogen partitioning (Eastmond and Rawsthorne, 2000; Eastmond and Graham, 2001; Brocard-Gifford et al., 2003). Arabidopsis seed germination is delayed but not blocked by a significant reduction in stored lipid content while a number of mutants deficient in the mobilisation of stored oil reserves are able to germinate successfully (Katavic et al., 1995; Focks and Benning, 1998; Lu and Hills, 2002; Penfield et al., 2005). These findings suggest that germination and the utilisation of stored lipids are independently regulated and that germination is regulated by other factors in addition to the accumulation and mobilisation of storage reserves (Pritchard et al., 2002; Fait et al., 2006).

Nutrient-sensing programmes allow cells to become aware of and respond appropriately to environmental and developmental perturbations. Nutrient-sensing responses have been described for carbohydrate sensing in bacteria and yeast, while the division of plant meristem cells has been shown to require a suitable nutritional state involving carbohydrate sensing (Postma et al., 1993; Rolland et al., 2002; Francis and Halford, 2006). Nitrate has also been implicated as a signalling compound in Arabidopsis germination. The

application of nitrate has been shown to stimulate the germination of dormant seeds and the accumulation of nitrate in the mother plants results in seeds with lower dormancy levels (Alboresi et al., 2005). The application of nitrate has been shown to alter the phytochrome-induced germination response in Arabidopsis seeds (Batak et al., 2005).

The production of metabolite profiles during seed development and germination showed that a series of distinct metabolic switches occur between reserve accumulation, seed desiccation, vernilisation and germination (Fait et al., 2006). It is possible that the metabolic state of the seedling might be one of the unknown factors involved in the regulation of seed dormancy and germination processes. Metabolite profiles produced at distinct developmental stages from imbibition to seedling establishment were examined in combination with corresponding transcriptome profiles to provide an overview of potential regulatory interactions between genes and metabolites during this developmental process.

6.1.1 Network cartography and correlation analysis

The PCA plots produced in the previous chapter to describe the relationships between samples based on gene expression and metabolite profiles gave an indication of trends in the data, but are open to user-specific interpretation. An algorithm based on spring embedding was used to visualise relationships between time points in the data set based on transcriptome and metabolite profiles in a clearer manner. The algorithm describes interactions between nodes of the network (e.g. time points) which are modelled as springs whose strength corresponds to the strength of the relationship, with stronger springs indicating a stronger relationship. Starting from a random initial configuration, the system is allowed to relax using Newtonian dynamics to reach a low energy equilibrium state with more strongly related nodes lying closer together. The network representation allows clusters and chains of highly related nodes to be identified, aiding the elucidation of higher order relationships. The algorithm can be used in the analysis of time series data by grouping tissue samples based on the similarity of either gene expression or metabolite data (Ebbels et al., 2006).

the analysis of time series data by grouping tissue samples based on the similarity of either gene expression or metabolite data (Ebbels et al., 2006).

In order to minimise artefacts due to the flooring of transcript intensity levels, all values of 20 and below were removed from the analysis. Only those genes which had at least one expression value for each of the sample days were included in the analysis. This rigorous filtering process produced a final set of 10,005 genes used for the Spring Embedding and correlation analysis. Such filtering was not required for metabolite levels as they had been quantified absolutely. Since material for both the transcript and metabolite profiling was collected in three independent groups, the values were averaged where possible for each sample day. Linear correlation coefficients were determined for corresponding transcript or metabolite levels between each of the days. Pearson correlations were determined between genes as a set and metabolites as a set and the time points clustered based on the resulting correlation coefficients using spring embedding (Ebbels et al., 2006). When samples were clustered based on gene expression correlation coefficients >0.7, two clusters were observed, one containing profiles from days 0 to 2 and the other days 3 to 8 (figure 6.1 A). When the time points were clustered based on metabolite coefficients >0.7, two clusters were again observed but day 2 samples clustered with those from days 3 to 8 (figure 6.1 B).

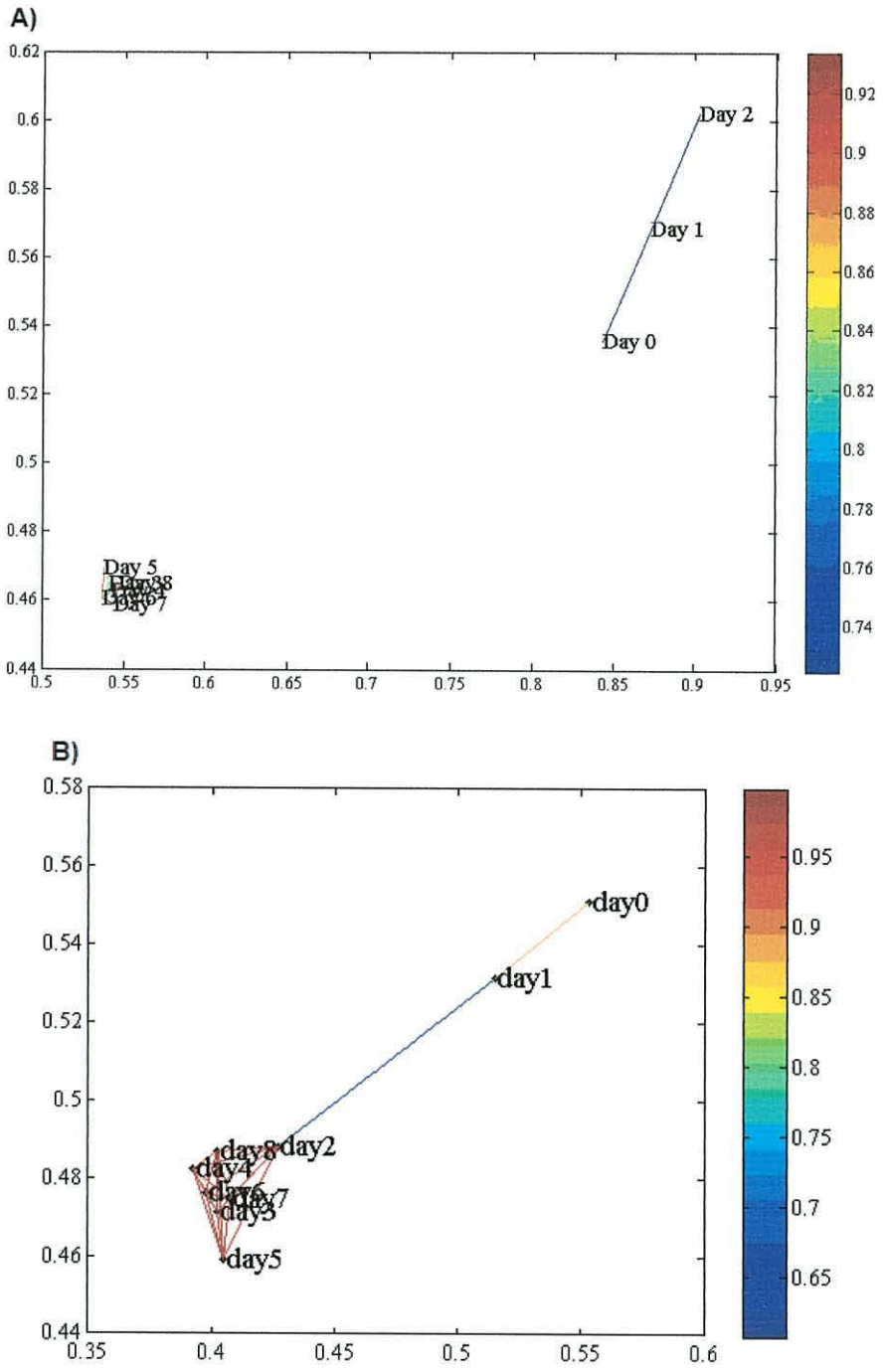


Figure 6.1: Spring Embedding model to describe the grouping of samples based on (a) gene expression profiles and (b) metabolite profiles. Time points are represented as nodes connected to related nodes by springs. The length and colour of the connecting line is representative of the relationship. Closely related nodes are joined by shorter lines and a dark red colour indicates a strong positive correlation while a dark blue line is representative of a weaker positive correlation. Samples based on gene expression correlation coefficients form two groups with samples from days 0, 1 and 2 grouping together and samples from days 3 to 8 grouping together. Samples grouped based on metabolite correlation coefficients also form two clusters but day 2 is grouped with days 3 to 8.

6.2 Correlations between transcript and metabolite levels

A variation in both metabolite and gene transcript levels is necessary to identify correlated behavioural patterns. The majority of the metabolites measured demonstrated altered levels throughout the time course allowing correlations to be identified with gene transcript levels. For metabolites that demonstrated altered levels throughout the time course either a linear or a biphasic profile \9i.e. one which increased, then decreased in levels or *visa versa*) was observed. A number of metabolite levels were seen to fluctuate over time, but demonstrated no overall change in behaviour; no correlations were identified between gene transcripts and metabolites that showed no overall change in levels. A spring embedding algorithm was used to visualise relationships between genes and metabolites based on the correlation of levels over time (Ebbels et al., 2006). Connected gene transcript and metabolite levels are identified by significantly non-zero correlation coefficients. A Bonferroni adjustment was made to the data set prior to analysis. Bonferroni adjustments are used based on the reasoning that a type I or α error will occur in data analysis (i.e. for an experiment where the null hypothesis holds true a significant result, $p<0.05$, will arise by chance 5% of the time). The chance of false positive occurrence increases in the case of multiple tests of statistical significance against the same data set, as in the identification of correlated gene transcript and metabolite levels in this experiment. For an experiment where n independent hypotheses are tested on a set of data, the Bonferroni adjustment sets a new level of statistical significance of the original α level divided by the number of outcome measures (α/n). It should be noted that the Bonferroni correction is a highly stringent correction and therefore the possibility of type II errors (false negative results) is also increased by the application of this correction (Perneger, 1998).

The statistical basis of this data integration was to calculate linear correlation coefficients between data values in or between data sets to produce similarity matrices that could be visualized using network cartography, where significance was determined by threshold cut-off values. In this first instance linear relationships have been forced, a valid assumption as long as a gene expression is directly responsive (i.e. changes linearly with a measurable degree of interaction

occurring within a 24 hour time period) to the level of a metabolite. Nevertheless, it is apparent that not all metabolites and transcript levels will change linearly, but may show multiphasic behavior, such as increase or decrease with developmental state. It may be possible to strengthen relationships if non-linearity is accounted for in determining correlations using non-parametric correlation statistics, for example. The statistical algorithms used to establish thresholds of significance for adjusting the similarity matrices are very stringent, and that we are more likely to lose significant relationships as opposed to including false positive ones. The identification of a correlated gene and metabolite does not provide any information relating to the causality in the relationship (i.e. metabolite affecting gene expression or visa versa). It is also difficult to determine whether the observed relationship results from a direct interaction between a gene and a metabolite, or whether a downstream signalling event is involved.

6.2.1 Identified gene-metabolite correlations

A total of 237 correlations (>0.7) from 20 metabolites and 210 genes were identified. As expected, both positive and negative correlations were identified. Table 6.1 lists the metabolites identified as showing a correlation with one or more genes along with the nature of the correlations. The metabolite profile is described as increasing, decreasing, or as biphasic throughout the developmental series.

Table 6.1: Table to show the number of genes correlated with each metabolite, and the nature of the correlations (positive or negative). The profile of the metabolite is described as increasing or decreasing or as biphasic throughout the developmental time course.

Metabolite	Metabolite profile	Related genes	Nature of Correlation
Glucose	Increases then decreases	1	Negative
Sucrose	Decreases	129	Positive
Fructose	Increases then decreases	1	Positive
Rhamnose	Increases then decreases	1	Negative
Malate	Increases then decreases	3	Negative (1), Positive (2)
Citrate	Decreases	7	Negative (6), Positive (1)
Fumarate	Increases	3	Negative (2), Positive (1)
Formate	Decreases then increases	1	Negative
Lactate	Increases then decreases	21	Negative
Glutamine	Increases then decreases	7	Negative
Alanine	Increases then decreases	4	Negative (1), Positive (3)
Valine	Increases then decreases	3	Negative
Threonine	Decreases	5	Negative
Trigonelline	Decreases	16	Negative (15), Positive (1)
Choline	Increases then decreases	1	Negative
unkD8.0	Decreases	6	Negative (2), Positive (4)
unkQ7.9	Increases then decreases	4	Negative
unkS7.36	Increases then decreases	3	Negative
unkQ5.18	Increases	2	Negative (1), Positive (1)
unkM1.85	Increases then decreases	19	Negative (18), Positive (1)

The 237 identified correlations between genes and metabolites listed in table 6.1 were visualised by spring embedding using metabolites as nodes and genes as edges (figure 6.2). The nature of the correlation is described by the colour of the connecting lines. Correlations represented by red lines indicate a strong positive correlation between a gene and metabolite and a dark blue line is indicative of a strong negative correlation.

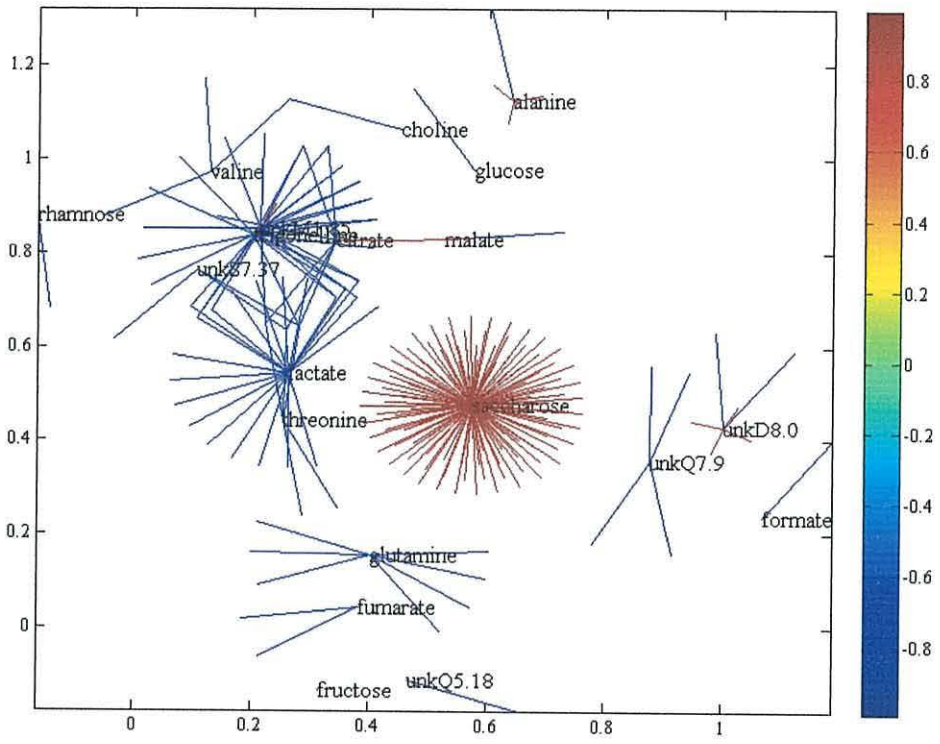


Figure 6.2: Spring embedding model to show relationships between genes and metabolites from days 0 to 8. Metabolites are represented as the central nodes from which correlated ($p<0.0001$) genes radiate outwards. The direction and strength of the association is represented by the colour of the line; a dark red colour indicates a strong positive correlation and a dark blue line is representative of a strong negative correlation. A total number of 237 correlations were identified between 20 genes and 209 metabolites.

Several representative examples of gene transcripts correlated with lactate levels are presented below (figure 6.3). The entire list of correlated gene transcripts and metabolites is presented in Appendix E.

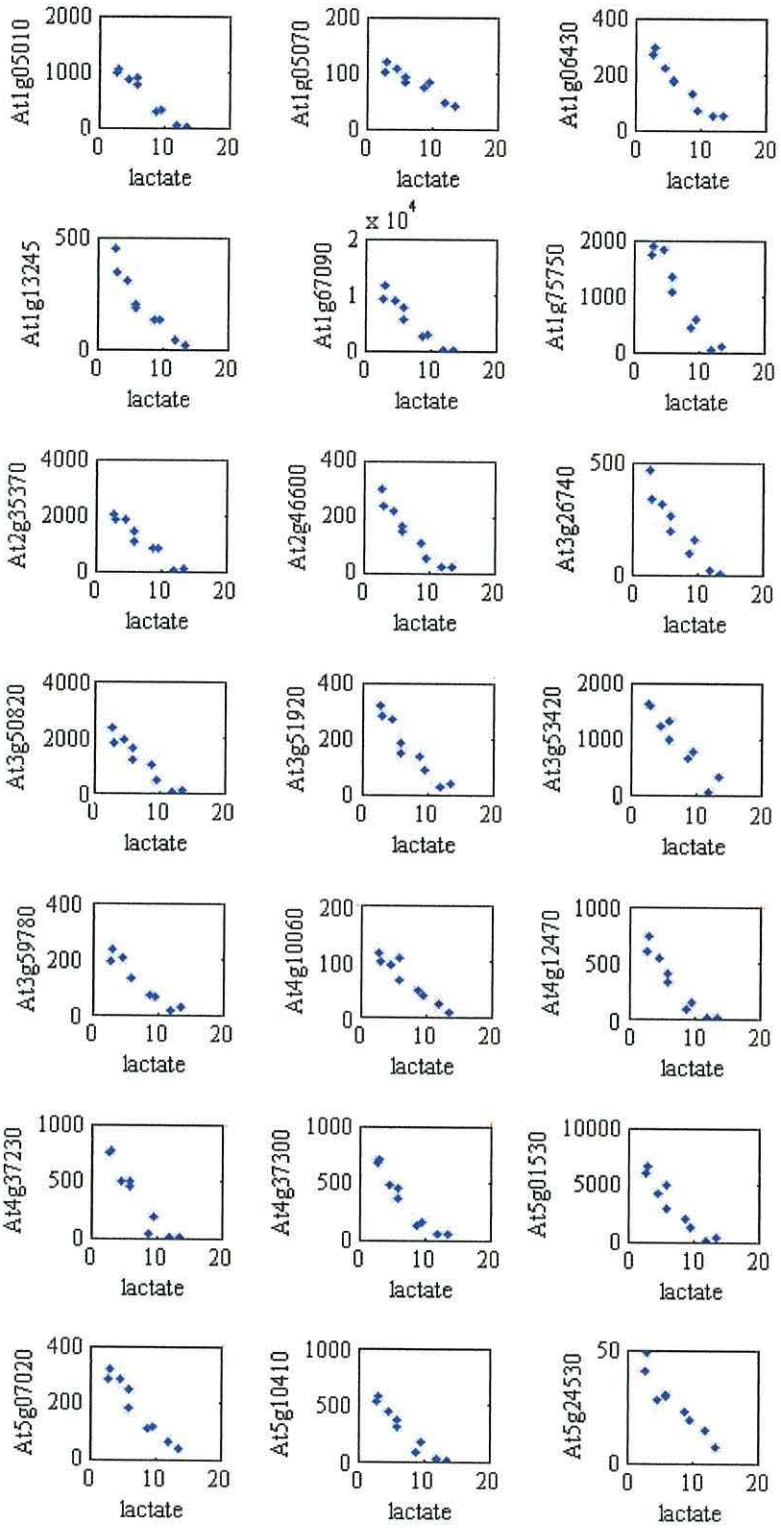


Figure 6.3: Scatter-plots to show correlations between 21 gene transcripts and lactate levels. Lactate levels are displayed along the bottom axes (ug/g fresh weight) and gene transcript levels are represented by the vertical axes (arbitrary hybridisation intensity scale).

6.2.2 Sucrose showed correlations with a large number of gene transcripts

Sucrose showed positive correlations with 129 gene transcripts indicating that transcript levels decrease with sucrose concentration throughout the time course. Of the 129 correlated gene transcripts, 44 had previously been identified as responsive to treatment with sucrose and are listed in table 6.2 (Zimmermann et al., 2004; Bläsing et al., 2005; Gonzali et al., 2006; Müller et al., 2007; chapter 4). Correlations that might be expected were also observed; sucrose levels showed a positive correlation with transcript levels of sucrose phosphate synthase, which catalyses the rate limiting step in sucrose biosynthesis (Huber and Huber, 1996).

Table 6.2: Gene transcripts correlated with sucrose levels which have previously been shown to be responsive to treatment with sucrose. The gene ID is provided with a brief description and the corresponding correlation coefficient (Corr. Coef.). The source reference (Ref.) is provided as follows: (1) Genevestigator, Zimmermann et al., 2004; (2) Bläsing et al., 2005; (3) Müller et al., 2007; (4) Gonzali et al., 2006; (5) chapter 4 (table continues on page 167).

Gene ID	Gene description	Corr. Coef.	Ref.
At1g10270	pentatricopeptide repeat-containing protein	0.97	4,5
At1g15440	transducin family protein / WD-40 repeat family protein	0.97	4
At1g18850	expressed protein	0.95	4
At1g26740	expressed protein, similar to 50S ribosomal protein L32 (*T. thermophilus*)	0.96	5
At1g27150	expressed protein	0.96	5
At1g29800	similar to zinc finger family protein	0.95	4
At1g51380	eukaryotic translation initiation factor 4A, putative	0.99	2, 4
At1g52930	brix domain-containing protein	0.99	4
At1g56110	nucleolar protein Nop56, putative snoRNA binding domain	0.98	4, 5
At1g69530	expansin, putative (EXP1)	0.97	5
At1g70350	expressed protein	0.95	5
At1g75200	flavodoxin family protein / radical SAM domain-containing protein	0.96	5
At1g77120	alcohol dehydrogenase	0.99	3, 5
At1g77510	protein disulfide isomerase-like protein, a member of the thioredoxin (TRX) superfamily.	0.96	2, 3, 5
At2g19640	SET domain-containing protein	0.96	5
At2g24500	zinc finger (C2H2 type) family protein	0.96	4
At2g37020	translin family protein	0.95	5

At3g08900	reversibly glycosylated polypeptide-3, contains non-consensus GA-donor splice site at intron 2	0.96	5
At3g12860	nucleolar protein Nop56, putative, putative snoRNA binding domain	0.97	4
At3g28345	ABC transporter family protein	0.97	5
At3g44750	histone deacetylase, putative, contains Zinc finger, C2H2 type domain	0.98	2, 4, 5
At3g49240	pentatricopeptide repeat-containing protein, contains	0.95	4
At3g52040	expressed protein	0.95	4
At3g53940	mitochondrial substrate carrier family protein	0.97	4
At3g57150	dyskerin, putative / nucleolar protein NAP57, putative	0.97	4
At4g04940	transducin family protein / WD-40 repeat family protein, contains seven G-protein beta WD-40 repeats	0.96	3
At4g16160	mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	0.96	1
At4g16515	expressed protein	0.95	1, 5
At4g17550	transporter-related; contains Pfam profile PF00083: major facilitator superfamily protein	0.96	5
At4g18910	aquaglyceroporin / NOD26-like major intrinsic protein 2	0.96	5
At4g24800	MA3 domain-containing protein	0.95	5
At5g02050	mitochondrial glycoprotein family protein / MAM33 family protein	0.96	2, 3, 4, 5
At5g04040	patatin-related, contains Patatin domain PF01734	0.98	4
At5g14180	lipase family protein, ab-hydrolase associated lipase region	0.95	2, 4
At5g17380	pyruvate decarboxylase family protein	0.97	3
At5g19550	aspartate aminotransferase, cytoplasmic isozyme 1 / transaminase A (ASP2)	0.97	5
At5g20280	sucrose-phosphate synthase, putative	0.95	1
At5g39850	40S ribosomal protein S9	0.97	5
At5g42150	expressed protein	0.97	4
At5g54300	expressed protein	0.95	5
At5g56030	identical to heat shock protein 81-2	0.96	3
At5g57655	xylose isomerase family protein	0.97	3, 4, 5
At5g61020	YT521-B-like family protein	0.95	5
At5g67360	cucumisin-like serine protease (ARA12)	0.95	4

Analysis of biological function ontology showed that the gene transcripts correlated with sucrose levels were involved in a variety of functions (figure 6.4). Of the 129 gene transcripts, only 29 had been assigned to an ontology class. Examination of the associated biological functions showed that the genes were involved in a variety of processes, primarily metabolic, cellular, localisation and regulatory processes.

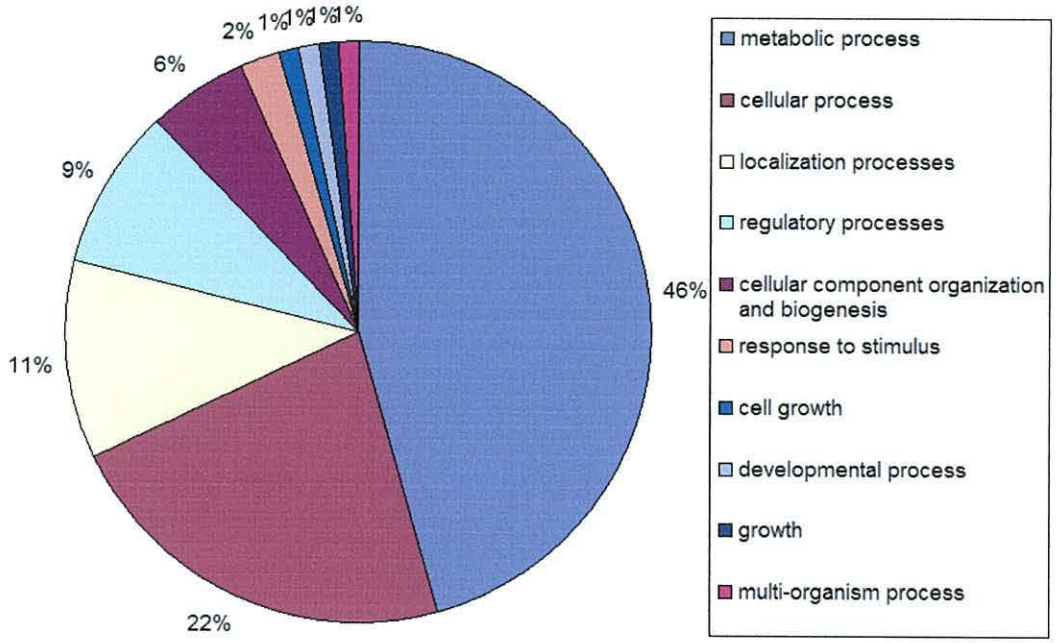


Figure 6.4: Gene ontology characterisation (biological function) of the gene transcripts correlated with sucrose levels.

6.2.3 Regulatory genes correlated with metabolites

Of the 210 gene transcripts correlated to metabolite levels, nearly 50 were identified with an involvement in transcriptional, translational or signalling processes. Three of these, (table 6.3) the phytochrome interacting factors PIF1 and PIF4 and abscisic acid insensitive 3 (ABI3) have been shown to have a regulatory role in seed germination (Oh et al., 2004, 2007; Koornneef et al., 1984). A further 30 of the 210 gene transcripts correlated to metabolite levels were identified in the TAIR database as responsive to a known developmental regulator such as gibberellin or auxin and/or with a role in germination. The full list is displayed in Appendix E.

Table 6.3: Transcription factors with an identified role in the regulation of germination showed correlations with metabolites. The gene ID and description is detailed with the corresponding metabolite and correlation coefficient.

Gene ID	Gene description	Correlation Coefficient
At2g20180	Basic helix-loop-helix (bHLH) family protein, contains Pfam domain, PF00010: Helix-loop-helix DNA-binding domain (PIF-1)	0.9493 Sucrose
At2g43010	phytochrome-interacting factor 4 (PIF4) / basic helix-loop-helix protein 9 (bHLH9) / short under red-light 2 (SRL2), identical to SP:Q8W2F3 Phytochrome-interacting factor 4 (Basic helix-loop-helix protein 9) (bHLH9) (Short under red-light 2) {Arabidopsis thaliana}; supporting cDNA gi:18026965:gb:AF251694.1:AF251694	-0.9529 unkM1.85
At3g24650	abscisic acid-insensitive protein 3 (ABI3), identical to abscisic acid-insensitive protein 3 GI:16146 SP:Q01593 from (Arabidopsis thaliana)	0.9721 Sucrose

6.2.4 Gene transcripts of similar related ontology correlate with metabolite levels

Several metabolites showed correlations with gene transcripts of a similar ontology. Lactate was negatively correlated with 21 gene transcripts, 7 of which were involved in photosynthesis-related functions. The unidentified peak M1.85 showed negative correlations with 18 gene transcripts, 7 of which were also identified with roles in photosynthesis. In order to identify potential regulatory mechanisms for these two groups of genes, promoter sequences 1.1 kb upstream, of the transcription initiation site were obtained for each of the sets of photosynthetic genes associated with lactate and M1.85 levels and analysed for over-represented promoter motifs as described previously (chapter 2, section 2). A summary of the potential DNA motifs identified in promoter sequences of co-expressed genes correlated with lactate and the unknown peak M1.85 is presented in tables 6.4 and 6.5, respectively. Motifs from the online CARE databases PLACE and PlantCARE are listed with a brief description and regions of sequence homology highlighted in **bold** or ***bold italics*** for sequences found in the reverse complement orientation. Unless otherwise stated, homologous

promoter motifs were identified in *Arabidopsis thaliana*. Where nucleotide base ambiguity occurs the IUPAC nomenclature system has been followed (Appendix A).

Table 6.4 The consensus sequences of potential DNA motifs found in the promoters of photosynthetic-related genes correlated with lactate. The corresponding motifs found in PlantCARE ([a]) and PLACE ([b]) are listed and a brief description of the motif is given.

Motif	Motif Submitted	Sequence Found	Descriptor
1	**TTGRATTG**	No match in database	-
2	**GCCACNNG**	***ACGTGGC***[a]	ACGT-containing ABA response element
		GCCACT[a]	CAT-box – meristem specific expression
		TC***CTCGTGGC***A[a]	G-box – light responsive element
		CAT***CGTGTGGC***[a]	GTGGC motif – light responsive element
		GCCACN[b]	Light responsive element
3	**CAGACA**	No match in database	-
4	**TCAAT**	**CAAT**[b]	Found in legumin promoter (*P. sativum*)

Four potential DNA sequence motifs were identified in the promoters of the 7 photosynthetic gene transcripts correlated with lactate levels. The potential GCCACN motif demonstrated sequence homology with a number of motifs in the PLACE and PlantCARE databases. Four DNA motifs associated with light responsiveness were identified. Motifs associated with meristem specific activity and ABA stimulus were also identified. Two of the potential DNA motifs identified in the promoter sequences did not show sequence homology to any of the CAREs in either the PlantCARE or the PLACE database.

Table 6.5: Consensus sequences of motifs found in the promoters of photosynthetic-related genes correlated with the unknown peak M1.85. The corresponding motifs found in PlantCARE ([a]) and PLACE ([b]) are listed with a brief description of the motif.

Motif	Motif Submitted	Sequence Found	Descriptor
1	**TNCTAG**	TTCTTCAAGCTTCA AGACAA**TCCTAGA** AATTAC[a]	HSE element – heat shock and ABA response
2	**GGTTNGA TTC**	No match in database	-
3	**TCAA**	**TCAA**CCAACTCC[a]	AC-II – regulation of phloem and xylem expression
		TTGACC[a]	W-box – wounding and pathogen response
		ATCTTATGTCA***TTG A***TGACGACCTCC[a]	Ocs – auxin, salicylic acid and oxidative stress response

Three potential promoter motifs were identified in the photosynthetic genes correlated to M1.85 levels, two of which showed homology to sequences in the PLACE and PlantCARE databases. The potential TNCTAG motif showed homology to sequences responsive to abscicic acid and elevated temperature. The TCAA motif sequence demonstrated homology to database DNA elements involved in stress responses and an element involved in the regulation of phloem and xylem expression.

6.3 Discussion

6.3.1 A reorganisation in metabolic content precedes gene expression changes

A series of distinct metabolic switches have previously been identified in Arabidopsis seed maturation and germination (Fait et al., 2006). The data presented here extends this analysis to provide an overview of metabolite and transcriptome profiles from imbibiton through to established, photosynthetically competent seedlings. In order to identify shifts in gene expression and metabolite levels, Pearson correlations were determined for gene transcript levels as a set and for metabolite levels as a set between sequential time points. When time points were clustered based on gene transcript levels using spring embedding (for correlation coefficients >0.7) two groupings were formed. Gene expression profiles from days 0 to 2 formed one group while a second group was formed by profiles from days 3 to 8 (figure 6.1 A). Interestingly, when samples were clustered based on metabolite correlation coefficients >0.7, two groups were identified. Samples from days 0 and 1 grouped together but samples from days 2 to 8 formed a second group (figure 6.1 B). The observed clustering of time points according to metabolite and gene transcript levels indicates that a metabolic switch has the potential to induce changes in gene expression, at least for the metabolites measured in this work.

Fait et al. (2006) observed a burst of metabolic activity during post-imbibition germination (corresponding to day 1 in this experiment). Grouping of the metabolite profiles produced during this experiment supports these findings, demonstrating that a relatively stable metabolic state is induced by day 0 and remains constant until day 1 (figure 6.1 B). In addition to the metabolic switches described by Fait et al. (2006), a further switch occurring between days 1 and 2 is indicated by this work. The main reorganisation in gene expression levels occurs between days 2 and 3 with the up-regulation of a large number of photosynthetic-related genes. The temporally distinct switches in gene and metabolite levels imply that a rearrangement in the metabolic state of the seedling might induce the establishment of photosynthetic machinery.

6.3.2 Sucrose levels correlate with a large number of genes

The spring embedding approach models interactions between nodes of a network (e.g. genes) whose strength corresponds to the strength of the relationship (Ebbels et al. 2006). This approach was applied to the visualisation of interactions of genes and metabolites uniquely and metabolites with genes. The correlation of 210 gene transcript and 20 metabolite levels throughout germination indicates the potential for regulatory relationships to exist between the identified pairs of genes and metabolites. Sucrose levels showed positive correlations with 129 gene transcripts, indicating that these gene transcripts decreased throughout germination. This is an expected result as sucrose has been well characterised as a signalling metabolite able to induce gene expression changes in a wide range of genes (Koch, 1996; Gibson, 2005). Furthermore, sucrose demonstrated a positive correlation with the transcript levels of sucrose phosphate synthase, which has been identified as catalysing the rate limiting step in sucrose biosynthesis (Huber and Huber, 1996). A number of transcriptome profiling experiments utilising microarrays have identified genes with an altered transcriptional response to the presence of sucrose. Further comparison of the gene transcripts correlated with sucrose levels with previous microarray experiments and online databases showed that 44 of the 129 genes had previously been identified as sucrose-responsive (Zimmermann et al., 2004; Bläsing et al., 2005; Gonzali et al., 2006; Müller et al., 2007; chapter 4). The correlation of sucrose levels with a large proportion of previously identified sucrose-responsive gene transcripts reinforces the validity of the use of correlation coefficients to identify relationships between gene transcript and metabolite levels. Although the possibility exists that a number of the identified correlations may have arisen spuriously, these results indicate that the majority of the gene transcripts correlated with sucrose levels is also likely to be sucrose-responsive and that relationships identified between gene transcripts and other metabolites are reliable.

The characterisation of the biological function ontology of the gene transcripts correlated with sucrose shows that the genes are primarily involved in metabolism and regulatory processes (figure 6.4). Sugar signalling and the associated metabolic and gene expression responses is a complex process with

many components yet to be identified (reviewed in Leon and Sheen 2003; Rook and Bevan 2003; Gibson 2005). Although correlated behaviour is not evidence that a regulatory relationship exists between a gene and a metabolite, it is possible that a number of these interactions might be involved in sugar sensing or responses. In order to fully understand sugar signalling mechanisms and the elicited responses, it is necessary to describe as fully as possible the components of the signalling pathways. The correlation of genes with an established role in regulatory processes such as the transcription factor zinc-finger family protein (At2g24500) or the brix-domain containing protein (At1g52930) which has been associated with ribosome biogenesis (Eisenhaber et al., 2001) indicates that novel components of sugar signalling pathways might be identified in this manner.

6.3.3 Genes associated with development correlate with metabolites

A number of gene transcripts correlated with metabolite levels are involved in signalling, transcriptional or translational processes. Of these, three transcription factors were of interest (table 6.3). The *PIF* and the *ABI3* genes have previously been identified as regulators of germination and are also known to interact with plant hormones and other regulatory factors. This demonstrates the potential network of interactions which occur between genes, metabolites, phytohormones and physical factors to regulate germination and seedling establishment.

PIF1 has been identified as a negative regulator of photomorphogenesis in seedlings (Huq et al., 2004; Oh et al., 2004). The transcriptional activity of *PIF1* has been shown to be light-responsive; in the dark *PIF1* is transcriptionally active and phytochrome-mediated light response reduces this activity, allowing photomorphogenesis to proceed (Huq et al., 2004). The observed correlation between *PIF1* and sucrose levels might be explained by cross-talk between sugar and light signalling pathways, again demonstrating how various factors might interact to regulate germination (reviewed in Smeekens et al., 2000; Rolland et al. 2002).

A phytochrome-mediated red-light signalling cascade has been shown to positively regulate seed germination through the degradation of *PIF4* and subsequent de-repression of giberellic acid biosynthesis (Oh et al., 2006). The

correlation with the unidentified peak M1.85 suggests that *PIF4*-mediated regulation of germination is a complex system which integrates the responses of light, hormones, gene expression and metabolites to control development.

The *ABI3* gene encodes the Arabidopsis homologue of the maize *VIVIPAROUS-1* (*VP1*) gene; the *abi3* mutant is able to germinate in the presence of exogenous abscisic acid concentrations that inhibit germination in the wild type seeds (McCarty et al., 1991; Koornneef et al., 1984). A combination of *VP1* and abscisic acid was shown to influence the expression of a large number of genes related to metabolism and transcriptional regulation in Arabidopsis in an *abi3* null mutant background (Suzuki et al., 2003). This suggests that interactions between *VP1* (or *ABI3*) and abscisic acid have the potential to induce metabolic and transcriptomic shifts in plant systems and that this is a situation which might occur in Arabidopsis seedlings during germination.

The gene transcript levels of *ABI3* showed a positive correlation with the levels of sucrose (table 6.3). A number of abscisic acid response and biosynthesis mutants have been identified with an altered response to sugars during Arabidopsis germination (Laby et al., 2000; Finklestein et al., 2002; Zeng and Kermode, 2004; Dekkers, 2006). Interactions between sucrose and *ABI3* are indicated by the sucrose insensitive phenotype displayed by the *abi3* mutants (Dekkers, 2006). This supports the correlated profiles observed in this experiment. A proposed signalling pathway model suggests that seed development and abscisic acid response are controlled by the interaction of *ABI3* with the abscisic acid response genes *ABI4* and *ABI5* and it is possible that the expression of the *ABI* genes is regulated by a further, unidentified, factor (Soderman et al 2000). Based on the assumption that a regulatory relationship between a gene transcript and the unknown factor will result in a rapid, linear and quantifiable response it is possible that sucrose, or a related signalling component, influences the expression of the abscisic acid response genes proposed to act in conjunction with each other.

The correlation of *ABI3* transcript levels with those of sucrose post-germination suggests that *ABI3* might be involved in the establishment of seedlings. These findings are supported by experiments which have shown that *ABI3* is essential for abscisic acid-induced post-germination early seedling developmental arrest and that *ABI3* may play a role in sugar-induced seedling

developmental arrest (Lopez-Molina et al., 2002). It is possible that interactions between sucrose and *ABI3* facilitate, in part, the successful establishment of Arabidopsis seedlings post-germination.

6.3.4 Genes of a similar ontology correlate with metabolites

A number of metabolites were observed to correlate with genes of a similar ontology. The levels of both lactate and the unknown peak M1.85 correlated with gene transcripts related to photosynthesis. In order to identify and compare potential regulatory mechanisms for these two groups of genes promoter sequences were analysed for the presence of statistically over-represented DNA motifs with the potential to act as CAREs. The motifs identified in the promoters of the photosynthetic genes correlated with lactate predominantly showed sequence homology to potential promoter elements identified as light-responsive. It has been shown that elevated concentrations of lactate found in certain types of tumour tissues induce alterations in gene expression levels (Roth and Droge, 1991; Walenta and Mueller-Klieser, 2004; Schmid et al., 2007). The carbohydrate-response elements (ChoRE) found in several glycolysis-related genes have been identified as potential components of signalling pathways linking lactate to the proteome of tumour cells (Walenta et al., 2004). A large degree of cross-talk exists between light sensing and carbohydrate signalling pathways; consequently it is difficult to distinguish between the effects of one stimulus in relation to the other (reviewed in Smeekens et al., 2000; Rolland et al. 2002). It is possible that the promoter elements identified as light responsive might also be responsive to carbohydrates and that a lactate signalling pathway similar to the proposed system in tumour cells might exist in plant systems.

Interestingly, the potential promoter motifs identified in the photosynthetic genes correlated with the levels of the unidentified peak at M1.85 predominantly included elements associated with response to various stresses and did not contain any light-responsive elements typically associated with the promoters of photosynthetic genes. The absence of potential promoter motifs with homology to light responsive elements in the promoters of the set of gene transcripts correlated with the unknown metabolite M1.85 suggests that a regulatory

mechanism other than light influences expression of these genes. It is possible that the compound found at M1.85 has an effect on the transcription of these genes. The difference between the potential promoter motifs identified in the two sets of photosynthetic genes indicates that two distinct regulatory mechanisms may operate at the level of the promoter to regulate the expression of these genes.

Chapter 7

Future Work

7.1 Consequences of disrupting acetate metabolism

7.1.1 Metabolic and transcriptomic consequences of disrupting acetate metabolism

Labelling studies using 2[^{13}C]acetate were used to trace the flow of organic acids from peroxisomes to the mitochondria. The resultant partitioning of label was used to describe a model for the utilisation of exogenously supplied 2[^{13}C]acetate. The proposed model may be incorporated into the findings of Pracharoenwattana et al. (2007) for the function of peroxisomal malate dehydrogenase as a means to recycle NADH produced from catabolism of stored fatty acids. Cytosolic malate dehydrogenase may act to produce oxaloacetate which re-enters the glyoxysome and subsequently be partitioned between malate and citrate synthesis to produce the labelling observed in the side chain carbons of glutamine. An interesting direction for future investigation is the use of further isotopic labelling studies to trace the pathway of carbon through glxoyxlate cycle intermediates to determine the metabolic fate of acetate in *acn1-2*. For example, feeding studies utilising labelled malate would reveal further information regarding the export of malate from the peroxisome and the potential for randomisation of carbons through conversion to fumarate.

7.1.2 Interactions between organic acid and carbohydrate signalling

Interactions between carbohydrate and a number of other signalling pathways including nitrogen (Coruzzi and Bush, 2001; Corruzzi and Zhou, 2001; Price et al, 2004), plant hormones (Zhou et al., 1998; Yuan and Wysocka-Diller, 2006) and light (Thum et al., 2004) amongst others (Smeekens, 2000, Rolland et al., 2002, Halford and Paul, 2003) are extremely well characterised. However, very little work has been performed into identifying the potential for cross-talk between carbohydrate and organic acid signalling. Transcriptome profiles were produced using the *acn1* mutant to investigate interactions between acetate and sucrose signalling. The expression of a number of sucrose responsive genes was altered in *acn1*. Analysis of the sequences of promoters of genes with an altered response identified a number of DNA motifs previously identified with an involvement in sugar, stress and developmental signalling pathways. An interesting direction for future work is the analysis of promoter elements associated with developmental regulators such as light and phytohormones through techniques such as promoter truncation in order to determine whether these elements are also responsive to acetate signalling.

7.2 Identification of metabolic and genetic switches

Metabolite and gene expression profiles were produced for discrete developmental time points from imbibed seeds to eight day old seedlings. Spring embedding was used to group samples based on metabolite and gene expression profiles. It was observed that a rearrangement in metabolites occurred between days 1 and 2, prior to the rearrangement in gene expression profiles which was observed between days 2 and 3. An interesting direction for future work would be to produce transcriptome and metabolite profiles at more frequent intervals during the early stages of development up to day three. Sampling tissue and the subsequent production of profiles every four hours would establish more accurately at what point metabolite clusters diverge from gene expression

clusters. This would provide a clearer overview of metabolic and genetic switches during development.

The metabolite profiles in this work were produced using ^{1}H-NMR. Due to the limited sensitivity of NMR, only the levels of the most abundant metabolites were measured (Hall et al., 2002; Sumner et al., 2003; Oksman-Caldentey and Saito, 2005). An interesting extension of this work is to investigate the identified groupings of samples when the number of metabolites measured is increased. The use of a mass spectrometry based technique such as electrospray ionisation mass spectrometry to profile a larger number of metabolites would confirm whether the observed reorganisation of metabolites still precedes genetic switch on a more global scale.

7.3 Correlated metabolite and gene expression levels

7.3.1 Identification of potential signalling metabolites

The metabolic regulation of gene expression is well described for a range of metabolites including amino acids, sugars and organic acids (Monroy and Schwartzbach, 1984; Sheen, 1990; Grierson et al., 1994; Koch, 1996; Tischner, 2000; Corruzzi and Bush, 2001; Gibson, 2005). Nutrient signalling has been shown to be involved in the division of meristem cells and a role for nitrate in the regulation of seed germination has been suggested (Batak et al., 2002; Alboresi et al., 2005; Francis and Halford, 2006). This raises the possibility that metabolic signalling might also play a role in germination and seedling establishment. In order to identify metabolites with the potential to act as regulators of gene expression, correlated gene expression and metabolite levels were identified throughout the time course. Although correlation of levels is not an absolute indication of a regulatory relationship, the identified pairs of genes and metabolites provide a focus for the basis of future work. An important assumption made in this work was that the expression of metabolite-responsive genes follows an immediate and linear relationship. The calculation of rank correlations which are sensitive to nonlinear monotonic relationships would

widen the set of potential interactions able to be detected. It would also be interesting to overlay the metabolite and gene expression profiles onto metabolic maps using programs such as MAPMAN (Thimm et al., 2004) in order to visualise data in the context of previously described networks.

7.3.2 Investigation into direction of relationships

A number of correlations were identified between metabolites and genes during development. A number of these correlations had previously been described in the literature, demonstrating the validity of correlation analysis for this purpose. Amongst the 210 genes correlated with metabolites, nearly 50 were involved in transcriptional, translational or signalling processes. The further investigation of genes such as these may provide more detailed information on the regulation of plant development. Of these genes three immediately present themselves as suitable candidates for investigation; the phytochrome interacting factors 1 and 4 (*PIF1*, *PIF4*) and abscisic acid insensitive 3 (*ABI3*) genes encode transcription factors which have previously been implicated in the regulation of seed germination. The phytochrome-mediated degradation of *PIF1* and *PIF4* are associated with the promotion of germination while *ABI3* exhibits insensitivity to the abscisic acid-induced inhibition of germination (Koornneef et al., 1984; Oh et al., 2006). The production and subsequent analysis of metabolite levels for *pif1*, *pif4* and *abi3* knock-out mutants during the developmental time course may provide information relating to the nature of the relationship; if the metabolite levels remain at a similar level in the mutant to the wild-type then the level of control does not rest with the gene. Although this is not conclusive evidence of a regulatory relationship, but this information may guide more detailed investigations of relationships, such as the elucidation of signalling pathways. The three genes identified as potential candidates for mutant analysis are of particular interest as they have already been identified as responsive to other developmental regulators; the further characterisation of potential regulatory relationships will help elucidate the details of signalling networks during development.

7.3.3 The application of correlation analysis to other systems

The analytical system detailed in this work has provided a benchmark platform for the integration of metabolite and gene expression profiles. The identification of a number of correlations between genes and metabolites previously detailed in the literature demonstrates the validity of spring embedding as an exploratory tool for the identification of potential regulatory mechanisms (Ebbels et al., 2006). Further applications for the identification of metabolites with the potential to influence gene expression using spring embedding are limited only by the ease of production of gene expression and metabolite profiles.

Appendix A

IUPAC codes

Ambiguity of bases in DNA sequences is represented by a code proposed by the International Union of Pure and Applied Chemistry (IUPAC):

A - adenine
C - cytosine
G - guanine
T - thymine
R - G or A
Y - T or C
K - G or T
M - A or C
S - G or C
W - A or T
B - G, C or T
D - G, A or T
H - A, C or T
V - G, C or A
N - A, C, G or T

Appendix B

Differentially expressed genes between *acn1* and Col-7

A combination of fold-change threshold and t-test ($p>0.05$) was used to identify 201 differentially expressed genes between *acn1* and Col-7.

Gene ID	Gene Title
At5g44120	12S seed storage protein (CRA1)
At2g19590	1-aminocyclopropane-1-carboxylate oxidase, putative / ACC oxidase, putative
At5g62490	ABA-responsive protein (HVA22b)
At1g48660	auxin-responsive GH3 family protein
At2g27480	calcium-binding EF hand family protein
At1g74820	cupin family protein
At3g26150	cytochrome P450 71B16, putative (CYP71B16)
At5g04110	DNA topoisomerase II family protein
At1g72070	DNAJ heat shock N-terminal domain-containing protein
At5g62210	embryo-specific protein-related
At5g07330	expressed protein
At3g12890	expressed protein
At1g33055	expressed protein
At1g75770	expressed protein
At2g11090	expressed protein
At5g39110	germin-like protein, putative
At2g25890	glycine-rich protein / oleosin
At3g45460	hypothetical protein
At3g44420	hypothetical protein
At1g23070	hypothetical protein
At1g52010	hypothetical protein
At2g29860	kelch repeat-containing F-box family protein
At1g03470	kinase interacting family protein
At4g36600	late embryogenesis abundant domain-containing protein / LEA domain-containing protein
At4g21020	late embryogenesis abundant domain-containing protein / LEA domain-containing protein
At3g17520	late embryogenesis abundant domain-containing protein / LEA domain-containing protein
At2g35300	late embryogenesis abundant group 1 domain-containing protein / LEA group 1 domain-containing protein
At2g36640	late embryogenesis abundant protein (ECP63) / LEA protein
At3g28470	myb family transcription factor (MYB35)
At2g37310	pentatricopeptide (PPR) repeat-containing protein
At5g51270	protein kinase family protein
At1g24090	RNase H domain-containing protein
At2g18080	serine carboxypeptidase S28 family protein
At3g29590	transferase family protein

At2g32790	ubiquitin-conjugating enzyme, putative
At4g35480	zinc finger (C3HC4-type RING finger) family protein
At1g68730	---
At5g10220	annexin 6 (ANN6)
At2g45110	beta-expansin, putative (EXPB4)
At3g44900	cation/hydrogen exchanger, putative (CHX4)
At1g14750	cyclin, putative (SDS)
At1g65670	cytochrome P450 family protein
At5g27680	DEAD/DEAH box helicase, putative
At3g51070	dehydration-responsive protein-related
At4g17505	expressed protein
At3g58540	expressed protein
At3g52770	expressed protein
At1g30800	expressed protein
At1g66000	expressed protein
At2g12240	hypothetical protein
At5g51860	MADS-box protein (AGL72)
At2g14210	MADS-box protein (ANR1)
At2g35030	pentatricopeptide (PPR) repeat-containing protein
At4g08670	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
At1g55610	protein kinase family protein
At1g33750	terpene synthase/cyclase family protein
At1g79680	wall-associated kinase, putative
At5g63360	---
At1g67920	---
At3g49460	60S acidic ribosomal protein-related
At3g28510	AAA-type ATPase family protein
At5g13350	auxin-responsive GH3 family protein
At1g66360	C2 domain-containing protein
At2g04620	cation efflux family protein
At2g34180	CBL-interacting protein kinase 13 (CIPK13)
At4g00310	expressed protein
At3g28310	expressed protein
At1g48580	expressed protein
At1g27590	expressed protein
At2g17920	expressed protein
At4g33610	glycine-rich protein
At1g56680	glycoside hydrolase family 19 protein
At3g43420	hypothetical protein
At1g22080	hypothetical protein
At1g75870	hypothetical protein
At2g06480	hypothetical protein
At1g26500	pentatricopeptide (PPR) repeat-containing protein
At5g19880	peroxidase, putative
At3g05170	phosphoglycerate/bisphosphoglycerate mutase family protein
At2g14270	protein phosphatase 2C, putative / PP2C, putative
At5g19920	transducin family protein / WD-40 repeat family protein
At4g05240	ubiquitin family protein
At1g49920	zinc finger protein-related
At5g12030	17.7 kDa class II heat shock protein 17.6A (HSP17.7-CII)
At5g57650	eukaryotic translation initiation factor-related
At2g35820	expressed protein
At4g36830	GNS1/SUR4 membrane family protein
At5g16460	hypothetical protein

At1g62530	hypothetical protein
At1g32560	late embryogenesis abundant group 1 domain-containing protein / LEA group 1 domain-containing protein
At3g06000	leucine-rich repeat family protein
At1g75030	pathogenesis-related thaumatin family protein
At3g03580	pentatricopeptide (PPR) repeat-containing protein
At2g24000	serine carboxypeptidase S10 family protein
At3g28210	zinc finger (AN1-like) family protein
At5g09470	mitochondrial substrate carrier family protein
At3g32350	---
At5g29080	hypothetical protein
At3g02480	ABA-responsive protein-related
At1g04560	AWPM-19-like membrane family protein
At5g66780	expressed protein
At1g21170	expressed protein
At1g71380	glycosyl hydrolase family 9 protein
At1g72100	late embryogenesis abundant domain-containing protein / LEA domain-containing protein
At5g06760	late embryogenesis abundant group 1 domain-containing protein / LEA group 1 domain-containing protein
At1g75830	plant defensin-fusion protein, putative (PDF1.1)
At2g02850	plastocyanin-like domain-containing protein / plantacyanin, putative
At2g47780	rubber elongation factor (REF) protein-related
At5g50600	short-chain dehydrogenase/reductase (SDR) family protein
At1g54870	short-chain dehydrogenase/reductase (SDR) family protein
At4g32440	agenet domain-containing protein
At1g71200	basic helix-loop-helix (bHLH) family protein
At5g57130	expressed protein
At5g42520	expressed protein
At2g23160	F-box family protein
At5g53640	F-box family protein
At2g03370	hypothetical protein
At1g65370	meprin and TRAF homology domain-containing protein / MATH domain-containing protein
At5g25370	phospholipase D, putative (PLDZETA)
At2g23990	plastocyanin-like domain-containing protein
At2g31000	protein kinase family protein
At2g33440	splicing factor family protein
At5g16480	tyrosine specific protein phosphatase family protein
At1g76590	zinc-binding family protein
At2g12220	---
At4g03480	ankyrin repeat family protein
At3g18090	DNA-directed RNA polymerase family protein
At1g23690	expressed protein
At5g44180	homeobox transcription factor, putative
At5g32610	hypothetical protein
At1g51820	leucine-rich repeat protein kinase, putative
At1g77990	sulfate transporter
At4g17505	expressed protein
At1g56680	glycoside hydrolase family 19 protein
At5g04110	DNA topoisomerase II family protein
At1g49920	zinc finger protein-related
At5g19920	transducin family protein / WD-40 repeat family protein
At5g19880	peroxidase, putative

At5g32610	hypothetical protein
At4g36830	GNS1/SUR4 membrane family protein
At4g36600	late embryogenesis abundant domain-containing protein / LEA domain-containing protein
At5g29080	hypothetical protein
At5g27680	DEAD/DEAH box helicase, putative
At5g66780	expressed protein
At5g63360	---
At5g62490	ABA-responsive protein (HVA22b)
At5g62210	embryo-specific protein-related
At5g57650	eukaryotic translation initiation factor-related
At5g51860	MADS-box protein (AGL72)
At5g51270	protein kinase family protein
At5g44120	12S seed storage protein (CRA1)
At5g44180	homeobox transcription factor, putative
At5g39110	germin-like protein, putative
At5g16460	hypothetical protein
At5g13350	auxin-responsive GH3 family protein
At5g12030	17.7 kDa class II heat shock protein 17.6A (HSP17.7-CII)
At5g10220	annexin 6 (ANN6)
At5g07330	expressed protein
At3g58540	expressed protein
At3g51070	dehydration-responsive protein-related
At3g49460	60S acidic ribosomal protein-related
At3g45460	hypothetical protein
At3g44420	hypothetical protein
At3g43420	hypothetical protein
At4g35480	zinc finger (C3HC4-type RING finger) family protein
At4g33610	glycine-rich protein
At4g21020	late embryogenesis abundant domain-containing protein / LEA domain-containing protein
At4g08670	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
At4g05240	ubiquitin family protein
At4g00310	expressed protein
At1g22080	hypothetical protein
At3g32350	---
At1g48660	auxin-responsive GH3 family protein
At1g72070	DNAJ heat shock N-terminal domain-containing protein
At1g54870	short-chain dehydrogenase/reductase (SDR) family protein
At1g32560	late embryogenesis abundant group 1 domain-containing protein / LEA group 1 domain-containing protein
At3g28210	zinc finger (AN1-like) family protein
At3g28510	AAA-type ATPase family protein
At3g28310	expressed protein
At3g29590	transferase family protein
At3g12890	expressed protein
At2g29860	kelch repeat-containing F-box family protein
At3g26150	cytochrome P450 71B16, putative (CYP71B16)
At3g28470	myb family transcription factor (MYB35)
At3g18090	DNA-directed RNA polymerase family protein
At3g17520	late embryogenesis abundant domain-containing protein / LEA domain-containing protein
At3g02480	ABA-responsive protein-related
At3g03580	pentatricopeptide (PPR) repeat-containing protein

At3g05170	phosphoglycerate/bisphosphoglycerate mutase family protein
At1g75030	pathogenesis-related thaumatin family protein
At1g67920	---
At1g66360	C2 domain-containing protein
At1g26500	pentatricopeptide (PPR) repeat-containing protein
At1g48580	expressed protein
At1g79680	wall-associated kinase, putative
At1g33055	expressed protein
At1g33750	terpene synthase/cyclase family protein
At1g77990	sulfate transporter
At1g74820	cupin family protein
At1g68730	---

Appendix C

Genes related to acetate metabolism

Transcript levels of genes associated with acetate metabolism were investigated between *acn1* and Col-7. No significant differences in expression were observed, other than those commented on in chapter 3.

Comatose	
At4g39850	
ACN1	
At3g16910	
Malate synthase	
At5g03860	malate synthase, putative
Isocitrate lyase	
At3g21720	isocitrate lyase, putative
Citrate synthase (mitochondrial)	
At2g44350	citrate synthase, miochondrial, putative
At3g60100	citrate synthase, miochondrial, putative
Citrate synthase (peroxisomal 1)	
At3g58740	
Citrate synthase (peroxisomal 2)	
At3g58750	
Citrate synthase (peroxisomal 3)	
At2g42790	
Aconitase 1	
At4g35830	
Aconitase 2	
At2g05710	
Aconitase 3	
At4g26970	
ATP: Citrate lyase	
At1g60810	ATP citrate-lyase -related
At1g09430	ATP-citrate synthase (ATP-citrate (pro-S-)-lyase/citrate cleavage enzyme), putative
At5g49460	ATP-citrate synthase, putative / ATP-citrate

	(pro-S-)-lyase, putative / citrate cleavage enzyme, putative
At3g06650	ATP-citrate synthase, putative / ATP-citrate (pro-S-)-lyase, putative / citrate cleavage enzyme, putative

Isocitrate dehydrogenase PLUS isoforms	
At5g14590	isocitrate dehydrogenase, putative / NADP+ isocitrate dehydrogenase, putative
At1g65930	isocitrate dehydrogenase, putative / NADP+ isocitrate dehydrogenase, putative
At5g03290	isocitrate dehydrogenase, putative / NAD+ isocitrate dehydrogenase, putative
At4g35260	isocitrate dehydrogenase subunit 1 / NAD+ isocitrate dehydrogenase subunit 1
At3g09810	isocitrate dehydrogenase, putative / NAD+ isocitrate dehydrogenase, putative
At4g35650	isocitrate dehydrogenase, putative / NAD+ isocitrate dehydrogenase, putative
At1g54340	isocitrate dehydrogenase, putative / NADP+ isocitrate dehydrogenase, putative
At2g17130	isocitrate dehydrogenase subunit 2 / NAD+ isocitrate dehydrogenase subunit 2

Malate dehydrogenase PLUS isoforms	
At1g04410	malate dehydrogenase, cytosolic, putative
At5g58330	malate dehydrogenase [NADP], chloroplast, putative
At5g09660	malate dehydrogenase, glyoxysomal
At2g22780	malate dehydrogenase, glyoxysomal, putative
At3g47520	malate dehydrogenase [NAD], chloroplast (MDH)
At3g15020	malate dehydrogenase [NAD], mitochondrial, putative
At1g53240	malate dehydrogenase [NAD], mitochondrial
At5g43330	malate dehydrogenase, cytosolic, putative
At3g53910	malate dehydrogenase-related
At5g56720	malate dehydrogenase, cytosolic, putative
At1g80560	3-isopropylmalate dehydrogenase, chloroplast, putative
At1g32480	isocitrate/isopropylmalate dehydrogenase family protein
At1g31180	3-isopropylmalate dehydrogenase, chloroplast, putative
At5g14200	3-isopropylmalate dehydrogenase, chloroplast, putative

Succinate dehydrogenase	
At5g66760	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial / flavoprotein subunit of complex II
At3g27380	succinate dehydrogenase, iron-sulphur

	subunit, mitochondrial (SDH2-1)
At5g40650	succinate dehydrogenase, iron-sulphur subunit, mitochondrial (SDH2-2)
At5g65165	succinate dehydrogenase, iron-sulphur subunit, mitochondrial (SDH2-3)
At2g18450	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial, putative / flavoprotein subunit of complex II, putative
At5g09600	succinate dehydrogenase cytochrome b subunit family protein
Fumarate	
At5g50950	fumarate hydratase, putative / fumarase, putative
At2g47510	fumarate hydratase, putative / fumarase, putative
Pyruvate dehydrogenase complex POH	
At5g50850	pyruvate dehydrogenase E1 component beta subunit, mitochondrial / PDHE1-B (PDH2)
At3g06483	pyruvate dehydrogenase (lipoamide) kinase (PDHK)
At1g01090	pyruvate dehydrogenase E1 component alpha subunit, chloroplast
At1g30120	pyruvate dehydrogenase E1 component beta subunit, chloroplast
At1g59900	pyruvate dehydrogenase E1 component alpha subunit, mitochondrial (PDHE1-A)
At1g24180	pyruvate dehydrogenase E1 component alpha subunit, mitochondrial, putative
Dihydrolipoyl transacetylase	
At1g34430	dihydrolipoamide S-acetyltransferase, putative
At1g54220	dihydrolipoamide S-acetyltransferase, putative
At3g52200	dihydrolipoamide S-acetyltransferase, putative
At3g25860	dihydrolipoamide S-acetyltransferase (LTA2)
At3g13930	dihydrolipoamide S-acetyltransferase, putative
Dihydrolipoyl dihydrogenase	
At4g16155	dihydrolipoamide dehydrogenase 2, plastidic / lipoamide dehydrogenase 2 (PTLPD2)
At1g48030	dihydrolipoamide dehydrogenase 1, mitochondrial / lipoamide dehydrogenase 1 (MTLPD1)
At3g17240	dihydrolipoamide dehydrogenase 2, mitochondrial / lipoamide dehydrogenase 2 (MTLPD2)
At3g16950	dihydrolipoamide dehydrogenase 1, plastidic

	/ lipoamide dehydrogenase 1 (PTLPD1)
Alpha-ketoglutarate dehydrogenase complex	
At5g65750	2-oxoglutarate dehydrogenase E1 component, putative / oxoglutarate decarboxylase, putative / alpha-ketoglutaric dehydrogenase, putative
At3g55410	2-oxoglutarate dehydrogenase E1 component, putative / oxoglutarate decarboxylase, putative / alpha-ketoglutaric dehydrogenase, putative
Succinyl-CoA ligase (succcinic thiokinase)	
At5g08300	succinyl-CoA ligase [GDP-forming] alpha-chain, mitochondrial, putative / succinyl-CoA synthetase, alpha chain, putative / SCS-alpha, putative
At2g20420	succinyl-CoA ligase [GDP-forming] beta-chain, mitochondrial, putative / succinyl-CoA synthetase, beta chain, putative / SCS-beta, putative
At5g23250	succinyl-CoA ligase [GDP-forming] alpha-chain, mitochondrial, putative / succinyl-CoA synthetase, alpha chain, putative / SCS-alpha, putative
Phosphoenoylpyruvate kinase	
At1g08650	phosphoenolpyruvate carboxylase kinase
At1g53310	phosphoenolpyruvate carboxylase kinase
At1g68750	phosphoenolpyruvate carboxylase kinase
At2g42600	phosphoenolpyruvate carboxylase kinase
At3g04530	phosphoenolpyruvate carboxylase kinase
At3g14940	phosphoenolpyruvate carboxylase kinase
At3g42628	phosphoenolpyruvate carboxylase kinase
At4g37870	phosphoenolpyruvate carboxylase kinase
At5g65690	phosphoenolpyruvate carboxylase kinase
Aspartate aminotransferase	
At4g31990	aspartate aminotransferase, chloroplast / transaminase A (ASP5) (AAT1)
At5g19550	aspartate aminotransferase, cytoplasmic isozyme 1 / transaminase A (ASP2)
At2g30970	aspartate aminotransferase, mitochondrial / transaminase A (ASP1)
At1g62800	aspartate aminotransferase, cytoplasmic isozyme 2 / transaminase A (ASP4)
At5g11520	aspartate aminotransferase, chloroplast / transaminase A (ASP3) (YLS4)
Glutamine synthetase	
At3g53180	glutamine synthetase, putative
At5g37600	glutamine synthetase, putative
At5g16570	glutamine synthetase, putative
At3g17820	glutamine synthetase (GS1)

At5g35630	glutamine synthetase (GS2)
At1g66200	glutamine synthetase, putative
At1g48470	glutamine synthetase, putative
Glutamate dehydrogenase	
At1g51720	glutamate dehydrogenase, putative
At3g03910	glutamate dehydrogenase, putative
At5g18170	glutamate dehydrogenase 1 (GDH1)
At5g07440	glutamate dehydrogenase 2 (GDH2)
At1g42490	pseudogene, glutamate dehydrogenase [fragment]

Appendix D

Genes with an altered response to sucrose in *acn1* compared to Col-7

The genes identified as sucrose responsive in Col-7 which demonstrated an altered response to sucrose in *acn1* and *visa versa* are listed in the table below. Gene IDs are provided along with a description of gene function and the cluster each gene was assigned to (chapter 4).

Gene ID	Gene Title	Cluster No
At5g64000	3'(2'),5'-bisphosphate nucleotidase, putative / inositol polyphosphate 1-phosphatase, putative	1
At5g37810	major intrinsic family protein / MIP family protein	1
At5g35770	sterile apetala (SAP)	1
At5g24410	glucosamine/galactosamine-6-phosphate isomerase-related	1
At4g25000	alpha-amylase, putative / 1,4-alpha-D-glucan glucanohydrolase, putative	1
At4g19940	F-box family protein	1
At2g33990	calmodulin-binding family protein	1
At3g16090	zinc finger (C3HC4-type RING finger) family protein	1
At1g34420	leucine-rich repeat family protein / protein kinase family protein	1
At1g43590	hypothetical protein	1
At1g10380	expressed protein	1
At1g03740	protein kinase family protein	1
At2g20800	pyridine nucleotide-disulphide oxidoreductase family protein	1
At2g25850	nucleotidyltransferase family protein	1
At1g77340	pentatricopeptide (PPR) repeat-containing protein	2
At5g15840	zinc finger protein CONSTANS (CO)	2
At5g57130	expressed protein	2
At5g53640	F-box family protein	2
At5g49680	cell expansion protein, putative	2
At5g41410	homeodomain protein (BEL1)	2
At5g12020	17.6 kDa class II heat shock protein (HSP17.6-CII)	2
At3g60980	pentatricopeptide (PPR) repeat-containing protein	2
At3g55190	esterase/lipase/thioesterase family protein	2
At3g49280	hypothetical protein	2
At4g19430	expressed protein	2
At4g09930	avirulence-responsive family protein / avirulence induced gene (AIG1) family protein	2
At3g10950	60S ribosomal protein L37a (RPL37aB)	2
At3g20140	cytochrome P450 family protein	2
At3g26790	transcriptional regulator (FUSCA3)	2
At3g02280	flavodoxin family protein	2
At1g63300	expressed protein	2

At5g34950	replication protein-related	2
At1g74200	leucine-rich repeat family protein	2
At2g04870	hypothetical protein	2
At1g09610	expressed protein	2
At1g76950	zinc finger protein (PRAF1) / regulator of chromosome condensation (RCC1) family protein	2
At2g42980	aspartyl protease family protein	2
At2g46790	pseudo-response regulator 9 (APRR9) / timing of CAB expression 1-like protein (TL1)	2
At5g15020	paired amphipathic helix repeat-containing protein	3
At5g33200	hypothetical protein	3
At5g64590	hypothetical protein	3
At5g55930	oligopeptide transporter OPT family protein	3
At5g55800	DC1 domain-containing protein	3
At5g49250	hypothetical protein	3
At5g24010	protein kinase family protein	3
At4g36110	auxin-responsive protein, putative	3
At4g11200	hypothetical protein	3
At4g08490	hypothetical protein	3
At4g01640	hypothetical protein	3
At3g25810	myrcene/ocimene synthase, putative	3
At2g24850	aminotransferase, putative	3
At2g01190	octicosapeptide/Phox/Bem1p (PB1) domain-containing protein	3
At4g15260	UDP-glucoronosyl/UDP-glucosyl transferase family protein	4
At5g28160	kelch repeat-containing F-box family protein	4
At5g46270	disease resistance protein (TIR-NBS-LRR class), putative	4
At5g44260	zinc finger (CCCH-type) family protein	4
At5g24150	squalene monooxygenase 1,1 / squalene epoxidase 1,1 (SQP1,1)	4
At5g10140	MADS-box protein flowering locus F (FLF)	4
At3g55970	oxidoreductase, 2OG-Fe(II) oxygenase family protein	4
At3g53720	cation/hydrogen exchanger, putative (CHX20)	4
At3g53940	mitochondrial substrate carrier family protein	4
At4g24050	short-chain dehydrogenase/reductase (SDR) family protein	4
At1g07330	hypothetical protein	4
At1g75890	family II extracellular lipase 2 (EXL2)	4
At2g16380	SEC14 cytosolic factor family protein / phosphoglyceride transfer family protein	4
At2g29310	tropinone reductase, putative / tropine dehydrogenase, putative	4
At2g47720	expressed protein	4
At2g02520	hypothetical protein	4
At5g28630	glycine-rich protein	5
At5g27680	DEAD/DEAH box helicase, putative	5
At5g66850	protein kinase family protein	5
At5g64510	expressed protein	5
At5g60070	ankyrin repeat family protein	5
At5g60130	transcriptional factor B3 family protein	5
At5g55010	hypothetical protein	5
At5g45530	expressed protein	5
At5g43890	flavin-containing monooxygenase family protein / FMO family protein	5
At5g35520	kinetochore protein-related	5
At5g18820	chaperonin, putative	5
At5g14070	glutaredoxin family protein	5
At5g13230	pentatricopeptide (PPR) repeat-containing protein	5
At5g03490	UDP-glucoronosyl/UDP-glucosyl transferase family protein	5

At3g58290	meprin and TRAF homology domain-containing protein / MATH domain-containing protein	5
At3g57480	zinc finger (C2H2 type, AN1-like) family protein	5
At3g54420	class IV chitinase (CHIV)	5
At3g43630	nodulin, putative	5
At4g20800	FAD-binding domain-containing protein	5
At4g20240	cytochrome P450, putative	5
At4g08610	hypothetical protein	5
At2g37950	zinc finger (C3HC4-type RING finger) family protein	5
At3g18710	U-box domain-containing protein	5
At3g22990	expressed protein	5
At3g05140	protein kinase family protein	5
At1g15045	glutamine amidotransferase-related	5
At1g43715	---	5
At1g21360	expressed protein	5
At1g64210	leucine-rich repeat transmembrane protein kinase, putative	5
At1g36100	myosin heavy chain-related	5
At2g31480	expressed protein	5
At2g06260	hypothetical protein	5
At1g78995	expressed protein	5
At1g60260	glycosyl hydrolase family 1 protein	5
At1g23840	expressed protein	5
At2g32710	kip-related protein 4 (KRP4) / cyclin-dependent kinase inhibitor 4 (ICK4)	5
At5g54740	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	6
At5g43840	heat shock transcription factor family protein	6
At5g40590	DC1 domain-containing protein	6
At5g22590	hypothetical protein	6
At5g02690	expressed protein	6
At3g46230	17.4 kDa class I heat shock protein (HSP17.4-CI)	6
At4g27450	expressed protein	6
At4g08880	Ulp1 protease family protein	6
At3g16910	AMP-dependent synthetase and ligase family protein	6
At1g12520	superoxide dismutase copper chaperone, putative	6
At1g09440	protein kinase family protein	6
At2g15010	thionin, putative	6
At1g26090	hypothetical protein	7
At5g26060	S1 self-incompatibility protein-related	7
At5g58400	peroxidase, putative	7
At4g00130	hypothetical protein	7
At1g43610	NLI interacting factor (NIF) family protein	7
At2g07698	ATP synthase alpha chain, mitochondrial, putative	7
At5g20480	leucine-rich repeat transmembrane protein kinase, putative	8
At5g32610	hypothetical protein	8
At5g62165	MADS-box protein (AGL42)	8
At5g42700	transcriptional factor B3 family protein	8
At5g11870	expressed protein	8
At3g51750	expressed protein	8
At4g10440	dehydration-responsive family protein	8
At4g05280	Ulp1 protease family protein	8
At4g03900	---	8
At3g31915	hypothetical protein	8
At1g49520	SWIB complex BAF60b domain-containing protein	8
At1g67810	Fe-S metabolism associated domain-containing protein	9

At1g57570	jacalin lectin family protein	9
At5g04840	bZIP protein	9
At5g66060	oxidoreductase, 2OG-Fe(II) oxygenase family protein	9
At5g65650	expressed protein	9
At5g63610	protein kinase, putative	9
At5g53820	expressed protein	9
At5g51570	band 7 family protein	9
At5g51270	protein kinase family protein	9
At5g40420	glycine-rich protein / oleosin	9
At3g51570	disease resistance protein (TIR-NBS-LRR class), putative	9
At3g48550	expressed protein	9
At4g22470	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	9
At4g19780	---	9
At4g18610	expressed protein	9
At3g12580	heat shock protein 70, putative / HSP70, putative	9
At1g72070	DNAJ heat shock N-terminal domain-containing protein	9
At3g12890	expressed protein	9
At3g14510	geranylgeranyl pyrophosphate synthase, putative / GGPP synthetase, putative / farnesyltranstransferase, putative	9
At3g03580	pentatricopeptide (PPR) repeat-containing protein	9
At1g35330	zinc finger (C3HC4-type RING finger) family protein	9
At1g07520	scarecrow transcription factor family protein	9
At1g66000	expressed protein	9
At1g68730	---	9
At1g75940	glycosyl hydrolase family 1 protein / anther-specific protein ATA27	9
At2g11600	hypothetical protein	9
At2g04720	---	9
At2g36590	proline transporter, putative	9
At2g21890	mannitol dehydrogenase, putative	9
At1g61210	WD-40 repeat family protein / katanin p80 subunit, putative	9
At1g52010	hypothetical protein	9
At2g37310	pentatricopeptide (PPR) repeat-containing protein	9
At2g45970	cytochrome P450, putative	9
At4g25560	myb family transcription factor (MYB18)	10
At4g17505	expressed protein	10
At4g16740	terpene synthase/cyclase family protein	10
At5g25120	cytochrome P450 family protein	10
At5g62080	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	10
At5g11940	subtilase family protein	10
At3g58280	hypothetical protein	10
At3g51070	dehydration-responsive protein-related	10
At3g50730	protein kinase, putative	10
At4g34138	UDP-glucoronosyl/UDP-glucosyl transferase family protein	10
At4g32000	protein kinase family protein	10
At4g25510	hypothetical protein	10
At4g05210	bacterial transferase hexapeptide repeat-containing protein	10
At1g30800	expressed protein	10
At3g09020	alpha 1,4-glycosyltransferase family protein / glycosyltransferase sugar-binding DXD motif-containing protein	10
At1g10920	disease resistance protein (CC-NBS-LRR class), putative	10
At1g70880	Bet v I allergen family protein	10
At1g64010	serpin, putative / serine protease inhibitor, putative	10
At1g65670	cytochrome P450 family protein	10

At2g12240	hypothetical protein	10
At5g32490	hypothetical protein	11
At5g60030	expressed protein	11
At5g54400	expressed protein	11
At5g45240	disease resistance protein (TIR-NBS-LRR class), putative	11
At5g37150	tRNA-splicing endonuclease positive effector-related	11
At3g53200	myb family transcription factor (MYB27)	11
At3g42940	expressed protein	11
At4g12180	---	11
At4g09270	hypothetical protein	11
At4g04690	F-box family protein (FBX15)	11
At1g36240	60S ribosomal protein L30 (RPL30A)	11
At1g32570	hypothetical protein	11
At1g05480	SNF2 domain-containing protein / helicase domain-containing protein	11
At2g35870	expressed protein	11
At2g18680	expressed protein	11
At2g03160	E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 (At19), putative	11
At4g14980	DC1 domain-containing protein	12
At5g28890	hypothetical protein	12
At4g36700	cupin family protein	12
At5g25960	hypothetical protein	12
At5g65160	tetratricopeptide repeat (TPR)-containing protein	12
At5g59680	leucine-rich repeat protein kinase, putative	12
At5g47270	disease resistance protein-related	12
At5g13680	IKI3 family protein	12
At5g07520	glycine-rich protein (GRP18)	12
At3g50650	scarecrow-like transcription factor 7 (SCL7)	12
At3g43170	hypothetical protein	12
At4g37960	expressed protein	12
At4g34880	amidase family protein	12
At4g27670	25.3 kDa small heat shock protein, chloroplast precursor (HSP25.3-P)	12
At4g24110	expressed protein	12
At4g19840	lectin-related	12
At4g10780	disease resistance protein (CC-NBS-LRR class), putative	12
At4g08370	proline-rich extensin-like family protein	12
At4g07460	hypothetical protein	12
At4g07670	protease-associated (PA) domain-containing protein	12
At1g30140	hypothetical protein	12
At1g66510	AAR2 protein family	12
At3g11050	ferritin, putative	12
At1g42540	glutamate receptor family protein (GLR3.3)	12
At3g29390	hydroxyproline-rich glycoprotein family protein	12
At3g20030	F-box family protein	12
At3g28870	hypothetical protein	12
At3g16890	pentatricopeptide (PPR) repeat-containing protein	12
At3g20840	ovule development protein, putative	12
At3g14580	pentatricopeptide (PPR) repeat-containing protein	12
At3g23630	adenylate isopentenyltransferase 7 / cytokinin synthase (IPT7)	12
At3g09710	calmodulin-binding family protein	12
At3g02060	DEAD/DEAH box helicase, putative	12
At3g10200	dehydration-responsive protein-related	12
At3g01570	glycine-rich protein / oleosin	12

At3g07690	NAD-dependent glycerol-3-phosphate dehydrogenase family protein	12
At1g77500	expressed protein	12
At1g06770	zinc finger (C3HC4-type RING finger) family protein	12
At1g43800	acyl-(acyl-carrier-protein) desaturase, putative / stearoyl-ACP desaturase, putative	12
At1g26420	FAD-binding domain-containing protein	12
At1g76070	expressed protein	12
At1g59670	glutathione S-transferase, putative	12
At1g65090	expressed protein	12
At2g42560	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	12
At2g24560	GDSL-motif lipase/hydrolase family protein	12
At1g60380	apical meristem formation protein-related	12
At1g27470	transducin-related / WD-40 repeat protein-related	12
At1g76910	hypothetical protein	12
At1g61550	S-locus protein kinase, putative	12
At1g03770	zinc finger (C3HC4-type RING finger) family protein	12
At2g13920	DC1 domain-containing protein	12
At2g07750	DEAD box RNA helicase, putative	12
At4g36600	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	13
At5g14840	hypothetical protein	13
At5g16600	myb family transcription factor (MYB43)	13
At5g07330	expressed protein	13
At4g27657	expressed protein	13
At4g21020	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	13
At1g54870	short-chain dehydrogenase/reductase (SDR) family protein	13
At3g28210	zinc finger (AN1-like) family protein	13
At3g13770	pentatricopeptide (PPR) repeat-containing protein	13
At3g15170	cup-shaped cotyledon1 protein / CUC1 protein (CUC1)	13
At2g19040	rapid alkalinization factor (RALF) family protein	13
At3g26150	cytochrome P450 71B16, putative (CYP71B16)	13
At3g25170	rapid alkalinization factor (RALF) family protein	13
At3g27550	group II intron splicing factor CRS1-related	13
At3g17520	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	13
At1g75030	pathogenesis-related thaumatin family protein	13
At2g21770	cellulose synthase, catalytic subunit, putative	13
At1g65480	flowering locus T protein (FT)	13
At2g35300	late embryogenesis abundant group 1 domain-containing protein / LEA group 1 domain-containing protein	13
At2g33850	expressed protein	13
At1g26310	MADS-box protein, putative	14
At5g63710	leucine-rich repeat transmembrane protein kinase, putative	14
At5g61260	chromosome scaffold protein-related	14
At5g56310	pentatricopeptide (PPR) repeat-containing protein	14
At3g58630	expressed protein	14
At3g49900	BTB/POZ domain-containing protein	14
At4g39190	expressed protein	14
At4g29170	Mnd1 family protein	14
At4g28020	expressed protein	14
At4g23830	leucine-rich repeat family protein	14
At1g32590	---	14
At3g21040	---	14

At3g25840	protein kinase family protein	14
At3g19300	protein kinase family protein	14
At3g17360	kinesin motor protein-related	14
At3g03690	glycosyltransferase family 14 protein / core-2/I-branching enzyme family protein	14
At1g74890	two-component responsive regulator / response regulator 15 (ARR15)	14
At1g62690	expressed protein	14
At1g16340	2-dehydro-3-deoxyphosphooctonate aldolase, putative / phospho-2-dehydro-3-deoxyoctonate aldolase, putative / 3-deoxy-D-manno-octulosonic acid 8-phosphate synthetase, putative	14
At1g59730	thioredoxin, putative	14
At2g04600	hypothetical protein	14
At1g77170	pentatricopeptide (PPR) repeat-containing protein	14
At2g18650	zinc finger (C3HC4-type RING finger) family protein	14
At2g34350	expressed protein	14
At2g40070	expressed protein	14
At5g62210	embryo-specific protein-related	15
At5g54120	expressed protein	15
At5g37730	expressed protein	15
At5g23405	high mobility group (HMG1/2) family protein	15
At5g19110	extracellular dermal glycoprotein-related / EDGP-related	15
At3g51320	pentatricopeptide (PPR) repeat-containing protein	15
At4g30720	expressed protein	15
At1g48660	auxin-responsive GH3 family protein	15
At3g12130	KH domain-containing protein / zinc finger (CCCH type) family protein	15
At1g33265	expressed protein	15
At3g29590	transferase family protein	15
At1g70720	invertase/pectin methylesterase inhibitor family protein	15
At1g33750	terpene synthase/cyclase family protein	15
At2g04840	F-box family protein	15
At2g35030	pentatricopeptide (PPR) repeat-containing protein	15
At3g30140	GTP-binding family protein	16
At1g56570	pentatricopeptide (PPR) repeat-containing protein	16
At5g19920	transducin family protein / WD-40 repeat family protein	16
At5g25580	expressed protein	16
At5g58160	formin homology 2 domain-containing protein / FH2 domain-containing protein	16
At5g51670	expressed protein	16
At5g24655	expressed protein	16
At5g10190	transporter-related	16
At3g54320	ovule development protein, putative	16
At4g22500	---	16
At4g03820	expressed protein	16
At3g23850	hypothetical protein	16
At3g26590	MATE efflux family protein	16
At3g26614	---	16
At3g26190	cytochrome P450 71B21, putative (CYP71B21)	16
At3g17720	pyridoxal-dependent decarboxylase family protein	16
At3g18090	DNA-directed RNA polymerase family protein	16
At3g06010	homeotic gene regulator, putative	16
At3g05470	formin homology 2 domain-containing protein / FH2 domain-containing protein	16
At1g73860	kinesin motor protein-related	16
At3g43350	helicase-related	16

At1g16410	cytochrome P450, putative	16
At2g17020	F-box family protein (FBL10)	16
At1g77050	DEAD/DEAH box helicase, putative	16
At1g27050	homeobox-leucine zipper family protein	16
At2g37090	glycosyl transferase family 43 protein	16
At2g37320	pentatricopeptide (PPR) repeat-containing protein	16
At2g46780	RNA recognition motif (RRM)-containing protein	16

Appendix E

Identified correlations between genes and metabolites

A total of 237 correlations (>0.7) from 20 metabolites and 210 genes were identified using spring embedding (Ebbels et al., 2006). The following table lists the identified correlations between genes and metabolites and provides the correlation coefficients.

Metabolite	Gene ID	Gene Descriptor	Correlation Coefficient
Alanine	At1g61770	DNAJ heat shock N-terminal domain-containing protein, similar to SP:Q9UBS4 DnaJ homolog subfamily B member 11 precursor Homo sapiens; contains Pfam profile PF00226 DnaJ domain	0.9548
Alanine	At2g03980	GDSL-motif lipase/hydrolase family protein, similar to Anther-specific proline-rich protein APG from Brassica napus (SP:P40603), Arabidopsis thaliana (GI:22599); contains Pfam profile PF00657: Lipase/Acylhydrolase with GDSL-like motif	0.9581
Alanine	At3g16030	lectin protein kinase family protein, contains Pfam domains PF00069: Protein kinase domain and PF01453: Lectin (probable mannose binding)	-0.9519
Alanine	At4g30280	putative xyloglucan endotransglycosylase/hydrolase, expressed in the mature or basal regions of both the main and lateral roots, but not in the tip of these roots where cell division occurs.	0.9533
Choline	At1g12140	flavin-containing monooxygenase family protein	-0.9547
Citrate	At1g22690	Gibberellin-responsive protein, putative	-0.9574
Citrate	At1g80440	Kelch repeat-containing F-box family protein	-0.9582
Citrate	At2g23600	Hydrolase, alpha/beta fold family protein	-0.9652
Citrate	At3g03640	Glycosyl hydrolase family 1 protein	0.9531
Citrate	At4g20260	DREPP plasma membrane polypeptide family protein (cold response)	-0.9565
Citrate	At5g14260	SET domain containing protein, similarity to Rubisco small subunit N-methyltransferase I	-0.9785

		(chloroplast/nucleus)	
Citrate	At5g50760	Auxin-responsive family protein (mitochondria)	-0.9714
Formate	At5g01350	expressed protein	-0.954
Fructose	At2g33520	expressed protein	0.9501
Fumarate	At2g31370	bZIP transcription factor	-0.9495
Fumarate	At3g57060	non-SMC condensin subunit, XCAP-D2/Cnd1 family protein,	0.953
Fumarate	At5g53180	polypyrimidine tract-binding protein, putative / heterogeneous nuclear ribonucleoprotein, putative	-0.9593
Glucose	At5g51840	expressed protein	-0.9678
Glutamine	At1g73970	expressed protein	-0.955
Glutamine	At3g09630	60S ribosomal protein L4/L1 (RPL4A), strong similarity to 60S ribosomal protein L1 GB:P49691	-0.9575
Glutamine	At3g19170	Zinc metalloprotease pitrilysin subfamily A. Signal peptide degrading enzyme targeted to mitochondria and chloroplasts. Expressed only in siliques and flowers	-0.9688
Glutamine	At4g16830	nuclear RNA-binding protein (RGGA), identical to nuclear RNA binding protein GI:6492264 from (Arabidopsis thaliana)	-0.9509
Glutamine	At5g14270	DNA-binding bromodomain-containing protein, contains bromodomain, INTERPRO:IPR001487	-0.9566
Glutamine	At5g46430	60S ribosomal protein L32 (RPL32B)	-0.9501
Glutamine	At5g46430	60S ribosomal protein L32 (RPL32B)	-0.9573
Lactate	At1g05010	1-aminocyclopropane-1-carboxylate oxidase / ACC oxidase / ethylene-forming enzyme	-0.9732
Lactate	At1g05070	expressed protein	-0.9502
Lactate	At1g06430	encodes a FtsH protease that is localized to the chloroplast	-0.9663
Lactate	At1g13245	expressed protein	-0.949
Lactate	At1g67090	ribulose bisphosphate carboxylase small chain 1A	-0.9643
Lactate	At1g75750	gibberellin-regulated protein 1	-0.964
Lactate	At2g35370	glycine cleavage system H protein 1, mitochondrial. Involved in photorespiration	-0.971
Lactate	At2g46600	calcium-binding protein	-0.9654
Lactate	At3g26740	transcripts are differentially regulated at the level of mRNA stability at different times of day controlled by the circadian clock. mRNAs are targets of the mRNA degradation pathway mediated by the downstream (DST) instability determinant.	-0.9505
Lactate	At3g50820	Encodes a protein which is an extrinsic subunit of photosystem II	-0.9614
Lactate	At3g51920	calmodulin-9	-0.961
Lactate	At3g53420	plasma membrane intrinsic protein	-0.9535

Lactate	At3g59780	expressed protein	-0.9549
Lactate	At4g10060	expressed protein	-0.9555
Lactate	At4g12470	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	-0.9513
Lactate	At4g37230	oxygen-evolving enhancer protein putative / 33 kDa subunit of oxygen evolving system of photosystem II,	-0.9525
Lactate	At4g37300	expressed protein	-0.9604
Lactate	At5g01530	chlorophyll A-B binding protein	-0.9506
Lactate	At5g07020	proline-rich family protein thylakoid membrane	-0.9669
Lactate	At5g10410	protein / clathrin assembly protein-related	-0.9676
Lactate	At5g24530	oxidoreductase,	-0.9489
Malate	At2g37530	expressed protein	-0.9771
Malate	At3g03640	glycosyl hydrolase family 1 protein	0.9567
Malate	At4g39510	cytochrome P450 family protein	0.9618
Rhamnose	At2g02100	plant defensin-fusion protein, putative (PDF2.2),	-0.9613
Sucrose	At1g07080	gamma interferon responsive lysosomal thiol reductase family protein / GILT family protein, similar to SP:P13284 Gamma-interferon inducible lysosomal thiol reductase precursor {Homo sapiens}; contains Pfam profile PF03227: Gamma interferon inducible lysosomal thiol reductase (GILT)	0.9924
Sucrose	At1g07950	surfeit locus protein 5 family protein / SURF5 family protein	0.9821
Sucrose	At1g09570	phytochrome A (PHYA), identical to SP:P14712 Phytochrome A {Arabidopsis thaliana}	0.9777
Sucrose	At1g10270	pentatricopeptide (PPR) repeat-containing protein, contains Pfam profile: PF01535 PPR repeat; similar to ESTs gb:R30192 and gb:AA651017	0.9651
Sucrose	At1g15440	transducin family protein / WD-40 repeat family protein, Strong similarity to gb X95263 Periodic tryptophan protein 2 gene (PWP2) from Homo sapiens and contains 6 WD40, G-beta repeat domains	0.9726
Sucrose	At1g16470	20S proteasome alpha subunit B (PAB1) (PRC3), identical to proteasome subunit alpha type 2 SP:O23708, GI:6093778; identical to cDNA proteasome subunit prc3 GI:2511573	0.9596
Sucrose	At1g16740	ribosomal protein L20 family protein, similar to ribosomal protein L20 GI:3603025 from (Guillardia theta)	0.9658
Sucrose	At1g18850	expressed protein	0.9549
Sucrose	At1g26740	expressed protein, similar to 50S ribosomal protein L32 (SP:P80339) {Thermus thermophilus}	0.9644

Sucrose	At1g26770	expansin, putative (EXP10), similar to expansin At-EXP1 GI:1041702 from (Arabidopsis thaliana); alpha-expansin gene family, PMID:11641069	0.9928
Sucrose	At1g27150	expressed protein	0.9558
Sucrose	At1g29350	expressed protein	0.9521
Sucrose	At1g29800	similar to zinc finger family protein	0.9518
Sucrose	At1g30970	zinc finger (C2H2 type) family protein, contains Pfam domain PF00096: Zinc finger, C2H2 type	0.9672
Sucrose	At1g31175	expressed protein	0.9788
Sucrose	At1g44960	expressed protein	0.954
Sucrose	At1g51380	eukaryotic translation initiation factor 4A, putative / eIF-4A, putative	0.9867
Sucrose	At1g52380	Ran-binding protein 1 domain-containing protein / RanBP1 domain-containing protein, weak similarity to SP:Q09717 Ran-specific GTPase-activating protein 1 (Ran binding protein 1) (RANBP1) (Spi1-binding protein) {Schizosaccharomyces pombe}; contains Pfam profile PF00638: RanBP1 domain	0.9837
Sucrose	At1g52930	brix domain-containing protein, contains Pfam domain, PF04427: Brix domain	0.9903
Sucrose	At1g52980	GTP-binding family protein, contains Pfam domain, PF01926: GTPase of unknown function	0.9703
Sucrose	At1g55370	expressed protein	0.9687
Sucrose	At1g56110	nucleolar protein Nop56, putative, similar to XNop56 protein (Xenopus laevis) GI:14799394; contains Pfam profile PF01798: Putative snoRNA binding domain	0.9846
Sucrose	At1g59920	expressed protein	0.9824
Sucrose	At1g61040	plus-3 domain-containing protein, contains Pfam profile PF03126: Plus-3 domain	0.9495
Sucrose	At1g69530	expansin, putative (EXP1), identical to expansin (At-EXP1) (Arabidopsis thaliana) GI:1041702; alpha-expansin gene family, PMID:11641069	0.9662
Sucrose	At1g70350	expressed protein	0.9496
Sucrose	At1g75200	flavodoxin family protein / radical SAM domain-containing protein, contains Pfam profiles PF00258: Flavodoxin, PF04055: radical SAM domain protein	0.9648
Sucrose	At1g76300	small nuclear ribonucleoprotein D3, putative / snRNP core protein D3, putative / Sm protein D3, putative, similar to SWISS-PROT:P43331 small nuclear ribonucleoprotein Sm D3 (snRNP core protein D3, Sm-D3) (Mouse)	0.9512

Sucrose	At1g77120	alcohol dehydrogenase (ADH), identical to alcohol dehydrogenase GI:469467 from (Arabidopsis thaliana)	0.9906
Sucrose	At1g77510	Encodes a protein disulfide isomerase-like (PDIL) protein, a member of a multigene family within the thioredoxin (TRX) superfamily.	0.9632
Sucrose	At1g80750	60S ribosomal protein L7 (RPL7A), similar to ribosomal protein L7 GB:AAA03081 GI:307388 from (Homo sapiens)	0.968
Sucrose	At2g07770	hypothetical protein	0.9627
Sucrose	At2g15130	plant basic secretory protein (BSP) family protein, similar to NtPRp27 (Nicotiana tabacum) GI:5360263; contains Pfam profile PF04450: Plant Basic Secretory Protein	0.9551
Sucrose	At2g17190	ubiquitin family protein, contains INTERPRO:IPR000626 ubiquitin domain	0.9637
Sucrose	At2g19080	metaxin-related, contains 1 transmembrane domain; similar to Metaxin 1 (component of a preprotein import complex) (Swiss-Prot:P47802) (Mus musculus);	0.9582
Sucrose	At2g19540	transducin family protein / WD-40 repeat family protein, contains WD-40 repeats (PF00400); similar to Glutamate-rich WD repeat protein (GRWD) (SP:Q9BQ67)(Homo sapiens)	0.9682
Sucrose	At2g19640	SET domain-containing protein, contains Pfam profile PF00856: SET domain	0.9558
Sucrose	At2g20180	basic helix-loop-helix (bHLH) family protein, contains Pfam domain, PF00010: Helix-loop-helix DNA-binding domain	0.9493
Sucrose	At2g20940	expressed protein	0.9861
Sucrose	At2g23110	expressed protein	0.9568
Sucrose	At2g24500	zinc finger (C2H2 type) family protein, contains Pfam profile: PF00096 zinc finger, C2H2 type	0.9628
Sucrose	At2g25740	ATP-dependent protease La (LON) domain-containing protein, low similarity to protease Lon (Pseudomonas fluorescens) GI:7644385; contains Pfam profile PF02190: ATP-dependent protease La (LON) domain	0.9543
Sucrose	At2g27020	20S proteasome alpha subunit G (PAG1) (PRC8), identical to proteasome subunit alpha type 3 SP:O23715, GI:12644056 from (Arabidopsis thaliana); identical to cDNA proteasome subunit prc8	0.9558

		GI:2511591	
Sucrose	At2g27200	GTP-binding family protein, contains Pfam domain, PF01926: GTPase of unknown function	0.9754
Sucrose	At2g27380	proline-rich family protein, contains proline-rich extensin domains, INTERPRO:IPR002965	0.9816
Sucrose	At2g33410	heterogeneous nuclear ribonucleoprotein, putative / hnRNP, putative	0.9771
Sucrose	At2g33730	DEAD box RNA helicase, putative, similar to SP:P23394 Pre-mRNA splicing factor RNA helicase PRP28 {Saccharomyces cerevisiae}; contains Pfam profiles PF00270: DEAD/DEAH box helicase, PF00271: Helicase conserved C-terminal domain	0.9771
Sucrose	At2g37020	translin family protein, similar to SP:Q62348 Translin {Mus musculus}; contains Pfam profile PF01997: Translin family	0.9522
Sucrose	At2g44860	60S ribosomal protein L24, putative	0.9587
Sucrose	At3g03950	expressed protein, contains Pfam profile PF04146: YT521-B-like family	0.9571
Sucrose	At3g06530	BAP28-related, similar to Protein BAP28 (Swiss-Prot:Q9H583) (Homo sapiens)	0.9948
Sucrose	At3g07910	expressed protein	0.9671
Sucrose	At3g08900	reversibly glycosylated polypeptide-3 (RGP3), nearly identical to reversibly glycosylated polypeptide-3 (Arabidopsis thaliana) GI:11863238; contains non-consensus GA-donor splice site at intron 2	0.9643
Sucrose	At3g09440	heat shock cognate 70 kDa protein 3 (HSC70-3) (HSP70-3), identical to SP:O65719 Heat shock cognate 70 kDa protein 3 (Hsc70.3) {Arabidopsis thaliana}	0.9733
Sucrose	At3g09700	DNAJ heat shock N-terminal domain-containing protein, contains Pfam profile PF00226 DnaJ domain; similar to a region of DNAJ domain-containing protein MCJ GB:AAD38506	0.9695
Sucrose	At3g09840	cell division cycle protein 48 (CDC48A) (CDC48), identical to SP:P54609 Cell division cycle protein 48 homolog {Arabidopsis thaliana}	0.9727
Sucrose	At3g11270	26S proteasome non-ATPase regulatory subunit 7, putative / 26S proteasome regulatory subunit S12, putative / MOV34 protein, putative, contains similarity to 26S proteasome regulatory subunit S12 (MOV34) SP:P26516 from (Mus musculus)	0.9754

Sucrose	At3g12860	nucleolar protein Nop56, putative, similar to XNop56 protein (Xenopus laevis) GI:14799394; contains Pfam profile PF01798: Putative snoRNA binding domain	0.9668
Sucrose	At3g13150	pentatricopeptide (PPR) repeat-containing protein, contains Pfam profile PF01535: PPR repeat	0.9601
Sucrose	At3g13460	expressed protein, contains Pfam profile PF04146: YT521-B-like family	0.9613
Sucrose	At3g15000	expressed protein, similar to DAG protein (required for chloroplast differentiation and palisade development) GB:Q38732 (Antirrhinum majus)	0.9508
Sucrose	At3g16810	pumilio/Puf RNA-binding domain-containing protein, contains Pfam profile:PF00806 Pumilio-family RNA binding domains	0.9735
Sucrose	At3g17080	self-incompatibility protein-related, similar to S1 self-incompatibility protein GB:CAA52380 (Papaver rhoeas) (Proc. Natl. Acad. Sci. U.S.A. 91 (6), 2265-2269 (1994))	0.9722
Sucrose	At3g19990	expressed protein	0.9544
Sucrose	At3g22330	DEAD box RNA helicase, putative	0.9606
Sucrose	At3g24650	abscisic acid-insensitive protein 3 (ABI3), identical to abscisic acid-insensitive protein 3 GI:16146 SP:Q01593 from (Arabidopsis thaliana), (Plant Cell 4 (10), 1251-1261 (1992))	0.9721
Sucrose	At3g26340	20S proteasome beta subunit E, putative	0.9621
Sucrose	At3g28345	ABC transporter family protein, similar to P-glycoprotein (Arabidopsis thaliana) GI:3849833; contains Pfam profiles PF00005: ABC transporter, PF00664: ABC transporter transmembrane region	0.9688
Sucrose	At3g29090	pectinesterase family protein, similar to pectinesterase precursor GB:Q43043 (Petunia integrifolia); contains Pfam profile: PF01095 pectinesterase	0.9521
Sucrose	At3g44750	histone deacetylase, putative (HD2A), contains Pfam domain, PF00096: Zinc finger, C2H2 type; identical to cDNA putative histone deacetylase (HD2A) GI:11066134	0.9792
Sucrose	At3g49240	pentatricopeptide (PPR) repeat-containing protein, contains Pfam profile PF01535: PPR repeat	0.9507
Sucrose	At3g52040	expressed protein	0.9528
Sucrose	At3g53940	mitochondrial substrate carrier family protein	0.9699
Sucrose	At3g57150	dyskerin, putative / nucleolar protein NAP57, putative, similar to SP:P40615	0.969

		Dyskerin (Nucleolar protein NAP57) {Rattus norvegicus}; contains Pfam profiles PF01509: TruB family pseudouridylate synthase (N terminal domain), PF01472: PUA domain; supporting cDNA gi:8901185:gb:AF234984.2:AF234984	
Sucrose	At3g57780	expressed protein	0.9798
Sucrose	At3g58660	60S ribosomal protein-related, contains weak similarity to 60S ribosomal protein L10A (CSA-19) (NEDD-6) (Swiss-Prot:P53026) (Mus musculus)	0.9824
Sucrose	At3g60820	20S proteasome beta subunit F1 (PBF1)	0.9754
Sucrose	At4g01560	brix domain-containing protein, contains Pfam domain, PF04427: Brix domain	0.9582
Sucrose	At4g03060	2-oxoglutarate-dependent dioxygenase, putative (AOP2), nearly identical to GI:16118891; contains Pfam profile PF03171: 2OG-Fe(II) oxygenase superfamily domain. The gene sequence is frameshifted, this could be a pseudogene or a sequencing error may exist; identical to cDNA AOP2 GI:16118890	0.9603
Sucrose	At4g04940	transducin family protein / WD-40 repeat family protein, contains seven G-protein beta WD-40 repeats	0.9636
Sucrose	At4g04950	thioredoxin family protein, similar to PKCq-interacting protein PICOT from (Mus musculus) GI:6840949, (Rattus norvegicus) GI:6840951; contains Pfam profile PF00085: Thioredoxin	0.9566
Sucrose	At4g14520	DNA-directed RNA polymerase II-related,	0.9625
Sucrose	At4g16160	mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein, contains Pfam PF02466: Mitochondrial import inner membrane translocase subunit Tim17	0.9559
Sucrose	At4g16250	phytochrome D (PHYD), nearly identical to SP:P42497 Phytochrome D {Arabidopsis thaliana}	0.9581
Sucrose	At4g16515	expressed protein	0.9489
Sucrose	At4g17550	transporter-related, similar to cAMP inducible 2 protein (Mus musculus) GI:4580997, glycerol 3-phosphate permease (Homo sapiens) GI:7543982; contains Pfam profile PF00083: major facilitator superfamily protein	0.9593
Sucrose	At4g18910	aquaglyceroporin / NOD26-like major intrinsic protein 2 (NLM2), contains Pfam profile: MIP PF00230; similar to SP:P08995 {Glycine max} Nodulin-26	0.9581

		(N-26); identical to cDNA aquaglyceroporin (nlm2 gene) GI:11071655, aquaglyceroporin (Arabidopsis thaliana) GI:11071656	
Sucrose	At4g20980	eukaryotic translation initiation factor 3 subunit 7, putative / eIF-3 zeta, putative / eIF3d, putative, similar to initiation factor 3d (Arabidopsis thaliana) GI:12407755, SP:O15371 Eukaryotic translation initiation factor 3 subunit 7 (eIF-3 zeta) (eIF3 p66) (eIF3d) {Homo sapiens}; contains Pfam profile PF05091: Eukaryotic translation initiation factor 3 subunit 7 (eIF-3)	0.9816
Sucrose	At4g24800	MA3 domain-containing protein, similar to programmed cell death 4 protein (Gallus gallus) GI:12958564; contains Pfam profile PF02847: MA3 domain	0.9543
Sucrose	At4g25500	arginine/serine-rich splicing factor RSP40 (RSP40), identical to SP:P92965 Arginine/serine-rich splicing factor RSP40 {Arabidopsis thaliana}	0.9721
Sucrose	At4g28510	prohibitin, putative	0.9752
Sucrose	At4g29510	protein arginine N-methyltransferase, putative, similar to protein arginine N-methyltransferase 1-variant 2 (Homo sapiens) GI:7453575	0.9718
Sucrose	At4g31300	20S proteasome beta subunit A (PBA1) (PRCD), identical to cDNA proteasome subunit prcd GI:2511593	0.9512
Sucrose	At4g32520	glycine hydroxymethyltransferase, putative / serine hydroxymethyltransferase, putative / serine/threonine aldolase, putative, similar to serine hydroxymethyltransferase (Chlamydomonas reinhardtii) GI:17066746; contains Pfam profile PF00464: serine hydroxymethyltransferase	0.9608
Sucrose	At4g35570	high mobility group protein delta (HMGdelta) / HMG protein delta, identical to HMG protein (HMGdelta) (Arabidopsis thaliana) GI:2832363	0.9803
Sucrose	At4g36020	cold-shock DNA-binding family protein, contains Pfam domains, PF00313: 'Cold-shock' DNA-binding domain and PF00098: Zinc knuckle	0.9577
Sucrose	At5g02050	mitochondrial glycoprotein family protein / MAM33 family protein, low similarity to SUAPRGA1 (Emericella nidulans) GI:6562379; contains Pfam profile PF02330: Mitochondrial glycoprotein	0.9638

Sucrose	At5g03740	zinc finger (C2H2 type) family protein, contains Pfam domain, PF00096: Zinc finger, C2H2 type	0.9598
Sucrose	At5g04040	patatin-related, contains Patatin domain PF01734	0.981
Sucrose	At5g07290	RNA recognition motif (RRM)-containing protein, Mei2-like protein - Arabidopsis thaliana, EMBL:D86122	0.9538
Sucrose	At5g10010	expressed protein	0.9484
Sucrose	At5g11390	expressed protein	0.9874
Sucrose	At5g11520	aspartate aminotransferase, chloroplast / transaminase A (ASP3) (YLS4), identical to SP:P46644 Aspartate aminotransferase, chloroplast precursor (EC 2.6.1.1) (Transaminase A) {Arabidopsis thaliana}; identical to cDNA YLS4 mRNA for aspartate aminotransferase (ASP3), partial cds GI:13122285	0.9731
Sucrose	At5g14180	lipase family protein, similar to SP:Q64194 Lysosomal acid lipase/cholesteryl ester hydrolase precursor (EC 3.1.1.13) {Rattus norvegicus}; contains Pfam profile PF04083: ab-hydrolase associated lipase region	0.9511
Sucrose	At5g17380	pyruvate decarboxylase family protein, similar to 2-hydroxyphytanoyl-CoA lyase (Homo sapiens) GI:6273457; contains InterPro entry IPR000399: Pyruvate decarboxylase	0.967
Sucrose	At5g17800	myb family transcription factor (MYB56), identical to putative transcription factor (MYB56) GI:3941473 from (Arabidopsis thaliana)	0.977
Sucrose	At5g19550	aspartate aminotransferase, cytoplasmic isozyme 1 / transaminase A (ASP2), identical to SP:P46645 Aspartate aminotransferase, cytoplasmic isozyme 1 (EC 2.6.1.1) (Transaminase A) {Arabidopsis thaliana}	0.9699
Sucrose	At5g20280	sucrose-phosphate synthase, putative, similar to sucrose-phosphate synthase (EC 2.4.1.14) isoform 1 - Citrus unshiu, EMBL:AB005023	0.9497
Sucrose	At5g26090	expressed protein	0.9565
Sucrose	At5g27540	GTP-binding protein-related, low similarity to Mig-2-like GTPase Mtl (Drosophila melanogaster) GI:7271872; contains Pfam profile PF00036: EF hand	0.9817
Sucrose	At5g35910	3'-5' exonuclease domain-containing protein / helicase and RNase D C-terminal domain-containing protein / HRDC domain-containing protein, low	0.9687

		similarity to SP:Q01780 Polymyositis/scleroderma autoantigen 2 {Homo sapiens}; contains Pfam profiles PF00570: HRDC domain, PF01612: 3'-5' exonuclease	
Sucrose	At5g38890	exoribonuclease-related, similar to SP:P53859 3'-5' exoribonuclease CSL4 (EC 3.1.13.-) {Saccharomyces cerevisiae}	0.9898
Sucrose	At5g39850	40S ribosomal protein S9 (RPS9C), 40S ribosomal protein S9 - Chlamydomonas sp.,EMBL:AU066528	0.9654
Sucrose	At5g41240	glutathione S-transferase, putative	0.9608
Sucrose	At5g42150	expressed protein	0.9685
Sucrose	At5g44140	prohibitin, putative, similar to SP:P24142 Prohibitin (B-cell receptor associated protein 32) (BAP 32) {Rattus norvegicus}; contains Pfam profile PF01145: SPFH domain / Band 7 family; non-consensus TT acceptor splice site at exon 2	0.9648
Sucrose	At5g45600	YEATS family protein, contains Pfam domain PF03366: YEATS family	0.9578
Sucrose	At5g47900	expressed protein	0.9599
Sucrose	At5g49950	embryogenesis-associated protein-related, contains weak similarity to Embryogenesis-associated protein EMB8 (Swiss-Prot:Q40863) (Picea glauca)	0.9596
Sucrose	At5g52820	WD-40 repeat family protein / notchless protein, putative, similar to notchless (Xenopus laevis) GI:3687833; contains Pfam PF00400: WD domain, G-beta repeat (8 copies)	0.9869
Sucrose	At5g54300	expressed protein, contains similarity to cotton fiber expressed protein 1 (Gossypium hirsutum) gi:3264828:gb:AAC33276	0.9494
Sucrose	At5g56030	heat shock protein 81-2 (HSP81-2), nearly identical to SP:P55737 Heat shock protein 81-2 (HSP81-2) {Arabidopsis thaliana}	0.9551
Sucrose	At5g57655	xylose isomerase family protein, contains similarity to Xylose isomerase (EC 5.3.1.5) (Swiss-Prot:P22842) (Thermoanaerobacter ethanolicus)	0.9698
Sucrose	At5g58590	Ran-binding protein 1, putative / RanBP1, putative, strong similarity to Ran binding proteins from Arabidopsis thaliana atranbp1a (Arabidopsis thaliana) GI:2058282, atranbp1b (Arabidopsis thaliana) GI:2058284; contains Pfam profile PF00638: RanBP1 domain	0.9664
Sucrose	At5g61020	YT521-B-like family protein, contains	0.9534

		Pfam profile PF04146: YT521-B-like family	
Sucrose	At5g62050	OXA1 protein (OXA1), identical to AtOXA1 (Arabidopsis thaliana) GI:6624207	0.9652
Sucrose	At5g65750	2-oxoglutarate dehydrogenase E1 component, putative / oxoglutarate decarboxylase, putative / alpha-ketoglutaric dehydrogenase, putative, similar to SP:P20967 2-oxoglutarate dehydrogenase E1 component, mitochondrial precursor (EC 1.2.4.2) (Alpha-ketoglutarate dehydrogenase) {Saccharomyces cerevisiae}; contains Pfam profiles PF02779: Transketolase, pyridine binding domain, PF00676: Dehydrogenase E1 component	0.9528
Sucrose	At5g66140	20S proteasome alpha subunit D2 (PAD2) (PRS1) (PRC6), identical to SP:O24616 Proteasome subunit alpha type 7-2 (EC 3.4.25.1) (20S proteasome alpha subunit D2) {Arabidopsis thaliana}	0.9661
Sucrose	At5g67360	cucumisin-like serine protease (ARA12), Asp48; almost identical to cucumisin-like serine protease (ARA12) GI:3176874 from (Arabidopsis thaliana)	0.9549
Threonine	At1g56220	dormancy/auxin associated family protein, similar to Auxin-repressed 12.5 kDa protein (Swiss-Prot:Q05349) (Fragaria ananassa); similar to auxin-repressed protein (GI:927034) (Fragaria x ananassa); similar to dormancy-associated protein (GI:2605887) (Pisum sativum)	-0.956
Threonine	At2g23600	hydrolase, alpha/beta fold family protein, similar to ethylene-induced esterase (Citrus sinensis)	-0.9548
Threonine	At4g33000	calcineurin B-like protein 10 (CBL10), identical to calcineurin B-like protein 10 (Arabidopsis thaliana) GI:29150248	-0.9522
Threonine	At5g14260	SET domain-containing protein, low similarity to ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit N-methyltransferase I (Spinacia oleracea) GI:3403236; contains Pfam profile PF00856: SET domain	-0.9914
Threonine	At5g50760	auxin-responsive family protein, similar to auxin-induced protein TGSAUR22 (GI:10185820) (Tulipa gesneriana)	-0.979
Trigonelline	At1g03130	Encodes a protein predicted by sequence similarity with spinach PsaD to be photosystem I reaction center subunit II	-0.964
Trigonelline	At1g05010	1-aminocyclopropane-1-carboxylate	-0.9589

		oxidase	
Trigonelline	At1g55670	Encodes subunit G of photosystem I	-0.9651
Trigonelline	At1g67090	Rubisco small subunti 1A	-0.9591
Trigonelline	At1g71970	Expresed protein	-0.9531
Trigonelline	At1g75750	Gibberellin regulated protein	-0.9665
Trigonelline	At2g06520	Encodes a protein with sequence similarity to the spinach photosystem II subunit PsbX	-0.9645
Trigonelline	At3g47470	chlorophyll A-B binding protein	-0.9545
Trigonelline	At4g10340	chlorophyll A-B binding protein CP26, chloroplast	-0.9647
Trigonelline	At4g20260	DREPP plasma membrane polypeptide family protein (cold response)	-0.9695
Trigonelline	At4g32590	ferredoxin-related, contains Pfam profile: PF00111 2Fe-2S iron-sulfur cluster binding domain	-0.95
Trigonelline	At4g33010	glycine dehydrogenase (decarboxylating), putative / glycine decarboxylase, putative / glycine cleavage system P-protein	-0.9543
Trigonelline	At5g07020	proline-rich family protein	-0.9758
Trigonelline	At5g50760	auxin-responsive family protein	-0.9545
Trigonelline	At5g57040	lactoylglutathione lyase family protein / glyoxalase I family protein	-0.9581
Trigonelline	At5g65690	phosphoenolpyruvate carboxykinase (ATP), putative / PEP carboxykinase, putative / PEPCK, putative	0.9647
unkD8.0	At1g04640	biotin/lipoate A/B protein ligase family protein, similar to lipoyltransferase (LIP2p) (Arabidopsis thaliana) GI:15887052; contains Pfam profile PF03099: Biotin/lipoate A/B protein ligase family	0.9498
unkD8.0	At1g21350	expressed protein	0.9627
unkD8.0	At2g17300	expressed protein	0.9629
unkD8.0	At4g00234	expressed protein	-0.9615
unkD8.0	At4g15550	UDP-glucose:indole-3-acetate beta-D-glucosyltransferase (IAGLU), identical to UDP-glucose:indole-3-acetate beta-D-glucosyltransferase (iaglu) GI:2149126 from (Arabidopsis thaliana)	-0.9499
unkD8.0	At4g34190	stress enhanced protein 1 (SEP1), identical to stress enhanced protein 1 (SEP1) GI:7384978 from (Arabidopsis thaliana)	0.9566
unkM1.85	At1g03130	photosystem I reaction center subunit II, chloroplast, putative / photosystem I 20 kDa subunit, putative / PSI-D, putative (PSAD2), similar to SP:P12353 Photosystem I reaction center subunit II, chloroplast precursor (Photosystem I 20 kDa subunit) (PSI-D) {Spinacia oleracea}; contains Pfam profile PF02531: PsaD	-0.9564
unkM1.85	At1g05010	1-aminocyclopropane-1-carboxylate	-0.9812

		oxidase / ACC oxidase / ethylene-forming enzyme (ACO) (EAT1), Identical to 1-aminocyclopropane-1-carboxylate oxidase (ACC oxidase) gb:X66719 (EAT1). ESTs gb:T43073, gb:T5714, gb:R90435, gb:R44023, gb:AA597926, gb:AI099676, gb:AA650810 and gb:29725 come from this gene	
unkM1.85	At1g22690	gibberellin-responsive protein, putative, similar to SP:P46688 Gibberellin-regulated protein 2 precursor {Arabidopsis thaliana}; contains Pfam profile PF02704: Gibberellin regulated protein	-0.9588
unkM1.85	At1g55670	photosystem I reaction center subunit V, chloroplast, putative / PSI-G, putative (PSAG), identical to SP:Q9S7N7; similar to SP:Q00327 Photosystem I reaction center subunit V, chloroplast precursor (PSI-G) (Photosystem I 9 kDa protein) {Hordeum vulgare}; contains Pfam profile PF01241: Photosystem I psaG / psaK	-0.9705
unkM1.85	At1g67090	ribulose bisphosphate carboxylase small chain 1A / RuBisCO small subunit 1A (RBCS-1A) (ATS1A)	-0.9549
unkM1.85	At1g71970	expressed protein	-0.9746
unkM1.85	At1g75750	gibberellin-regulated protein 1 (GASA1) / gibberellin-responsive protein 1, identical to SP:P46689 Gibberellin-regulated protein 1 precursor {Arabidopsis thaliana}; supporting cDNA gi:887938:gb:U11766.1:ATU11766	-0.9729
unkM1.85	At2g23600	hydrolase, alpha/beta fold family protein, similar to ethylene-induced esterase	-0.9644
unkM1.85	At2g43010	phytochrome-interacting factor 4 (PIF4) / basic helix-loop-helix protein 9 (bHLH9) / short under red-light 2 (SRL2), identical to SP:Q8W2F3 Phytochrome-interacting factor 4 (Basic helix-loop-helix protein 9) (bHLH9) (Short under red-light 2) {Arabidopsis thaliana}; supporting cDNA gi:18026965:gb:AF251694.1:AF251694	-0.9529
unkM1.85	At3g47470	chlorophyll A-B binding protein 4, chloroplast / LHCI type III CAB-4 (CAB4), identical to SP:P27521 Chlorophyll A-B binding protein 4, chloroplast precursor (LHCI type III CAB-4) (LHCP) {Arabidopsis thaliana}	-0.9592
unkM1.85	At4g14550	auxin-responsive AUX/IAA family	-0.9588

		protein, identical to IAA14 (GI:972931) (Arabidopsis thaliana); similar to SP:Q38825 Auxin-responsive protein IAA7 (Indoleacetic acid-induced protein 7) {Arabidopsis thaliana}	
unkM1.85	At4g16210	enoyl-CoA hydratase/isomerase family protein, similar to 3-hydroxybutyryl-CoA dehydratase (Crotonase) from Clostridium acetobutylicum (SP:P52046), FadB1x (enoyl-CoA hydratase) from Pseudomonas putida (GI:13310130); contains Pfam profile PF00378 enoyl-CoA hydratase/isomerase family protein	0.9675
unkM1.85	At4g20260	DREPP plasma membrane polypeptide family protein, contains Pfam profile: PF05558 DREPP plasma membrane polypeptide	-0.9847
unkM1.85	At4g37230	oxygen-evolving enhancer protein, chloroplast, putative / 33 kDa subunit of oxygen evolving system of photosystem II, putative, similar to Oxygen-evolving enhancer protein 1, chloroplast precursor (OEE1) (33 kDa subunit of oxygen evolving system of photosystem II) (OEC 33 kDa subunit) (33 kDa thylakoid membrane protein) (SP:P14226) {Pisum sativum}	-0.9646
unkM1.85	At4g37300	expressed protein	-0.9585
unkM1.85	At5g02380	metallothionein protein 2B (MT-2B), identical to SWISS-PROT:Q38805 metallothionein-like protein 2B (MT-2B) (Arabidopsis thaliana)	-0.954
unkM1.85	At5g07020	proline-rich family protein	-0.9645
unkM1.85	At5g10410	epsin N-terminal homology (ENTH) domain-containing protein / clathrin assembly protein-related	-0.9641
unkM1.85	At5g65690	phosphoenolpyruvate carboxykinase (ATP), putative	0.9791
unkQ5.18	At4g02720	expressed protein	0.977
unkQ5.18	At4g36400	FAD linked oxidase family protein	-0.9541
unkQ7.9	At1g12910	flower pigmentation protein (AN11), contains 3 WD-40 repeats (PF00400); identical to GB:AAC18912 from (Arabidopsis thaliana) (Genes Dev. 11 (11), 1422-1434 (1997))	-0.9546
unkQ7.9	At1g76850	expressed protein	-0.9512
unkQ7.9	At2g25930	hydroxyproline-rich glycoprotein family protein, identical to cDNA nematode responsive protein GI:2213418	-0.9752
unkQ7.9	At2g28350	auxin-responsive factor (ARF10), similar to auxin response factor 10 GI:6165644 from (Arabidopsis thaliana); identical to cDNA auxin	-0.9667

		response factor 10 (ARF10) mRNA, partial cds GI:6165643	
unkS7.37	At3g15030	TCP family transcription factor, putative, similar to TCP3 GB:AAC24010 (Arabidopsis thaliana)	-0.9586
unkS7.37	At5g14260	SET domain-containing protein, low similarity to ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit N-methyltransferase I (Spinacia oleracea) GI:3403236; contains Pfam profile PF00856: SET domain	-0.9703
unkS7.37	At5g50760	auxin-responsive family protein, similar to auxin-induced protein TGSAUR22 (GI:10185820) (Tulipa gesneriana)	-0.9497
Valine	At1g12140	flavin-containing monooxygenase family protein / FMO family protein	-0.9576
Valine	At2g44410	expressed protein	-0.9499
Valine	At5g48500	expressed protein	-0.9506

References

Affymetrix. (2001). Statistical algorithms reference guide. www.affymetrix.com.

Affymetrix. (2004). GeneChip expression analysis technical manual. www.affymetrix.com.

Aharoni, A., Vorst, O. (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell **12,** 647-661.

Air, G.M. (1977). Nucleotide-Sequence of DNA of Bacteriophage-Phix174. Proceedings of the Australian Biochemical Society **10,** 60-60.

Alboresi, A., Gestin, C., Leydecker, M-T., Bedu, M., Meyer, C., Truong, H-N. (2005). Nitrate, a signal relieving seed dormancy in Arabidopsis. Plant Cell and Environment **28,** 500-512.

Ali-Rachedi, S., Bouinot, D., Wagner, M-H., Bonnet, M., Sotta, B., Grappin, P., Jullien, M. (2004). Changes in endogenous abscisic acid levels during dormancy release and maintenance of mature seeds: studies with the Cape Verde Islands ecotype, the dormant model of Arabidopsis thaliana. Planta **219,** 479-488.

Almoguera, C., Jordano, J. (1992). Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular-weight heat-shock protein and Lea mRNAs. Plant Molecular Biology. **19,** 781-792.

Alwine, J.C., Kemp, D.J., Stark, G.R. (1977). Method for Detection of Specific Rnas in Agarose Gels by Transfer to Diazobenzyloxymethyl-Paper and Hybridization with DNA Probes. Proceedings of the National Academy of Sciences of the United States of America **74,** 5350-5354.

Arabidopsis Genome Initiative. (2001). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. **408,** 796-815.

Archard, P., Vriezen, W.H., van der Straeten, D., Harderd, N.P. (2003). Ethylene regulates *Arabidopsis* development via the modulation of DELLA protein growth repressor function. Plant Cell. **15,** 2816–2825.

Archard, P., Cheng, H., De Grauwe, L., Decat, J., Schoutteten, H., Moritz, T., van der Straeten, D., Peng, J., Harberd, N.P. (2006). Integration of plant responses to environmentally activated phytohormonal signals. Science. **311,** 91-94.

Attucci, S., Carde, J.P., Raymond, P., Saint Gès, V., Spiteri, A., and Pradet, A.(1990). Oxidative phosphorylation by mitchondria extracted from dry sunflower seeds. Plant Physiology. **95,** 390-398.

Bachmair, A., Novatchkova, M., Potuschak, T., Eisenhaber, F. (2001). Ubiquitylation in plants: a post-genomic look at a post- translational modification. Trends in Plant Science, **6,** 463-470.

Bailey, T.L., Elkan, C. (2005). Unsupervised learning of multiple motifs in biopolymers using expectation maximisation. Machine Learning. **21,** 51-80.

Bao XM, Focke M, Pollard M, Ohlrogge J (2000) Understanding *in vivo* carbon precursor supply for fatty acid synthesis in leaf tissue. Plant Journal **22,** 39-50

Batak, I., Devic, M., Giba, Z., Grubisic, D., Poff, K.L., Konjevic, R. (2002). The effects of potassium nitrate and NO-donors on phytochrome A- and phytochrome B-specific induced germination of Arabidopsis thaliana seeds. Seed Science Research **12,** 253-259.

Beevers, H. (1961). Metabolic Production of Sucrose from Fat. Nature **191,** 433-&.

Belton, P.S., Colquhoun, I.J., Kemsley, E.K., Delgadillo, I., Roma, P., Dennis, M.J., Sharman, M., Holmes, E., Nicholson, J.K., Spraul, M. (1998). Application of chemometrics to the ^{1}H NMR spectra of apple juices: discrimination between apple varieties. Food Chemistry. **61,** 207–213.

Beltran-Pena, E., Ortiz-Lopez, A., Jiminez, E.S. (1995). Synthesis of Ribosomal-Proteins from Stored Messenger-Rnas Early in Seed-Germination. Plant Molecular Biology **28,** 327-336.

Bender-Machado, L., Bäuerlein, M., Carrari, F., Schauer, N., Lytovchenko, A., Gibon, Y., Kelly, A.A., Loureiro, M.E., Müller-Röber, B., Willmitzer, L. Fernie AR (2004) Expression of a yeast acetyl-CoA hydrolase in the mitochondrion of tobacco plants inhibits growth and restricts photosynthesis. Plant Mol. Biol. **55**, 645-662.

Bentsink, L., Koornneef, M. (2002). Seed dormancy and germination. In *The Arabidopsis Book* (eds Somerville, C.R., Meyerowitz, E.M.), pp 1-18. American Society of Plant Biologists.

Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miler, N., Weems, D., Rhee, S.Y. (2004). Functional annotation of the Arabidopsis genome usinf controlled vocabularies. Plant Physiology. **135,** 745-755.

Berry, T.A., and Bewley, J.D. (1992). A role for the surrounding fruit tissues in preventing germination of tomato (*Lycopersicon esclentum*) seeds. A consideration of the osmotic environment and abscisic acid. Plant Phsyiology. **100.** 951-957.

Bewley, J.D. (1997). Seed germination and dormancy. Plant Cell **9,** 1055-1066.

Bewley, J.D., and Black M. (1994). Seeds: Physiology of development and germination (New York: Plenum Press).

Bewley, J.D. and Fountain, D.W. (1972) A distinction between the actions of abscisic acid, gibberellic acid and cytokinins in a light sensitive lettuce seed. Planta. **102.** 368-371

Bewley, J.D., and Marcus, A. (1990). Gene expression in seed development and germination. Prog. Nucleic Acid Res. Mol. Biol. **38.** 165-193.

Bialek, K., Michalczuk, L. and Cohen, J.D. (1992) Auxin biosynthesis during seed germination in *Phaseolus vulgaris.* Plant Physiology. **100.** 509-517.

Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S., Sumner, L.W. (2004). Potential of metabolomics as a functional genomics tool. Trends in Plant Science **9,** 418-425.

Birboim, H.C., Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research. **7,** 1513-1523.

Bishop, G.J., and Koncz, C. (2002). Brassinosteroids and plant steroid hormone signaling. Plant Cell **14,** S97-S110.

Black, M. (1991). Involvement of ABA in the physiology of developing and maturing seeds. In abscisic acid: Physiology and Bio-chemistry, W.J. Davies and H.G. Jones, eds (Oxford, UK: Bios Scientific). pp 90-124.

Blasing, O.E., Gibon Y., Gunther, M., Hohne, M., Morcuende, R., Osuna, D., Thimm, O., Usadel, B., Scheible, W-R., Stitt, M. (2005). Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. Plant Cell **17,** 3257-3281.

Blechert, S., Brodschelm, W., Hölder, S., Kammerer, L., Kutchan, T.M., Mueller, M.J., Xia, Z-Q. and Zenk, M.H. (1995). The octadecanoic pathway: Signal molecules for the regulation of secondary pathways. Proceedings of the National Academy of Sciences of the United States of America. **1995,** 4099-4105.

Boccalandro, H.E. (2001). Ultraviolet B radiation enhances a phytochrome-B-mediated photomorphogenic response in Arabidopsis. Plant Physiology **126,** 780-788.

Bohlmann, J. (2002). Gene expression of 5-epi-aristolochene synthase and formation of capsidiol in roots of Nicotiana attenuata and N-sylvestris. Phytochemistry **60,** 109-116.

Botha, F.C., Potgieter, G.P., and Botha, A.M. (1992) Respiratory metabolism and gene expression during seed germination. Journal of Plant Growth Regulation. **11.** 211-224.

Bowyer, P., de Lucas, J.R., Turner, G. (1994). Regulation of the expression of the isocitrate lyase gene (*acuD*) of *Aspergillus nidulans.* Molecular and General Genetics MGG. **242,** 484-489.

Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., and Gorlach, J. (2001). Growth stage-based phenotypic analysis of arabidopsis: A model for high throughput functional genomics in plants. Plant Cell **13,** 1499-1510.

Breitling, R. (2006). Biological microarray interpretation: The rules of engagement. Biochimica et Biophysica acta. **1759,** 319-327.

Brown, P.O., Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nature Genetics **21,** 33-37.

Canvin, D.T., and Beevers, H. (1961). Sucrose Synthesis from Acetate in the Germinating Castor Bean: Kinetics and Pathway. The Journal of Biological Chemistry **236,** 988-995.

Casal, J.J., Luccioni, L.G., Oliverio, K.A., Boccalandro, H.E. (2003). Light, phytochrome signalling and photomorphogenesis in Arabidopsis. Photochemical & Photobiological Sciences **2,** 625-636.

Caspary, F., Hartig, A. and Schüller, H.J. (1997). Constituative and carbon source-responsive promoter elements are involved in the regulated expression of the Saccharomyces cerevisiae malate synthase gene MLS1. Molecular Genetics and Genomics. **255,** 619-627.

Chandra-Shekara, A.C., Venugopal, S.C., Barman, S.R., Kachroo, A., Kachroo, P. (2007). Plastidial fatty acid levels regulate resistance gene-dependent defense signaling in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **104,** 7277-7282.

Chang, W.M. (2000). A simultaneous assay of the expression of individual members of stress proteins under chronic and acute stimulation in human lung cancer cells, H460. Cell Biology International **24,** 939-940.

Chen, F., Bradford, K.J. (2000). Expression of an expansin is associated with endosperm weakening during tomato seed germination. Plant Physiology. **124,**

1265–1274.

Chen, F., Dahal, P., Bradford, K.J. (2001). Two Tomato Expansin Genes Show Divergent Expression and Localization in Embryos during Seed Development and Germination. Plant Physiology. **127,** 928-936.

Chen, M. (2004). Light signal transduction in higher plants. Annual Review of Genetics **38,** 87-117.

Cheng, C-L., Acedo, G.N., Cristinsin, M. and Conkling, M.A. (1992). Sucrose mimics the light induction of Arabidopsis nitrate reductase gene transcription. Proceedings of the National Academy of Sciences of the United States of America. **89,** 1861-1864.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research **31,** 3497-3500.

Chevalier, C., Le Querrec, F., Raymond, P. (1996). Sugar levels regulate the expression of ribosomal protein genes encoding protein S28 and ubiquitin-fused protein S27a in maize primary root tips. Plant Science. **117,** 95-105.

Chim, C.S., Liang, R., Tam, C.Y.Y., Kwong, Y.L. (2000). Methylation of p15 and p16 genes in acute promyelocytic leukemia: Potential diagnostic and prognostic significance. Journal of Clinical Oncology. **19**, 2033-2040.

Chourey, K., Ramani, S., Apte, S.K. (2003). Accumulation of LEA proteins in salt (NaCl) stressed young seedlings of rice (*Oryza sativa*L.) cultivar Bura Rata and their degradation during recovery from salinity stress. Plant Physiology. **160,** 1165-1174.

Chourey, P.S., Nelson, O.E. (1976). Enzymatic Deficiency Conditioned by Shrunken 1 Mutations in Maize. Biochemical Genetics **14,** 1041-1055.

Clerkx, E.J.M., Blankestijn-De Vries, H., Ruys, G.J., Groot, S.P.C. and Koorneef, M. (2003) Characterisation of green seed, an enhancer of *abi3-1* in Arabidopsis that affects seed longevity. Plant Physiology. **132** 1077-1084.

Clouse, S.D. (1997). Molecular genetic analysis of brassinosteroid action. Physiologia Plantarum. **8,** 702-709.

Clouse, S.D., Sasse, S.D. (1998). Brassinosteroids: Essential regulators of plant growth and development. Annual Review of Plant Physiology and Plant Molecular Biology **49,** 427-451.

Cohn, M.A. and Butera, D.L. (1982) Seed dormancy in red rice (*Oryza sativa*). 2.Response to cytokins. Weed Science. **30.** 200-205.

Comai, L. (1989). Coordinate Expression of Transcriptionally Regulated Isocitrate Lyase and Malate Synthase Genes in Brassica-Napus L. Plant Cell **1,** 293-300.

Cornah JE, Germain V, Ward JL, Beale MH, Smith SM (2004) Lipid utilization, gluconeogenesis, and seedling growth in Arabidopsis mutants lacking the glyoxylate cycle enzyme malate synthase. J Biol Chem **279**: 42916-42923

Coruzzi, G., and Bush, D.R. (2001). Nitrogen and carbon nutrient and metabolite signaling in plants. Plant Physiology **125,** 61-64.

Coruzzi, G.M., and Zhou, L. (2001). Carbon and nitrogen sensing and signaling in plants: emerging 'matrix effects'. Current Opinion in Plant Biology **4,** 247-253.

Cosgrove, D.J. (1997). Relaxation in a high-stress environment: the molecular bases of extensible cell walls and cell enlargement. Plant Cell. **9,** 1031-1041.

Cosgrove, D.J. (2000). Loosening of plant cell walls by expansins. Nature. **407,** 321-326.

Craigon, D.J., James, N., Okyere, J., Higgins, J., Jotham, J., May, S. (2004). NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. Nucleid Acids Research. **32,** D575-D577.

Crawford, R.M.M. (1977). Tolerance of Anoxia and Ethanol-Metabolism in Germinating Seeds. New Phytologist **79,** 511-517.

Creelman, R.A., Mullet, J.A. (1997). Oligosaccharins, brassinolides, and jasmonates: Nontraditional regulators of plant growth, development, and gene expression. Plant Cell **9,** 1211-1223.

Crocker, W., Knight, L.I., Rose, R.C. (1913) A delicate seedling test. Science **37,** 380-381.

Crowe, J.H., and Crowe, L.M. (1992). Membrane integrity in anydrobiotic organisms:Toward a mechanism for stabilizing dry seeds. In Water and Life, G.N. Somero, C.B. Osmand, and C.L. Bolis, Eds (Berlin: Springer-Verlag), pp. 87-103

Cui, X., Churchill, G.A. (2003). Statistical tests for differential expression in cDNA microarray experiments. Genome Biology. **4,** 210.

da Silva, E.A.A., Toorop, P.E., van Aelst, A.C. Hilhorst, H.W.M. (2004). Abscisic acid controls embryo growth potential and endosperm cap weakening during coffee (*Coffea arabica* cv. Rubi) seed germination. Planta. **220,** 251-261.

Dahal, P., Nevins, D.J., Bradford, K.J. (1997). Relationship of endo-β-D-mannanase activity and cell wall hydrolysis in tomato endosperm to germination rates. Plant Physiology. **113,** 1243-1252.

D'haeseleer, P. (2006) How does DNA sequence motif discovery work? Nature Biotechnology. **24,** 959-961.

Dekkers, B.J.W. (2004). Glucose delays seed germination in Arabidopsis thaliana. Planta **218,** 579-588.

Dekkers, B.J.W. (2006) Sugar signalling during germination and early seedling establishment in Arabidopsis. Ph.D. Thesis, Utrecht University.

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

Dennis, R.A., Rhodey, M., McCammon, M.T. (1999) Yeast mutants of glucose metabolism with defects in the coordinate regulation fo carbon assimilation. Arch. Biochem. Biophys. **365,** 279-288.

Derkx, M.P.M. Karssen (1993). Variability in Light Gibberellin and Nitrate Requirement of Arabidopsis-Thaliana Seeds Due to Harvest Time and Conditions of Dry Storage. Journal of Plant Physiology **141,** 574-582.

Dommes, J., Van de Walle, C. (1990). Polysome formation and incorporation of new ribosomes into polysomes during germination to the embryonic axis of maize. Plant Physiology. **79,** 289-296.

Droog, F.N.J., Hooykaas, P.J.J., Libbenga, K.R., van der Zaal, E.J. (1993). Proteins Encoded by an Auxin-Regulated Gene Family of Tobacco Share Limited but Significant Homology with Glutathione S-Transferases and One Member Indeed Shows Invitro Gst Activity. Plant Molecular Biology **21,** 965-972.

Dubreucq, B., bergerm N., Vincent, E., Boisson, M., Pelletier, G., Caboche, M.,

Lepiniec, L. (2000). The Arabidopsis AtEPR1 extensin-like gene is specifically expressed in endosperm during seed germination. Plant Journal **23,** 643-652.

Dunn, J.J., Studier, F.W. (1983). Complete Nucleotide-Sequence of Bacteriophage-T7 DNA and the Locations of T7 Genetic Elements. Journal of Molecular Biology **166,** 477-535.

Dure, L., Waters, L. (1965). Long-Lived Messenger Rna - Evidence from Cotton Seed Germination. Science **147,** 410-&.

Eastmond, P.J. (2000). The multifunctional protein AtMFP2 is co-ordinately expressed with other genes of fatty acid beta-oxidation during seed germination in Arabidopsis thaliana (L.) Heynh. Biochemical Society Transactions **28,** 95-99.

Eastmond, P.J., Hooks, M., and Graham, I.A. (2000). The Arabidopsis acyl-CoA oxidase gene family. Biochemical Society Transactions **28,** 755-757.

Eastmond, P.J., Graham, I.A. (2000). The multifunctional protein AtMFP2 is co-ordinately expressed with other genes of fatty acid β-oxidation during seed germination in *Arabidopsis thaliana* (L.) Heynh. Biochemical Society Transactions. **28,** 95-99.

Eastmond, P.J., Graham, I.A. (2001). Re-examining the role of the glyoxylate cycle in oilseeds. Trends in Plant Science **6,** 72-77.

Eastmond, P.J., and Rawsthorne, S. (2000). Coordinate changes in carbon partitioning and plastidial metabolism during the development of oilseed rape embryo. Plant Physiology **122,** 767-774.

Ebbles, T.M., Buxton, B.F. and Jones, D.T. (2006). springScape: Visualisation of microarray and contextual bioinformatics data using spring embedding and an 'information landscape'. Bioinformatics **22,** e99-e107

Eccleston, V.S., Ohlrogge, J.B. (1998). Expression of lauroyl-acyl carrier protein thioesterase in Brassica napus seeds induces pathways for both fatty acid oxidation and biosynthesis and implies a set point for triacylglycerol accumulation. Plant Cell **10,** 613-621.

Ecker, J.R. (1995). The Ethylene Signal-Transduction Pathway in Plants. Science **268,** 667-675.

Edwards, D., and Batley, J. (2004). Plant bioinformatics: from genome to phenome. Trends in Biotechnology **22,** 232-237.

Espelund, M., Jakobsen, K.S. (1992). Cloning and Direct Sequencing of Plant Promoters Using Primer-Adapter Mediated PCR on DNA Coupled to a Magnetic Solid Phase. Biotechniques. **13,** 74-81.

Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A.R., Galili, G. (2006). Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiology **142,** 839-854.

Farmer E.E. (1994). Fatty acid signalling in plants and their associated microorganisms. Plant Molecular Biology **26,** 1423-1437.

Fatland BL, Nikolau BJ, Wurtele ES (2005) Reverse genetic characterization of the cytosolic acetyl-CoA generation by ATP-citrate lyase in Arabidopsis. Plant Cell **17**, 182-203

Fenner, M., Thompson, K. (2005). The ecology of seeds. Cambridge University Press.

Fiehn, O. (2002). Metabolomics - the link between genotypes and phenotypes. Plant Molecular Biology **48,** 155-171.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., Willmitzer. (2000). Metabolite profiling for plant functional genomics. Nature Biotechnology. **18,** 1157-1161.

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., van de Voorde, A., van, Heuverswyn, H., van Herreweghe, J., Volckaert, G., Ysebaert, M. (1978). Complete Nucleotide-Sequence of Sv40 DNA. Nature **273,** 113-120.

Finch-Savage, W.E., Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. New Phytologist **171,** 501-523.

Finklestein, R.R., Gampala, S.S.L and Rock, C.D. (2002). Abscisic acid signalling in seeds and seedlings. The Plant Cell S15-S45.

Fischer-Iglesias, C. and Neuhaus, G. (2001) Zygotec embryogenesis – Hormonal control of embryo development. pp. 223-247 in Bhojwani, S.S: Soh, W.Y. (Eds) Current trends in the embryology of angiosperms. Dordrecht, Kluwer Academic.

Francis, D., Halford, N.G. (2006). Nutrient sensing in plant meristems. Plant Molecular Biology **60,** 981-993.

Frey, A., Godin, B., Bonnet, M., Sotta, B., Marion-Poll, A. (2004). Maternal synthesis of abscisic acid controls seed development and yield in Nicotiana plumbaginifolia. Planta **218,** 958-964.

Focks, N., Benning, C. (1998). *wrinkled1:* A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiology. **118,** 91-101.

Folta, K.M., Pontin, M.A., Karlin-Neumann, G., Bottini, R., Spalding, E.P. (2003). Genomic and physiological studies of early cryptochrome 1 action demonstrate roles for auxin and gibberellin in the control of hypocotyl growth by blue light. Plant Journal **36,** 203-214.

Fox, S.R., Hill, L.M., Rawsthorne, S., Hills, M.J. (2000). Inhibition of the glucose-6-phosphate transporter in oilseed rape (Brassica napus L.) plastids by acyl-CoA thioesters reduces fatty acid synthesis. Biochemical Journal **352,** 525-532.

Fu, Q. (2005). Proteomic analysis and extensive protein identification from dry, germinating Arabidopsis seeds and young seedlings. Journal of Biochemistry and Molecular Biology **38,** 650-660.

Fu, H.Y., Park, W.D. (1995). Sink-Associated and Vascular-Associated Sucrose Synthase Functions Are Encoded by Different Gene Classes in Potato. Plant Cell **7,** 1369-1385.

Fu, X.D., Harberd, N.P. (2003). Auxin promotes Arabidopsis root growth by modulating gibberellin response. Nature **421,** 740-743.

Fulda, M., Shockey, J., Werber, M., Wolter, F.P., Heinz, E. (2002). Two long-chain acyl-CoA synthetases from Arabidopsis thaliana involved in peroxisomal fatty acid beta-oxidation. Plant Journal **32,** 93-103.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., Garrels, J.I. (1999). A sampling of the yeast proteome. Molecular and Cellular Biology. **19,** 7357-7368.

Gachon, C., Mingham, A., Charrier, B. (2004). Real-time PCR: what relevance to plant studies? Journal of Experimental Botany **55,** 1445-1454.

Galau, G.A., Hughes, D.W., Dure III, L. (1986). Abscisic acid induction of cloned cotton late embryogenesis-abundant (Lea) mRNAs. Plant Molecular Biology. **7,** 155-170.

Geigenberger, P. (2003). Response of plant metabolism to too little oxygen. Current Opinion in Plant Biology **6,** 247-256.

Germain, V., Rylott, E.L., Larson, T.R., Sherson, S.M., Bechtold, N., Carde, J-P., Bryce, J.H., Graham, I.A., Smith, S.M. (2001). Requirement for 3-ketoacyl-CoA thiolase-2 in peroxisome development, fatty acid beta-oxidation and breakdown of triacylglycerol in lipid bodies of Arabidopsis seedlings. Plant Journal **28,** 1-12.

Gibson, S.I. (1994). Another player joins the complex field of sugar-regulated gene expression in plants. Proceedings of the National Academy of Sciences of the United States of America **96,** 4746-4748.

Gibson, S.I. (2005). Control of plant development and gene expression by sugar signaling. Current Opinion in Plant Biology **8,** 93-102.

Gilbert, W., Muller-Hill, B. (1966). Isolation of the lac repressor. Proceedings of the National Academy of Sciences of the United States of America. **56,** 1891-1898.

Giovanelli J, Mudd SH, Datko AH (1980) Sulfur amino acids in plants. *In* BJ Miflin, ed, The Biochemistry of Plants, Vol 5. Academic Press, New York, pp 375-402

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

Goeschl, J.D. (1966). Ethylene as a Factor Regulating Growth of Pea Epicotyls Subjected to Physical Stress. Plant Physiology **41,** 877-&.

Goldberg, R.B., de Paiva G., Yadegari, R. (1994) Plant embryogenesis: Zygote to seed. Science. **266.** 605-614.

Gonai, T., Kawahara, S., Tougou, M., Satoh, S., Hashiba, T., Hirai, N., Kawaide, H., Kamiya, Y., Yoshioka, T. (2004). Abscisic acid in the thermoinhibition of lettuce seed germination and enhancement of its catabolism by gibberellin. Journal of Experimental Botany. **55,** 111-118.

Gonzalez-Guzman, M., Abia, D., Salinas, J., Serrano, R., Rodrigues, P.L. (2004) Two new alleles of the abscisic aldehyde oxidase 3 gene reveal its role in abscisic acid biosynthesis in seeds. Plant Physiology. **135.** 325-333.

Gonzali, S., Loreti, E., Solanelli, C., Novi, G., Alpi, A., Perata, P. (2006). Identification of sugar-modulated genes and evidence for in vivo sugar sensing in Arabidopsis. Journal of Plant Research **119,** 115-123.

Graham, I.A., Leaver, C.J., Smith, S.M. (1992). Induction of Malate Synthase Gene-Expression in Senescent and Detached Organs of Cucumber. Plant Cell **4,** 349-357.

Greenstein, B., Greenstein, A. (1996). Medical biochemistry at a glance. Blackwell Science Ltd. pp 42-43

Grierson, C., Du, J.S., de Torres Zabala, M., Beggs, K., Smith, C., Holdsworth, M. Bevan, M. (1994). Separate cis sequences and trans factors direct metabolic and developmental regulation of a potato tuber storage protein gene. The Plant Journal. **5,** 815-26.

Grzelczak, Z.F., Sattolo, M.H., Hanley-Bowdoin, L.K., Kennedy, T.D., Lane, B.G. (1982). Synthesis and turnover of proteins and mRNA in germinating wheat embryos. Canadian Journal of Biochemistry. **60,** 389-397.

Gygi, S.P., Rochon, Y., Franza, B.R., Aebersol, R. (1999). Correlation between

protein and mRNA abundance in yeast. Molecular and Cellular Biology. **19,** 1720, 1730.

Hadacek, F. (2002). Secondary metabolites as plant traits: Current assessment and future perspectives. Critical Reviews in Plant Sciences **21,** 273-322.

Halford, N.G., and Paul, M.J. (2003). Carbon metabolite sensing and signalling. Plant Biotechnology Journal **1,** 381-398.

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L.W., and Bino, R. (2002). Plant Metabolomics: The Missing Link in Functional Genomics Strategies. The Plant Cell **14,** 1437 - 1440.

Han, B., Hughes, D.W., Galau, G.A., Bewley, J.D., Kermode, A.R. (1997). Changes in late-embryogenesis-abundant (LEA) messenger RNAs and dehydrins during maturation and premature drying of Ricinus communis L seeds. Planta **201,** 27-35.

Hanfrey C, Sommer S, Mayer MJ, Burtin D, Michael AJ (2001) Arabidopsis polyamine biosynthesis: absence of ornithine decarboxylase and the mechanism of arginine decarboxylase activity. Plant Journal **27,** 551-560

Hare, P.D. and Cress (1997). Metabolic implications of stress-induced proline accumulation in plants. Plant Growth Regulation **21,** 79-102.

Hare, P.D. (2003). A regulatory role for proline metabolism in stimulating Arabidopsis thaliana seed germination. Plant Growth Regulation **39,** 41-50.

Hartig, A., Simon, M.M., Schuster, T., Daugherty, J.R., Yoo, H.S. and Cooper, T.G. (1992).Differentially regulated malate synthase genes participate in carbon and nitrogen metabolism of S. cerevisiae. Nucleic Acids Research. **20,** 5677-5686.

Hays, D.B., Yeung, E.C., and Pharis, R.P. (2002) The role of gibberellins in embryo axis development. Journal of Experimental Botany. **53.** 1747-1751

Heath, J.D. (1986). Analysis of Storage Proteins in Normal and Aborted Seeds from Embryo-Lethal Mutants of Arabidopsis-Thaliana. Planta **169,** 304-312.

Heazlewood, J.L., and Millar, A.H. (2003). Integrated plant proteomics - putting the green genomes to work. Functional Plant Biology **30,** 471-482.

Hepher, A., Roberts, J.A. (1985). The Control of Seed-Germination in Trollius-Ledebouri - the Breaking of Dormancy. Planta **166,** 314-320.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research **27,** 297-300.

Higuchi, R., Dollinger, G., Walsh, P.S., Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. Nature Biotechnology. **10,** 413-417.

Higo, K., Ugawa, Y., Iwamoto, Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research. **27,** 297-300.

Hilhorst, H.M.W. (1995). A Critical update on seed dormancy. I. Primary dormancy. Seed Science Research. **5.** 61-73.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., and Saito, K. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the

United States of America **101,** 10205-10210.

Ho, S.L., Chao, Y.C., Tong, W.F., and Yu, S.M. (2001). Sugar coordinately and differentially regulates growth- and stress-related gene expression via a complex signal transduction network and multiple control mechanisms. Plant Physiology **125,** 877-890.

Hobbie, L.J. (2006). Auxin and cell polarity: the emergence of AXR4. Trends in Plant Science **11,** 517-518.

Hobbie, L.J. (2007) Auxin. *In*: Encyclopedia of life sciences. John Wiley and Sons, Ltd.

Holsbeeks, I., Lagatie, O., van Nuland, A., van de Velde, S., Thevelein, J.M. (2004). The eukaryotic plasma membrane as a nutrient-sensing device. Trends in Biochemical Sciences **29,** 556-564.

Hooks, M.A., Turner, J.E., Murphy, E.C., Graham, I.A. (2004). Acetate non-utilizing mutants of Arabidopsis: evidence that organic acids influence carbohydrate perception in germinating seedlings. Molecular Genetics and Genomics **271,** 249-256.

Hooks, M.A., Turner, J.E., Murphy, E.C., Johnston, K.A., Burr, S., Jaroslawski, S. (2007). The Arabidopsis ALDP protein homologue COMATOSE is instrumental in peroxisomal acetate metabolism. Biochemical Journal. **406,** 399-406.

Horii, A. (2007). Enhancement of seed vigour following insecticide and phenolic elicitor treatment. Bioresource Technology **98,** 623-632.

Howell, S.H. (1998). Molecular genetics of plant development. Cambridge University Press.

Hu, J.F., Pham, J., Dey, I., Li, T., Vu, T.H., Hoffman, A.R. (2000). Allele-specific histone acetylation accompanies genomic imprinting of the insulin-like growth factor II receptor gene. Endocrinology. **141**, 4428-4435.

Huber, S.C., Huber, J.L. (1996). Role and Regulation of sucrose-phosphate synthase in higher plants. Annual Review of Plant Physiology and Plant Molecular Biology. **47.** 431-444

Huq, E., Al-Sady, B., Hudson, M., Kim, C., Apel, K., Quail, P.H. (2004). PHYTOCHROM-INTERACTING FACTOR 1 is a critical bHLH regulator of chlorophyll biosynthesis. Science. **305,** 1937-1941.

Hutton, D., Stumpf, P.K. (1969). Fat Metabolism in Higher Plants. XXXVII. Characterization of the beta-Oxidation Systems From Maturing and Germinating Castor Bean Seeds. Plant Physiology. **44,** 508-516.

Ideker, T., Galitski, T., Hood, L. (2001a). A new approach to decoding life: Systems biology. Annual Review of Genomics and Human Genetics **2,** 343-372.

Ideker, T., Thorsson, V., Ranish, J., Christmas, R., Buhler, J., Bumgarner, R., Aebersold, R., Hood, L. (2001b). Integrated genomic and proteomic analysis of a systematically perturned metabolic network. Science. **292,** 929-934.

Innis, M.A., Gelfand, D.H. (1990). Optimization of PCRs. In: Innis MA , Gelfand DH , Sninsky JJ , White TJ , editors. PCR protocols: a guide to methods and applications. San Diego: Academic Press. p 3-12.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. Nature **21**, 931-945

Ishibashi, N., Yamauchi, D., Minamikawa, T. (1990). Stored Messenger-Rna in

Cotyledons of Vigna-Unguiculata Seeds - Nucleotide-Sequence of Cloned Cdna for a Stored Messenger-Rna and Induction of Its Synthesis by Precocious Germination. Plant Molecular Biology **15,** 59-64.

Jacob, F., Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology. **3,** 318-356.

Jiao, Y.L., Ma, L., Strickland, E., Deng, X.W. (2005). Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and Arabidopsis. Plant Cell **17,** 3239-3256.

Jiao, Y.L., Lau, O.S., Deng, X.W. (2007). Light-regulated transcriptional networks in higher plants. Nature Reviews Genetics **8,** 217-230.

Jung, S.H., Lee, J-Y., Lee, D-H. (2003). Use of SAGE technology to reveal changes in gene expression in Arabidopsis leaves undergoing cold stress. Plant Molecular Biology **52,** 553-567.

Karssen, C.M. (1976). Uptake and Effect of Abscisic-Acid During Induction and Progress of Radicle Growth in Seeds of Chenopodium-Album. Physiologia Plantarum **36,** 259-263.

Karssen, C.M., Brinkhorst-van der Swan, D.L.C., Breekland, A.E., Koornneef, M. (1983). Induction of Dormancy During Seed Development by Endogenous Abscisic-Acid - Studies on Abscisic-Acid Deficient Genotypes of Arabidopsis-Thaliana (L) Heynh. Planta **157,** 158-165.

Karssen, C.M. (1995). Hormonal regulation of seed development, dormancy, and germination studied by genetic control. In Seed Development and Germination, J. Kigel and G. Galili, eds (New York: Marcel Dekker), pp. 333-350

Katavic, V., Reed, D.W., Taylor, D.C., Giblin, E.M., Barton, D.L., Zou, J., Mackenzie, S.L., Covello, P.S., Kunst, L. (1995). Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in *Arabdopsis thaliana* affecting diacylglycerol acyltransferase activity. Plant Physiology. **108,** 399-409.

Kausch, A.P., Owen, T.P., Zachwieja, S.J., Flynn, A.R., Sheen, J. (2001). Mesophyll-specific, light and metabolic regulation of the C(4)PPCZm1 promoter in transgenic maize. Plant Molecular Biology **45,** 1-15.

Ke J, Behal RH, Back SL, Nikolau BJ, Wurtele ES, Oliver DJ (2000) The role of pyruvate dehydrogenase and acetyl-Coenzyme A synthetase in fatty acid synthesis in developing arabidopsis seeds. Plant Physiology. 123, 497-508

Kermode, A.R. (1990). Regulatory mechanisms involved in the transition from seed development to germination. Critical Reviews in Plant Sciences. **9,** 155-195.

Kerr, M.K., Martin, M., and Churchill, G.A. (2000). Analysis of variance for gene expression microarray data. Journal of Computational Biology **7,** 819-837.

Kerr, M.K., and Churchill, G.A. (2001). Bootstrapping cluster analysis: Assessing the reliability of conclusions from microarray experiments. Proceedings of the National Academy of Sciences of the United States of America **98,** 8961-8965.

Kim, B.C., Tennessen, D.J., Last, R.L. (1998). UV-B-induced photomorphogenesis in Arabidopsis thaliana. Plant Journal **15,** 667-674.

Kim, J.A., Yun, J., Lee, M.S., Kim, Y.S., Woo, J.C., Park, C.M. (2005). A basic helix-loop-helix transcription factor regulates cell elongation and seed germination. Molecules and Cells **19,** 334-341.

King, J.E., Gifford, D.J. (1997). Amino Acid Utilization in Seeds of Loblolly Pine during Germination and Early Seedling Growth (I. Arginine and Arginase Activity). Plant Physiology. **113,** 1125-1135.

Kitano, H. (2002) Systems biology: A brief overview. Science. **295,** 1662-1663.

Koch, K.E. (1996). Carbohydrate-modulated gene expression in plants. Annual Review of Plant Physiology and Plant Molecular Biology **47,** 509-540.

Kolb, A., Busby, S., Buc, I.I., Garges, S., Adhya, S. (1993). Transcriptional Regulation by Camp and Its Receptor Protein. Annual Review of Biochemistry. **62,** 749-795.

Koornneef, M., Bentsink, Hilhorst, H. (2002). Seed dormancy and germination. Current Opinion in Plant Biology **5,** 33-36.

Koornneef, M., Reuling, G., Karssen, C.M. (1984). The isolation and characterisation of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. Physiologia Plantarum. **61,** 377-383.

Koornneef, M., and Karssen, C.M. (1994). Seed dormancy and germination. In *Arabidopsis,* E.M. Meyerowitz and C.R. Somerville. Eds (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press). pp 313-334.

Krock, B., Schmidt, S., Hertwick, C., Baldwin, I.T. (2002). Vegetation-derived abscisic acid and four terpenes enforce dormancy in seeds of the post-fire annual, *Nicotiana attenuate*. Seed Science Research. **12,** 239-252.

Kucera, B., Cohn, M.A., Leubner-Metzger, G. (2005). Plant hormone interactions during seed dormancy release and germination. Seed Science Research **15,** 281-307.

Kunkel, B.N., and Brooks, D.M. (2002). Cross talk between signaling pathways in pathogen defense. Current Opinion in Plant Biology **5,** 325-331.

Kushiro, T., Okamoto, M., Nakabayshi, K., Yamagishi, K., Kitamura, S., Asami, T. Hirai, N., Koshiba, T., Kamiya, Y. and Nmabara, E. (2004). The Arabidopsis cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: Key enzymes in ABA catabolism. EMBO Journal. **23.** 1647-1656.

Laby, R.J., Kincaid, S., Kim, D. and Gibson, S.I. (2000) The Arabidopsis sugar-insensitive mutants sis4 and sis5 are defective in abscisic acid synthesis and response. The Plant Journal **23,** 587-596.

Lange, G.A. (1996). Plant dormancy. (Oxford, UK: CAB International).

Lange, T. (1998). Molecular biology of gibberellin synthesis. Planta. **204,** 409-419.

Lange, T., Graebe, J.E. (1993). Enzymes of gibberellin synthesis. Methods in Plant Biochemistry. **9,** 403-430.

Laux, T., Jurgens, G. (1997). Embryogenesis: A new start in life. Plant Cell **9,** 989-1000.

Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., Wootton, J.C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science. **262,** 208-214.

Lee, Y., Choi, D., Kende, H. (2001). Expansins: ever-expanding numbers and functions. Current Opinion in Plant Biology. **4,** 527-532.

Leon, P., and Sheen, J. (2003). Sugar and hormone connections. Trends in Plant Science **8,** 110-116.

Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze,

P., and Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research **30,** 325-327.

Leubner-Metzger, G. (2001). Brassinosteroids and gibberellins promote tobacco seed germination by distinct pathways. Planta. **213,** 758-763.

Leubner-Metzger, G. (2002). Seed after-ripening and over-expression of class I ß-1,3-glucanase confer maternal effects on tobacco testa rupture and dormancy release. Planta. **215,** 959-968.

Leubner-Metzger, G. (2003). Functions and regulation of beta-1,3-glucanases during seed germination, dormancy release and after-ripening. Seed Science Research **13,** 17-34.

Leubner-Metzger, G., Frűndt, C., Vőgeli-Lange, R., Meins, F., Jr. (1995). Class I β-1, 3- glucanases in the endosperm of tobacco during germination. Plant Physiology. **109,** 751-759.

Lewin, B., (2000). Genes VII. Oxford University Press.

Li, J., Nagpal, P., Vitart, V., McMorris, T.C., Chory, J. (1996). A role for brassinosteroids in tight-dependent development of Arabidopsis. Science. **272,** 398-401.

Li L., Thipyapong P., Breeden D.C., Steffens, J.C. (2003). Overexpression of a bacterial branched-chain-keto acid dehydrogenase complex in Arabidopsis results in accumulation of branched-chain acyl-CoAs and alteration of free amino acid composition in seeds. Plant Science. **165,** 1213-1219

Liang, P., Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science. **14,** 967-971

Lindsey, K. (1993). Embryogenesis - a Question of Pattern. Journal of Experimental Botany **44,** 359-374.

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nature Genetics. **21,** 20-24.

Liu, B., Joshi, H.C., Wilson, T.J., Silflow, C.D., Palevitz, B.A., Snustad, D.P. (1994). gamma-tubulin in Arabidopsis: gene sequence, immunoblot and immunofluorescencs studies. Plant Cell. **6,** 303-314.

Liu, P., Koizuka, N., Martin, R.C., Nonogaki, H. (2005). The BME3 (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. The Plant Journal **44,** 960-971.

Ljung, K., Ostin, A., Lioussanne, L. Sandberg, G (2001). Developmental regulation of indole-3-acetic acid turnover in scots pine seedlings. Plant Physiology. **125.** 464-475

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

Lopez-Molina, L., Chua, N.H. (2000). A null mutation in a bZIP factor confers ABA-insensitivity in Arabidopsis thaliana. Plant and Cell Physiology **41,** 541-547.

Love, A.J., Yun, B.W., Laval, V., Loake, G.J., Milner, J.J. (2005). *Cauliflower mosaic virus*, a compatible pathogen of Arabidopsis, engages three distinct defense-signaling pathways and activates rapid systemic generation of reactive

oxygen species. Plant Physiology. **139,** 935-948.

Lu, C., Hills, M.J. (2002). Arabidopsis mutants deficient in diacylglycerol acyltransferase display increased sensitivity to abscisic acid, sugars and osmotic stress during germination and seedling development. Plant Physiology. **129,** 1-7.

Luo, J., Nishiyama, Y., Fuell, C., Taguchi, G., Elliott, K., Hill, L., Tanaka, Y., Kitayama, M., Yamazaki, M., Bailey, P., Parr, A., Michael, A.J., Saito, K., Martin, C. (2007). Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from Arabidopsis thaliana. The Plant Journal. **50,** 678-695.

Madison, J.T., Thompson, J.F., Muenster, A.E. (1981). Turnover of storage protein in seeds of soya bean and pea. Annals of Botany. **47,** 65-73.

Maeting, I., Schmidtab, G., Sahma, H., Revueltac, J.L., Stierhofd, Y-D., Stahmanna, K-P. (1999). Isocitrate lyase of *Ashbya gossypii* – transcriptional regulation and peroxisomal localization. FEBS Letters. **444,** 15-21

Mahalingham, R., Gomez-Buitrago, A., Eckardt, N., Shah, N., Guevara-Garcia, A., Day, P., Raina, R., Fedoroff, N.V. (2003). Characterizing the stress/defence transcriptome of Arabidopsis. Genome Biology. **4,** R20.

Mansfield, S.G. Briarty (1996). The dynamics of seedling and cotyledon cell development in Arabidopsis thaliana during reserve mobilization. International Journal of Plant Sciences **157,** 280-295.

Martínez-Villaluenga C, Kuo Y-H, Lambein F,· Juana Frías, Vidal-Valverde C (2006) Kinetics of free protein amino acids, free non-protein amino acids and trigonelline in soybean (Glycine max L.) and lupin (Lupinus angustifolius L.) sprouts. European Food Research Technology. **224**, 177-186

Matsumura, H., Nirasawa, S., Terauchi, R. (1999). Transcript profiling in rice (Oryza sativa L.) seedlings using serial analysis of gene expression (SAGE). The Plant Journal. **20,** 719-726.

McCammon, M.T. (1996). Mutants of Saccharomyces cerevisiae with defects in acetate metabolism: Isolation and characterization of Acn(-) mutants. Genetics **144,** 57-69.

McCarty, D.R., Hattori, T., Carson, C.B., Vasil, V., Lasar, M., Vasil, I.K. (1991). The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. Cell. **66,** 895-905.

McCulloch, R.K., Choong, C.S., Hurley, D.M. (1995). An evaluation of competitor type and size for use in the determination of mRNA by competitive PCR. PCR Methods Applications. **4,** 219-226.

McKee, T., McKee, J.R. (2003). Biochemistry: The molecular basis of life. (3 ed.). McGraw-Hill Higher Education, International Edition.

Meinke, D.W. (1991). Embryonic Mutants of Arabidopsis-Thaliana. Developmental Genetics **12,** 382-392.

Mercke, P., Kappers, I.F., Verstappen, F.W.A., Vorst, O., Dicke, M., and Bouwmeester, H.J. (2004). Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants. Plant Physiology **135,** 2012-2024.

Mettler, I.J. and Beevers, H. (1980) Oxidation of NADH in glyoxysomes by a malate-aspartate shuttle. Plant Physiology **66,** 555-560.

Millar, D.P. (1996). Fluorescence studies of DNA and RNA structure and dynamics. Current Opinion in Structural Biology **6,** 322-326.

Moing, A., Maucourt, M., Renaud, C., Gaudillere, M., Brouquisse, R., Lebouteiller, B., Gousset-Dupont, A., Vidal, J., Granot, D., Denoyes-Rothan, B., Lerceteau-Kohler, E., and Rolin, D. (2004). Quantitative metabolic pro. ling by 1-dimensional H-1-NMR analyses: application to plant genetics and functional genomics. Functional Plant Biology **31,** 889-902.

Mok, D.W.S and Mok, M.C. (2001) Cytokinin metabolism and action. Annual Review of Plant Physiology and Plant Molecular Biology. **52.** 89-118.

Monroy, A.F., Schwartzbach, S.D. (1984). Catabolite repression of chloroplast development in Euglena. Proceedings of the National Academy of Sciences of the United States of America. **81,** 2786-2790.

Morohashi, Y. (1986). Patterns of mitchocondrial development in reserve tissue of germinated seeds: A Survey. Plant Physiology. **66,** 653-658.

Morohashi, Y. and Bewley, J.D. (1980). Development of mitchocondrial activities in pea cotyedons during and following germination of the axis. Plant Physiology. **66**, 70-73.

Müller, K., Tintelnot, S., Leubner-Metzger, G. (2006). Endosperm-limited Brassicaceae seed germination: abscisic acid inhibits embryo-induced endosperm weakening of *Lepidium sativum* (cress) and endosperm rupture of cress and *Arabidopsis thaliana*. Plant and Cell Physiology. **47,** 864-877.

Müller, R., Morant, M., Jarmer, H., Nilsson, L., Nielson, T.H. (2007). Genome-wide analysis of the Arabidopsis leaf transcriptome reveals interaction of phosphate and sugar metabolism. Plant Physiology. **143,** 156-171.

Mütz, K. (1998). Deposition of storage proteins. Plant Molecular Biology. **38,** 77-99.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassay with tobacco tissue cultures. Physiologia Plantarum **15,** 473-496.

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., Nambara, E. (2005). Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. Plant Journal **41,** 697-709.

Nambara, E., and Marion-Poll, A. (2003) ABA action and interactions in seeds. Trends in Plant Science. **8,** 213-217

Naot, D., Ben-Hayyim, G., Eshdat, Y., Holland, D. (1995). Drought, heat and salt stress induce the expression of a citrus homologue of an atypical late-embryogenesis Lea5 gene. Plant Molecular Biology. **3,** 619-622.

NDong, C., Danyluk, J., Wilson, K.E., Pocock, T., Huner, N.P.A. and Sarhan, F. (2002) Cold-regulated cereal chloroplast late embryogenesis abundant-like protein. Molecular characterization and functional analyses. Plant Physiology **129,** 1368–1381

Nemhauser, J.L., Mockler, T.C., Chory, J. (2004). Interdependency of brassinosteroid and auxin signaling in Arabidopsis. Plos Biology **2,** 1460-1471.

Nemhauser, J.L., Hong, F., Chory, J. (2006). Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. Cell **126,** 467-475.

Noteborn, H.P.J.M., Lommen, A., van der Jagt, R.C., Weseman, J.M. (2000).

Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. Journal of Biotechnology. **77,** 103–114.

Oehler, S. (1990). The 3 Operators of the Lac Operon Cooperate in Repression. Embo Journal **9,** 973-979.

Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y. and Yamaguchi, S. (2003) Gibberellin biosynthesis and response during Arabidopsis seed germination. Plant Cell. **15.** 1591-1604.

Oh, E. (2004). PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in Arabidopsis thaliana. Plant Cell **16,** 3045-3058.

Oh, E., Yamaguchi, S., Kamiya, Y., Bae, G., Chung, W-I., Choi, G. (2006). Light activates the degradation of PIL5 protein to promote seed germination through gibberellin in Arabidopsis. The Plant Journal. **47,** 124-139.

Oh, E., Yamguchi, S., Hu, J., Yusuke, J., Jung, B., Paik, I., Lee, H-S., Sun, T-P., Kamiya, Y., Choi, G. (2007). PIL5, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in Arabidopsis seeds. Plant Cell **19,** 1192-1208.

Oh, M.K., Rohlin, L., Kao, K.C., Liao, J.C. (2002). Global expression profiling of acetate-grown Escherichia coli. Journal of Biological Chemistry **277,** 13175-13183.

Oksman-Caldentey, K.M., Saito, K. (2005). Integrating genomics and metabolomics for engineering plant metabolic pathways. Current Opinion in Biotechnology **16,** 174-179.

Oliver, M.J., Dowd, S.E., Zaragoza, J., Mauget, S.A. and Payton, P.R. (2004) The rehydration transcriptome of the desiccation-tolerant bryophyte Tortula ruralis: transcript classification and analysis. BMC Genomics **5** 89

Pastori, G.M., Foyer, C.H. (2002). Common components, networks, and pathways of cross-tolerance to stress. The central role of "redox" and abscisic acid-mediated controls. Plant Physiology. **129,** 460-468.

Pastuglia, M., Azimzadeh, J., Goussot, M., Camilleri, C., Belcram, K., Evrard, J-L., Schmit, A-C., Guerche, P., Bouchez, D. (2006). gamma-tubulin is essential for microtubule organization and development in Arabidopsis. Plant Cell. **18,** 1412-1425.

Penfield, S., Graham, S., and Graham, I.A. (2005). Storage reserve mobilization in germinating oilseeds: Arabidopsis as a model system. Biochemical Society Transactions **33,** 380-383.

Penfield, S., Josse, E.M., Kannangara, R., Gilday, A.D., Halliday, K.J., and Graham, I.A. (2003). Cold and light control seed germination through the bHLH transcription factor SPATULA. Current Biology **15,** 1998-2006.

Penfield, S., Pinfield-Wells, H.M., and Graham, I.A. (2006). Storage reserve mobilisation and seedling establishment in Arabidopsis. The Arabidopsis Book, American Society of Plant Biologists.

Penfield, S., Rylott, E.L., Gilday, A.D., Graham, S., Larson, T.R., and Graham, I.A. (2004). Reserve mobilization in the Arabidopsis endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. Plant Cell **16,** 2705-2718

Perneger, T.V. (1998). What's wrong with Bonferroni adjustments. British Medical Journal. **316.** 1236-1238.

Petruzzelli, L., Muller, L., Hermann, K., Leubner-Metzger, G. (2003). Distinct expression patterns of beta-1,3-glucanases and chitinases during the germination of Solanaceous seeds. Seed Science Research **13,** 139-153.

Postma, P.W., Lengeler, J.W., Jacobson, G.R. (1993). Phosphoenolpyruvate - Carbohydrate Phosphotransferase Systems of Bacteria. Microbiological Reviews **57,** 543-594.

Pracharoenwattana I, Cornah JE, Smith SM (2005). Arabidopsis peroxisomal citrate synthase
is required for fatty acid respiration and seed germination. Plant Cell **17,** 2037-2048

Pracharoenwattana I, Cornah JE, Smith SM (2007) Arabidopsis peroxisomal malate dehydrogenase functions in β-oxidation but not in the glyoxylate cycle. Plant Journal. **50,** 381-390

Price, J., Laxmi, A., St. Martin, S.K., Jang, J-C. (2004). Global transcription profiling reveals multiple sugar signal transduction mechanisms in Arabidopsis. Plant Cell **16,** 2128-2150.

Pritchard, S.L., Charlton, W.L., Baker, A., Graham, I.A. (2002). Germination and storage reserve mobilization are regulated independently in Arabidopsis. Plant Journal **31,** 639-647.

Raghavan, V. (2000). Developmental biology of flowering plants. Developmental biology of flowering plants, i-xxii, 1-354.

Rausch T., Wachter A. (2005) Sulfur metabolism: a versatile platform for launching defence operations. Trends in Plant Science. **10**: 503-509

Raz, V. (2001). Sequential steps for developmental arrest in Arabidopsis seeds. Development **128,** 243-252.

Reddy, A.S.N. (2001) Nuclear pre-mRNA splicing in plants. *Critical Reviews in Plant* Sciences. **20**, 523-571.

Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan. J., Ghosh, P.K., Celma, M.L., Weissman, S.M. (1978). Genome of Simian-Virus 40. Science **200,** 494-502.

Reymond, P., Farmer, E.E. (1998). Jasmonate and salicylate as global signals for defense gene expression. Current Opinion in Plant Biology. **1,** 404-411.

Rockwell, N.C., Su, Y.S., Lagarias, J.C. (2006). Phytochrome structure and signaling mechanisms. Annual Review of Plant Biology **57,** 837-858.

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 Journal **23,** 131-142.

Rolland, F., Moore, B., Sheen, J. (2002). Sugar sensing and signaling in plants. Plant Cell **14,** S185-S205.

Rombauts, S., Dehais, P., Van Montagu, M., and Rouze, P. (1999). PlantCARE, a plant cis-acting regulatory element database. Nucleic Acids Research **27,** 295-296.

Rombauts, S., Florquin, K., Lescot, M., Marchal, K., Rouze, P., and Van de Peer, Y. (2003). Computational approaches to identify promoters and cis-regulatory

elements in plant genomes. Plant Physiology **132,** 1162-1176.

Rook, F., Bevan, M.W. (2003). Genetic approaches to understanding sugar-response pathways. Journal of Experimental Botany. **54,** 495-501.

Rousselin, P., Kraepiel, Y., Maldiney, R., Miginiac, E. and Caboche, M. (1992) Characterization of three hormone mutants of *nicotiana plumnagnifolia*: evidence for a common ABA deficiency. Theoretical and Applied Genetics. **85.** 213-221.

Rose, M., Albig, W. and Entian, K.D. (1991). Glucose repression in Saccharomyces cerevisiae is directly associated with hexose phosphorylation by hexokinase PI and PIII. European Journal of Biochemistry. **199,** 511-518.

Roth, S., Droge, W. (1991). Regulation of interleukin 2 production, interleukin 2 mRNA expression and intracellular glutathione levels in ex vivo derived T lymphocytes by lactate. European Journal of Immunology. **21,** 1933-1937.

Ruas, M., Peters, G. (1998) The p16(INK4a)/CDKN2A tumor suppressor and its relatives. Biochimica Et Biophysica Acta-Reviews on Cancer. **1378**, F115-F177.

Rylott, E.L., Hooks, M.A., and Graham, I.A. (2001). Co-ordinate regulation of genes involved in storage lipid mobilization in Arabidopsis thaliana. Biochemical Society Transactions **29,** 283-287.

Saiki, R.K., Walsh, P.S., Levenson, C.H., Erlich, H.A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proceedings of the National Academy of Sciences of the United States of America. **86,** 6230-6234.

Sakai, H., Aoyama, T., Oka, A. (2000). Arabidopsis ARR1 and ARR2 response regulators operate as transcriptional activators. The Plant Journal. **6,** 703-711.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: a Laboratory Manual. (Cold Spring Harbor Laboratory Press).

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., Petersen, G.B. (1982). Nucleotide-Sequence of Bacteriophage-Gamma DNA. Journal of Molecular Biology **162,** 729-773.

Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America. **74,** 5463-5467.

Scheible, W.R., Lauerer, M., Schulze E-D., Caboche, M., Stitt, M. (1997). Accumulation of nitrate in the shoot acts as a signal to regulate shoot-root allocation in tobacco. Plant Journal **11,** 671-691.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995). Quantitative Monitoring of Gene-Expression Patterns with a Complementary-DNA Microarray. Science **270,** 467-470.

Schena, M. (1996). Genome analysis with gene expression microarrays. Bioessays **18,** 427-431.

Schmid., S.A., Gaumann, A., Wondrak, M., Eckermann, C., Schulte, S., Mueller-Klieser, W., Wheatley, D.N. and Kunz-Schughart, L.A. (2007). Lactate adversely affects the in vitro formation of endothelial cell tubular structures through the action of TGF-β1. Experimental Cell Research. **313,** 2531-2549.

Schöler, A., Schüller, H.J. (1993). Structure and Regulation of the Isocitrate Lyase Gene Icl1 from the Yeast Saccharomyces-Cerevisiae. Current Genetics **23,** 375-381.

Schopfer, P., and Placy, C. (1985). Control of seed germination by abscisic acid. III. Effect of embryo growth potential (minimum turgor pressure) and growth coefficient (cell wall extensibility) in *brassica napus* L. Plant Physiology. **77,** 676-686.

Schwechheimer, C. and Bevan, M. (1998). The regulation of transcription factor activity in plants. Trends in Plant Science. **3**, 378-383.

Sheen, J. (1990). Metabolic regulation of transcription in higher plants. Plant Cell **2,** 1027-1038.

Sheen, J. (1994). Feedback control of gene expression. Photosynthetic Research. **39,** 427-438.

Shevchenko, A., Wilm., M., Vorm, O., Mann, M. (1996). Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Analytical Chemistry **68,** 850-858.

Smeekens, S. (2000). Sugar-induced signal transduction in plants. Annual Review of Plant Physiology and Plant Molecular Biology **51,** 49-81.

Smeekens, S., and Rook, F. (1997). Sugar sensing and sugar-mediated signal transduction in plants. Plant Physiology **115,** 7-13.

Soderman, E.M., Brocard, I.M., Lynch, T.J. and Finklestein, R.R. (2000) Regulation and function o the Arabidopsis ABA-insensitive4 gene in seed and abscisic acid response signalling networks. Plant Physiology. **124,** 1752-1765.

Solfanelli, C., Poggi, A., Loreti, E., Alpi, A., Perata, P. (2006). Sucrose-specific induction of the anthocyanin biosynthetic pathway in Arabidopsis. Plant Physiology. **140,** 637-646.

Sowokinos, J.R., Varns, J.L. (1992). Induction of sucrose synthase in potato tissue culture: effect of carbon source and metabolic regulators on sink strength. Journal of Plant Physiology. **139,** 672-679.

Speed, T.P. (ed.) (2003). Statistical analysis of gene expression microarray data. Chapman & Hall/CRC, Boca Raton.

Steadman, K.J. (2004). Dormancy release during hydrated storage in Lolium rigidum seeds is dependent on temperature, light quality, and hydration status. Journal of Experimental Botany **55,** 929-937.

Steber, C.M., and McCourt, P. (2001) A role for brassinosteroids in germination in Arabidopsis. Plant Physiology. **125.** 763-769.

Stekel, (2003). Microarray bioinformatics. Oxford University Press.

Sterk, P., Booji, H., Schellekens, G.A., van Kammen, A., de Vries, S.C. (1991). Cell-Specific Expression of the Carrot Ep2 Lipid Transfer Protein Gene. Plant Cell **3,** 907-921.

Sullivan, J.A. (2003). From seed to seed: the role of photoreceptors in Arabidopsis development. Developmental Biology **260,** 289-297.

Sumner, L.W., Mendes, P., and Dixon, R.A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry **62,** 817-836.

Sutton, F., Ding, X., Kenefick, D.G. (1992). Group 3 LEA Gene *HVA1* Regulation by Cold Acclimation and Deacclimation in Two Barley Cultivars with Varying

Freeze Resistance. Plant Physiology. **99,** 338-340.

Suzuki, M., Ketterling, M.G., Li, Q-B., McCarty, D.R. (2003). Viviparous1 alters global gene expression patterns through regulation of abscisic acid signalling. Plant Physiology. **132,** 1664-1677.

Swain, S.M., Reid, J.B., Kamiya, Y. (1997). Gibberellins are required for embryo growth and seed development in pea. Plant Journal **12,** 1329-1338.

Swain, S.M., Singh, D.P. (2005). Tall tales from sly dwarves: novel functions of gibberellins in plant development. Trends in Plant Science **10,** 123-129.

Swindell, W.R. (2006). The Association Among Gene Expression Responses to Nine Abiotic Stress Treatments in Arabidopsis thaliana. Genetics. **174,** 1811-1824.

Szekeres, M., Nemeth, K., Koncz-Kalman, Z., Mathur, J., Kauschmann, A., Altmann, T., Redei, G.P., Nagy, F., Schell, J., Koncz, C. (1996). Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in arabidopsis. Cell **85,** 171-182.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., Golub, T.R. Interpreting patterns of gene expression with self organizing maps: Methods and application to hematopoietic differentiation. Proceedings of the National Academy of Sciences of the United States of America **96,** 2907-2912.

Thelen, J.J. and Ohlorgge J.B. (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic engineering. **4,** 419-431

Thijs, G., Lescot, M., Marchal, K., Rombauts, S., De Moor, B., Rouze, P., Moreau, Y. (2001). A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. Bioinformatics **17,** 1113-1122.

Thijs, G., Marchal, K., Lescot, M., Rombauts, S., De Moor, B., Rouze, P., Moreau, Y. (2002). A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. Journal of Computational Biology **9,** 447-464.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Krufer, P., Selbig, J., Muller, L.A., Rhee, S. Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant Journal. **37,** 914-939.

Thomas, T.L. (1993). Gene expression during plant embryogenesis and germination: an overview. Plant Cell. **5,** 1401-1410.

Thomma, B.P.H.J., Eggermont, K., Penninckx, I.A.M.A., Mauch-Mani, B., Vogelsange, R., Cammue, B.P.A., Broekaert, W.F. (1998). Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. Proceedings of the National Academy of Sciences of the United States of America **95,** 15107-15111.

Thompson, A.J. (2000). Ectopic expression of a tomato 9-cis-epoxycarotenoid dioxygenase gene causes over-production of abscisic acid. Plant Journal **23,** 363-374.

Thompson JF (1980) Arginine synthesis, proline synthesis, and related processes. *In* BJ Miflin,ed, The Biochemistry of Plants, Vol 5. Academic Press, New York, pp 375-402

Thum, K.E., Shin, M.J., Palenchar, P.M., Kouranov, A., Coruzzi, G.M. (2004). Genome-wide investigation of light and carbon signaling interactions in Arabidopsis. Genome Biology **5**.

Tilton, G., Shockey, J. and Browse, J. (2000) Two families of acyl-CoA thioesterases in Arabidopsis. Biochemical Society Transactions. **28,** 946-947

Tischner, R. (2000). Nitrate uptake and reduction in higher and lower plants. Plant, Cell and Environment. **23,** 1005-1024.

Todd, R.B., Andrianopoulos, A., Davis, M.A. and Hynes, M.J. (1998). FacB, the Aspergillus nidulans activator of acetate utiliation genes, binds dissimilar DNA sequences. EMBO Journal **17,** 2042-2054.

Toorop, P.E., van Aelst, A.C., Hilhorst, H.W.M. (2000). The second step of the biphasic endosperm cap weakening that mediates tomato (Lycopersicon esculentum) seed germination is under control of ABA. Journal of Experimental Botany **51,** 1371-1379.

Trewavas, A. (2006) A brief history of systems biology. The Plant Cell **18,** 2420-2430.

Tsuda, K., Tsvetanov, S., Takumi, S., Mori, N., Atanassov, A., Nakamura, C. (2000). New members of a cold-responsive group-3 *Lea/Rab*-related *Cor* gene family from common wheat (*Triticum aestivum* L.). Genes and Genetic Systems. **75,** 179-188.

Tunnacliffe, A. and Wise, M.J. (2007). The continuing conundrum of the LEA proteins. Naturwissenschaften **94,** 791-812.

Turner, J.E. (2003). Characterisation of the acetate non utilisation mutant 1 of Arabidopsis. Ph.D. Thesis, University of Wales, Bangor.

Turner, J.E., Greville, K., Murphy, E.C., and Hooks, M.A. (2005). Characterization of Arabidopsis fluoroacetate-resistant mutants reveals the principal mechanism of acetate activation for entry into the glyoxylate cycle. Journal of Biological Chemistry **280,** 2780-2787.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., and Fernie, A.R. (2003). Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. Embo Reports **4,** 989-993.

Urbanczyk-Wochniak, E., Baxter, C., Kolbe, A., Kopka, J., Sweetlove, L.J., and Fernie, A.R. (2005). Profiling of diurnal patterns of metabolite and transcript abundance in potato (Solanum tuberosum) leaves. Planta **221,** 891-903.

van Helden, J. (2000). Regulatory sequence analysis tools. Nucleic Acids Research. **31** 3593-3596

van Engelen, F.A., Hartog, M.V., Thomas, T.L., Taylor, B., Sturm, A., van Kammen, A., de Vries, S.C. (1993). The Carrot Secreted Glycoprotein Gene Ep1 Is Expressed in the Epidermis and Has Sequence Homology to Brassica S-Locus Glycoproteins. Plant Journal **4,** 855-862.

van Roermund, C.W.T., Elgersma Y., Singh N., Wanders, J.A., Tabak, H.F. (1995) The membrane of peroxisomes in Saccharmyces cerevisiae is impermeable to NAD(H) and acetyl-CoA under in vivo conditions. EMBO J **14**: 3480-3486

Vanlerberghe, G.C., McIntosh, L. (1996). Signals regulating the expression of the nuclear gene encoding alternative oxidase of plant mitochondria. Plant Physiology **111,** 589-595.

Velayos, A., Eslava, A.P., Iturriaga, E.A. (2000). A bifunctional enzyme with lycopene cyclase and phytoene synthase activities is encoded by the carRP gene of Mucor circinelloides. European Journal of Biochemistry. **267**, 5509-5519.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. (1995). Serial Analysis of Gene-Expression. Science **270,** 484-487.

Vickery, M.L. and Vickery, B. (1981) Secondary plant metabolism. The MacMillan Press Ltd.

Vierstra, R.D. (1993). Protein degradation in plants. Annual Review of Plant Physiology and Plant Molecular Biology. **44,** 385-410.

Vleeshouwers, L.M., Bouwmeester, H.J., and Karssen, C.M. (1995). Redefining seed dormancy: An attempt to integrate physiology and ecology. Journal of Ecology **83**, 1031-1037

Voigt, B., Bewley, J.D., (1996). Developing tomato seeds when removed from the fruit produce multiple forms of germinative and post-germinative endo-β-mannase. Responses to desiccation, abscisic acid and osmoticum. Planta. **200,** 71-77.

von Roepenack-Lahaye, E., von Degenkolb, T., Zerjeski, M. (2004). Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. Plant Physiology **134,** 548-559.

Walenta, S. and Mueller-Klieser, W.F. (2004). Lactate: mirror and motor of tumor malignancy. Seminars in radiation oncology. **14,** 267-274.

Walenta, S., Schroeder, T. and Mueller-Klieser, W. (2004). Lactate in solid malignnt tumors: potential basis of a metabolic classification in clinical oncology. Curret Medicinal Chemistry. **11,** 2195-2204.

Wang, C., Wang, X. (2001). A novel phospholipase D of Arabidopsis that is activated by oleic acid and associated with the plasma membrane. Plant Physiology **127,** 1102-1112.

Wang, R.C., Okamoto, M., Xing, X., Crawford, N.M. (2003). Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. Plant Physiology **132,** 556-567.

Ward, J.L., Harris, C., Lewis, J., Beale, M.H. (2003). Assessment of ^{1}H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. Phytochemistry. **62,** 949–957.

Wareing,P.F., Saunders, P.F. (1971). Hormones and dormancy. Annual Review of Plant Physiology. **22,** 261-288.

Wasteneys, G.O. (2002). Microtubule organization in the green kingdom: chaos or self-order? Journal of Cell Science. **115,** 1345-1354.

Weckwerth, W. (2003). Metabolomics in systems biology. Annual Review of Plant Biology **54,** 669-689.

Weckwerth, W., Wenzel, K., and Fiehn, O. (2004). Process for the integrated extraction identification, and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics **4,** 78-83.

Weigel, D., Ahn, J.H., 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.Y., Xia, Y.J., Dixon, R.A., Harrison,

M.J., Lamb, C.J., Yanofsky, M.F., and Chory, J. (2000). Activation tagging in Arabidopsis. Plant Physiology **122,** 1003-1013.

Welbaum, G.E., and Bradford, K.J. (1990). Water relations of seed development and germination in muskmelon (Cucumis Melo L). V. Water relations of imbibition and germination. Plant Physiology. **92.** 1046-1062.

Welin, B.V., Olson, A., Nylander, M., Palva, E.T. (1994) Characterization and differential expression of dhn/lea/rab-like genes during cold acclimation and drought stress in Arabidopsis thaliana. Plant Molecular Biology **26,** 131–144

West, M.A.L., Harada. (1993). Embryogenesis in Higher-Plants - an Overview. Plant Cell **5,** 1361-1369.

West, M.A.L. (1994). Leafy Cotyledon1 Is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in Arabidopsis. Plant Cell **6,** 1731-1745.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J-C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., Hochstrasser, D.F. (1996). From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Bio-Technology **14,** 61-65.

Wise, M.J. and Tunnacliffe, A. (2004). POPP the question: what do LEA proteins do? Trends in Plant Science **9,** 13-17.

Xiong, L., Gong, Z., Rock, C., Subramanian, S., Guo, Y., Xu, W., Galbraith, D., Zhu, J. (2001). Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in Arabidopsis. Developmental Cell. **1,** 771-781.

Xu, D., Duan, X., Wang, B., Hong, B., Ho, T-HD. and Wu, R. (1996) Expression of a late embryogenesis abundant protein gene, HVA1, from barley confers tolerance to water defecit and salt stress in transgenic rice. Plant Physiology **110**, 249–257

Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y., Yamaguchi, S. (2004). Activation of Gibberellin biosynthesis and response pathways by low temperature during imbibition of Arabidopsis thaliana seeds. Plant Cell **16,** 367-378.

Yamaguchi, S., Kamiya, Y. (2001). Gibberrellins and light-stimulated seed germination. Journal of Plant Growth Regulation. **20,** 369-376.

Yang, Y.H., and Speed, T., (2003). Design and analysis of comparative microarray experiments. *In:* Statistical analysis of gene expression microarray data. Ed. Speed. Chapman and Hall CRC.

Yanofsky, C. (1981). Attenuation in the Control of Expression of Bacterial Operons. Nature **289,** 751-758.

Yuan, K., Wysocka-Diller, J. (2006). Phytohormone signalling pathways interact with sugars during seed germination and seedling development. Journal of Experimental Botany **57,** 3359-3367.

Zeng, Y., Kermode, A.R. (2004). A gymnosperm *ABI3* gene functions in a severe abscisic acid-insensitive mutant of Arabidopsis (*abi3-6*) to restore the wild-type phenotype and demonstrates a strong synergistic effect with sugar in the inhibition of post-germinative growth. Plant Molecular Biology **56,** 731-746.

Zhou, L., Jang, J.C., Jones, T.L. and Sheen, J. (1998). Glucose and ethylene signal transduction cross-talk revealed by an Arabidopsis glucose-insensitive mutant. Proceedings of the National Academy of Sciences of the United States of

America. **95,** 10294-10299.

Zinn, K., Dimaio, D., Maniatis, T. (1983). Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell. **34,** 865-87.

Zimmerman, P., Hirsch-Hoffmann, M., Hennig, L., Gruissem, W. (2004). GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiology **136,** 2621-2632.

Zolman, B.K., Monroe-Augustus, M., Thompson, B., Hawes, J.W., Kruckenburg, K.A., Matsuda, S.P.T. and Bartel, B. (2001) *chy1*, an Arabidopsis mutant with impaired β-oxidation, is defective in peroxisomal β-hydroxyisobutyryl-CoA hydrolase. Joural of Biological Chemistry. **276,** 31037-31046.

References

World Wide Web (WWW) addresses listed in the text:
(correct as of 1 September 2007)

Primer 3: Primer design website
http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

The Arabidopsis Information Resource
www.arabidopsis.org

Dr. D Galbraith's Homepage University of Arizona, U.S.A.
http://ag.arizona.edu/microarray/

Regulatory Sequence Analysis Tools website, Université Libre de Bruxelles.
http://rsat.ulb.ac.be/rsat/

Plant CARE (Cis Acting Regulatory Element) database
http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

Affymetrix
http://affymetrix.arabidopsis.info/

PLACE (Plant *cis*-acting regulatory elements) database
http://www.dna.affrc.go.jp/PLACE/index.html

ClustalW alignment at the European Bioinformatics Institute
http://www.ebi.ac.uk

TAIR gene ontology bulk analysis tool
http://www.arabidopsis.org/tools/bulk/go/index.jsp