Activation of seed-specific genes in leaves and roots of the desiccation tolerant plant, *Xerophyta humilis*.

Sally-Ann Walford

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Molecular and Cell Biology

UNIVERSITY OF CAPE TOWN

November 2008

Sally-Ann Walford

Activation of seed-specific genes in leaves and roots of the desiccation tolerant plant, *Xerophyta humilis*

November 2008

The ability of tissues to survive almost complete loss of cellular water is a trait found throughout the plant kingdom. While this desiccation tolerance is common in seeds of most angiosperms it is rare in their vegetative tissues. *Xerophyta humilis* (Bak.) Dur and Schintz belongs to a small group of resurrection angiosperms and it possesses the ability to withstand extreme desiccation of greater than 90% in both its seeds and vegetative tissues and return to active metabolism upon rehydration. We have tested the hypothesis that vegetative desiccation tolerance in angiosperms has evolved as an adaptation of seed desiccation tolerance. Microarray analysis was used to investigate whether there was significant overlap between the profiles of mRNA transcript abundance between desiccated leaves, roots and seeds from *X. humilis*. A normalized *X. humilis* library of 3105 cDNAs was used to compare the abundance of mRNA transcripts in hydrated leaves and roots with that in desiccated leaves, roots and mature seeds. A total of 2702 cDNAs were sequenced representing 1452 unique contigs. Cluster analysis revealed a large group of genes whose transcripts levels have increased in desiccated vegetative tissue and seed to levels that are greater than those seen in the hydrated tissues. A second group included genes showing up-regulation in desiccated leaf and seed and constitutive expression in root tissue, suggesting a protective mechanism constitutively present in root but only activated in leaf and seed during desiccation. Remaining clusters included genes that were down-regulated during desiccation and tissue specific genes. The *X. humilis* cDNA microarray data was compared with Atgen affymetrix data from *Arabidopsis thaliana*. The profiles of mRNA transcript abundance were compared between dehydrated and osmotically stressed vegetative tissue and mature seeds from this desiccation sensitive plant. Cluster analysis of the *A. thaliana* leaf and root water stress, and

seed development data sets revealed significant overlap between tissue types but no significant correlation between desiccated vegetative tissue and mature seed in *A. thaliana*. In this analysis, a number of genes show an increase in transcript abundance in mature *A. thaliana* seeds, and the orthologous genes accumulate transcript in desiccated vegetative tissues of *X. humilis*. The *X. humilis* microarray results were confirmed by real-time PCR and support the hypothesis that the activation of seed-specific gene expression in leaves and roots may have resulted in the acquisition of desiccation tolerance in *X. humilis*.

Acknowledgements

I gratefully acknowledge The National Research Foundation (NRF) and the Deutscher Akademischer Austausch Dienst (DAAD) for financial assistance.

I would like to thank Prof. Nicci Illing, Prof. Jill Farrant and Dr. Katherine Denby for their guidance and support throughout my doctorate.

A big thank you to all friends and family who have endured this journey with me. I am forever indebted.

Table of Contents

Chapter 1

Desiccation tolerance in resurrection angiosperms and orthodox seeds

1.1. Introduction

Water is a central component of life providing turgor pressure at the cellular level and hydrophilic and hydrophobic associations at the molecular level. It also maintains inter- and intra-molecular distances affecting protein conformation and partitions molecules within organelles. Additionally, as a reactant and a product of many reactions it is a key part of metabolism (Walters et al., 2002). All organisms living on land in dry air are exposed to the threat of drying out, a stress that most plants are not able to survive. Many plants, however, are able to tolerate minor fluctuations in water availability and can avoid or moderate water loss by employing morphological or physiological adaptations. Some plants evade water-deficit by undergoing growth and reproduction within a few weeks after significant rainfall and survive the dry periods as desiccation-tolerant seeds or spores (Bewley, 1979; Niklas, 1997). Many structural features such as thickening of cell walls, sunken stomata, low surface area/volume ratios, mucilaginous sheaths and leaf hairs serve to reduce water loss (Lerner, 1999) and some plants are able to remain hydrated by mining deepwater resources through their extensive root systems (Niklas, 1997). Plants can also avoid water loss by growing together in clumps or dense mats or by growing in sheltered, shady habitats thereby alleviating both dramatic changes in water availability and exposure to high temperatures (Lerner, 1999). While these measures are employed by a number of plants throughout the plant kingdom to lessen the severity of water loss they do not make these plants tolerant to desiccation and with prolonged drought these plants will die (Scott, 2000). There are only a few species that are truly desiccation tolerant and able to survive extreme levels of water-deficit. While desiccation tolerance is common in organs such as seeds and pollen only a few angiosperms possess mature foliage or vegetative tissue that is able to survive extreme water deficit. A large number of algae, lichen and bryophyte species (Proctor, 1990; Oliver, 1996; Oliver and Bewley, 1997; Kappen and Valadares, 1999; Oliver et al., 2000), some fern and fern-allies (Gaff, 1977; Bewley and Krochko, 1982) and a small group

of angiosperms, termed resurrection plants are able to withstand severe desiccation (Gaff, 1977; Porembski and Barthlott, 2000; Alpert and Oliver, 2002). The small group of taxonomically diverse, desiccation-tolerant angiosperms, are isolated to areas with restricted water availability and mostly colonize rocky outcrops with shallow soil (Porembski and Barthlott, 2000; Proctor and Pence, 2002). Resurrection plants are widely distributed and are found in all continents except Antarctica. The majority identified to date have been reported in tropical and subtropical zones in southern Africa, Australia, India and parts of South America (Gaff, 1977, 1987; Gaff and Bole, 1986) and although rarely found in Europe, a few species have been found in Western Balkan mountains (Stefanov et al., 1992).

.

1.2. Evolution of desiccation tolerance

The facts of the early evolution of desiccation tolerance are uncertain. It is believed that it was a primitive characteristic in the vegetative and reproductive stages of early land plants which allowed survival in dry air and colonization of land (Oliver et al, 2000, Oliver et al., 2005). This primitive mechanism, present in first spores, involved constitutive cellular protection and active cellular repair but it is thought that this trait was lost early in the evolution of tracheophytes (Oliver et al., 2000). As plants colonized land further they began to maintain internal water levels through the evolution of vascular tissues, and desiccation tolerance was lost in the vegetative tissues but conserved in seeds and spores. These reproductive structures were still subjected to desiccation in order to ensure survival upon dispersal. Oliver et al. (2000) propose that such propagules evolved as a modification of vegetative desiccation tolerance, using existing genes, and it is possible that desiccation-sensitive adult plants may still contain the genes needed for desiccation tolerance in a silenced form or redeployed in other functions (Bartels and Salamini, 2001; Zeng and Kermode, 2004). This developmentally induced seed desiccation tolerance program may then have been adapted by the vegetative tissue of some angiosperms, possibly through changes in regulatory genes (Bartels and Salamini, 2001), in response to environmental pressures associated with drying (Oliver et al., 2000). It is thought that desiccation tolerance may have re-evolved several times in different lineages allowing for the spread of a range of species

into very arid niches with limited water availability (Oliver et al., 2000; Porembski and Barthlott, 2000; Proctor and Pence, 2002).

1.3. Desiccation tolerance in seeds and resurrection angiosperms

The survival and dispersal of most seeds relies on their ability to tolerate desiccation (Bewley and Black, 1994; Dickie and Pritchard, 2002). In these seed types (termed orthodox) desiccation tolerance is acquired during seed development. During early development seeds are sensitive to drying but become tolerant at later stages (Kermode, 1990, 1995) and this tolerance is lost again upon germination (Kermode and Finch-Savage, 2002). Orthodox seed development can be divided into three main phases beginning with histodifferentiation, which is followed by a period of expansion and terminates with maturation of the embryo. Histodifferentiation is characterized by pattern formation and morphogenesis by mitotic division and cell differentiation occurring concurrently with the formation of the triploid endosperm or haploid megagametophyte. Once cell division ceases, cell expansion begins with reserve deposition occurring primarily in the storage tissues. Towards the end of reserve accumulation seeds enter a maturation stage during which orthodox types accumulate protection against desiccation and then the final stage is entered as maturation drying, when as much as 95% water is lost, and developmental arrest occurs as metabolism is switched off . Mature seeds are shed in a quiescent state (Bewley and Black, 1994; Kermode and Finch-Savage, 2002). In this state the plant embryo can survive diverse environmental stresses and are able to resume full metabolic activity, growth and development when conditions are favourable for germination. Water is taken up by the seed during imbibition promoting the recommencement of metabolic processes and the emergence of the radicle through the seed coat. At this stage the seedling loses its ability to survive desiccation (Kermode and Finch-Savage, 2002). All orthodox seeds can withstand dehydration to around 5%, even when maturation drying is not completed prior to shedding. Any seed that does not behave this way is not orthodox and have been described as being recalcitrant (Roberts, 1973). Recalcitrant seeds do not undergo maturation drying and remain desiccation sensitive both during development and after they are shed.

Truly orthodox seed species are able to survive the effects of complete water loss and the vegetative tissues of resurrection plants are able to acquire the same degree of extreme tolerance (Bewley, 1979; Gaff, 1989; Oliver et al., 1998). Vegetative tissues, like seeds, require time to adapt to water conditions of desiccation stress, presumably to induce protective mechanisms (Steponkus et al., 1995; Oliver et al., 1998; Farrant et al., 1999), and are able to remain in an air-dry state for prolonged periods by reaching a quiescent state comparable to that in seeds (Bartels, 2005). Both seeds and the vegetative tissue of resurrection angiosperms shut down metabolism on drying and recover on remoistening after periods of desiccation, suggesting parallels between vegetative desiccation recovery and the maturation and imbibition and germination of seeds (Proctor and Pence, 2002). A major difference between seeds and desiccation tolerant vegetative tissue, however, is that tolerance and desiccation are developmentally programmed in seeds but environmentally induced in mature plant tissue (Proctor and Pence, 2002; Berjak et al., 2007).

Survival of extreme desiccation poses several challenges including minimization of mechanical damage associated with turgor loss, preservation of the functional integrity of proteins, nucleic acids and membranes, minimization of toxin accumulation and free radical damage and, upon rehydration, activation of repair mechanisms (Bewley, 1979). As a result, the acquisition of desiccation tolerance is complex and in order for the cells of seeds and resurrection plants to survive desiccation they must prepare for this stress in advance. If drying occurs too rapidly some orthodox seeds have been shown to lose viability (Kermode, 1995; Kermode and Finch-Savage, 2002). Similarly rapid drying in the vegetative tissues of resurrection angiosperms result in the loss of the ability to recover from the stress (Gaff, 1989; Bewley and Oliver, 1992; Oliver and Bewley, 1997; Farrant et al., 1997). This would suggest that this ability to survive severe desiccation relies on a dynamic progression towards complete tolerance that requires specific biochemical modifications and the production of desiccation-related molecules (Phillips et al., 2002). As diverse mechanisms have been suggested to be involved in the acquisition of desiccation tolerance and maintenance of the integrity of dehydrated orthodox seeds, it is possible that any of these may be absent or ineffective in recalcitrant seeds.

The integrated mechanisms that allow whole seed or plant to survive desiccation are still not fully understood but recent molecular and biochemical research has shown that desiccation tolerant angiosperms and orthodox seeds tolerate desiccation by means of a set of synergistically acting protective mechanisms induced during drying. These include the synthesis of protective molecules, such as compatible solutes, non-reducing sugars that replace water (Hoekstra et al., 2001; Buitink et al., 2002) and act physically as volumetric and osmotic spacers preventing membrane adhesion (Koster and Bryant, 2005), proteins such as late embryogenesis abundant (LEA) proteins (Cuming, 1999) and small heat shock proteins (Wehmeyer and Vierling, 2000) that stabilize macromolecules and membranes and together with sugars form glasses that stabilize the entire subcellular milieu in the dry state (Koster and Leopold, 1988; and reviewed recently by Berjak et al., 2007). Additionally they have the ability to avoid damage from detrimental reactive oxygen species with antioxidants countering the damage and by repressing metabolism in a coordinated fashion (Leprince et al., 2000; Hoekstra et al., 2001; Walters et al., 2002).

Similarities in the responses between these two desiccation tolerant systems has led us to propose that resurrection plants may have adapted or re-evolved the developmentally regulated desiccation program of seeds in response to water deficit stresses encountered in their environments (Amuti and Pollard, 1977; Bewley and Krocko, 1982; Öpik, 1985; Koster and Leopold, 1988; Vertucci and Williams and Leopold, 1989; Bianchi et al., 1991; Leprince et al., 1993; Leopold et al., 1994; Farrant, 1995; Leprince and Walters-Vertucci, 1995; Vertucci and Farrant, 1995; Ingram and Bartels, 1996; Farrant et al., 1997; Oliver and Bewley, 1997; Farrant and Walters, 1998; Ghasempour et al., 1998; Pammenter and Berjak, 1999; Vicre et al., 1999; Farrant, 2000; Mundree and Farrant, 2000; Scott, 2000; Bartels and Salamini, 2001; Illing et al., 2005; Berjak, 2006). A closer investigation of the stresses imposed during extreme water loss and the counteractive mechanisms induced would certainly suggest an overlap in response to water deficit in these two systems. The stresses as a consequence of water loss and the induction and regulation of counter responses are expanded below with particular emphasis on the overlap between seed and desiccation tolerant vegetative tissues.

1.4. Stresses during water deficit and protective measures

Changes on a cellular and molecular scale occur to counteract the various stresses associated with water loss. The first stress is a mechanical one associated with the loss of turgor which affects the whole cell and the molecules within the cell (Steponkus, 1979; Steponkus and Lynch, 1989; Walters et al., 2002; Steponkus et al., 2005). Metabolism is also adversely affected and reactive oxygen species are generated as metabolic pathways cease to function efficiently. Water loss also results in protein denaturation and loss of enzymatic activity and consequently metabolic regulation and repair mechanisms are required to minimize or reverse the damage that occurs (Piatkowski et al. 1990, Ingram and Bartels 1996, Mundree and Farrant 2000; Leprince et al., 2000; Hoekstra et al., 2001; Bartels and Salamini 2001; Alpert and Oliver 2002; Berjak, 2006). Although there are many underlying mechanisms in common between seeds and resurrection plants, there are also differences as a consequence of different tissue types. For example, vegetative tissues are subjected to photo-oxidative stress, a stress not encountered by seeds.

1.4.1. Mechanical stress – consequences of cell volume reduction

Physical and physiological properties of the cell change as water is removed and one of the first changes is a reduction of cell volume. As the plant cell shrinks the plasmalemma may tear if it pulls away from the cell wall which contracts to a lesser extent than protoplasm (Walters et al., 2002). Additionally, an excessive decrease in the surface area of the plasmalemma will cause the cell to rupture upon rehydration and it is therefore critical to conserve the membrane surface during contraction (Steponkus, 1979; Steponkus and Lynch, 1989; Steponkus et al., 2005).

Both angiosperm resurrection plants and orthodox seeds appear to employ two general mechanisms to avoid mechanical stress; active and reversible cell wall folding and increased vacuolation with the replacement of water in these vacuoles with non-aqueous substances. The desiccated cells of orthodox seeds and some resurrection angiosperms (e.g. *Craterostigma* species) accommodate the changes in cell volumes by folding cell walls and the plasmalemma

remains intact and closely associated with these folded walls (Öpik, 1985; Vicre et al., 1999; 2003; 2004a). The components of cell walls of these resurrection plants do not appear to change during drying and it is believed that slight modifications to intrinsic wall characteristics allow stable reversible folding (Vicre et al., 1999; 2003; 2004a, b; Moore et al., 2006). In some of the *Craterostigma* species, xyloglucans already present in the cell wall may be cleaved into shorter more flexible units to assist wall folding and Ca^{2+} ions are also believed to have an important role in stabilizing walls in the dry state by cross linking wall polymers like acid pectins (Vicre et al., 1999; 2004b, Farrant, 2007). Expansins are believed to disrupt non-covalent bonds between polysaccharides to cause cell wall loosening (McQueen-Mason and Cosgrove, 1995) and highly mobile proteins like arabinose allow wall flexibility and have high water absorbing capacity which would be important for rehydration (Moore et al., 2006).

The large vacuole in the cells of some resurrection plants, such as the *Xerophyta* species, is replaced by a number of smaller vacuoles which fill the cytoplasm to maintain cell volume, reduce organelle compression and prevent membrane collapse and possible plasmalemma rupture. It appears that water is replaced by solid material in these vacuoles which is believed to reduce the degree to which they must contract and therefore minimize the tension in tonoplast membranes during drying (Farrant, 2000; Mundree and Farrant, 2000). The content of these vacuoles has been analyzed from a number of species and shown to contain proline, sucrose, protein and polyphenolics (van der Willigen et al., 2004; Moore et al., 2005a, b; 2007a). The accumulation of dry matter reserves is also observed during development in orthodox seeds and may account for the increasing tolerance to water loss in developing embryos (Vertucci and Farrant, 1995; Farrant et al., 1997; Farrant and Walters, 1998). The similar counteractive measures to cell volume reduction and mechanical stress in both seed and desiccation tolerant vegetative tissues imply an overlap in response to water deficit at a cellular level between these tissues.

1.4.2. Denaturation and subcellular disruptions

The reduction in cell volume as a consequence of water loss results in concentration of cellular contents promoting inappropriate molecular interactions leading to aggregation and denaturation of proteins and membrane damage. Metabolic pathways may be uncoupled as a consequence of protein misfolding with the resultant production of damaging free radical species. To ensure continued viability of cells during desiccation, the controlled downregulation of metabolism together with the synthesis of antioxidants are essential to the prevention of and clean up dangerous reactive of oxygen species (Piatkowski et al. 1990, Ingram and Bartels 1996, Mundree and Farrant 2000; Leprince et al., 2000; Hoekstra et al., 2001; Bartels and Salamini 2001; Alpert and Oliver 2002; Berjak, 2006). Additionally, the production of compatible solutes, sugars and desiccation-related proteins is believed to have a role in alleviating these stresses in seeds and vegetative tissue of resurrection plants.

1.4.2.1. Compatible solutes and sugars

The harmful effects of water loss are minimized by the production of compatible solutes which can occur at high intracellular concentrations without impeding normal cellular metabolism (Ramanjulu and Bartels, 2002). These are typically polyols, sugars, amino acids, betaines and related compounds (Bohnert et al., 1996; Ramanjulu and Bartels, 2002) and are believed to maintain turgor pressure and to stabilize proteins during the early stages of desiccation (Carpenter et al., 1990).

Sugars are thought to function as compatible solutes early during the drying process and finally as a replacement for water when the bulk cytoplasmic water is lost (Hoekstra et al., 2001). As structural water is removed protein conformation is compromised causing denaturation and aggregation of these molecules. Additionally, the phospholipid bilayers of membranes are disrupted promoting fusion between membranes that have become closely appressed causing irreversible structural and functional damage. It has been postulated that the hydroxyl groups of sugars may substitute for water to maintain hydrophilic interactions between and within macromolecules ensuring the correct spacing between phospholipid molecules thereby

preserving the native structure of proteins and membranes (Crowe et al., 1992, 1998). It is, however, a controversial hypothesis which has not been fully accepted.

Sugars are fundamental to the formation of cytoplasmic glass (Leopold et al., 1994) which prevents crystallization of the cytoplasm in dry cells (Vertucci and Leopold, 1986). Glass formation has been demonstrated in seeds (Williams and Leopold, 1989; Leopold et al., 1994; Leprince and Walters-Vertucci, 1995), pollen (Buitink et al., 1996) and vegetative tissue of resurrection species (Wolkers et al., 1998a) during drying. As desiccation tolerant tissues dry, the concentration of solutes increases which causes an increase in the viscosity of the cytoplasm and eventual vitrification to a glassy state (Burke, 1986). This glass is similar to solid brittle material but retains the physical properties of the liquid state (Franks et al., 1991) and is believed to increase the stability of enzymes (Chang et al., 1996) and prevent conformational changes in proteins (Pretrelski et al., 1993). Additionally, due to the increased viscosity in this state, molecular movement is slowed which restricts diffusion of reactive compounds within cells and promotes a state of metabolic quiescence (Slade and Levine, 1991; Roos, 1995). It has also been shown in model systems that glasses are capable of preventing the fusion of membranes (Leopold et al., 1994; Buitink et al., 2002). Although it is suggested that glass formation may not be a primary requirement for achieving desiccation tolerance it may be fundamental to survival of the dried state as it protects cellular constituents and retards degradative processes that may occur during storage (Buitink et al., 2000a, b, c). In addition to sugars, other molecules are believed to be involved in glass formation including proteins, such as LEA proteins (Late Embryogenic Abundant) (Leprince and Walters-Vertucci, 1995; Sun and Leopold, 1997; Wolkers et al., 1998b; Buitink et al., 2000a, b). LEA proteins are able to change the hydrogen-bonding properties of model sugar systems indicating a possible involvement in intracellular glass formation (Wolkers et al., 2001). These proteins will be discussed in more detail below.

Sucrose (a disaccharide), raffinose and stachyose (oligosaccharides), are the most commonly accumulated sugars in orthodox seed and the resurrection plants examined to date (Amuti and Pollard, 1977; Bewley and Krocko, 1982; Koster and Leopold, 1988; Bianchi et al., 1991; Leprince et al., 1993; Ghasempour et al., 1998; Vertucci and Farrant, 1995; Ingram and Bartels,

1996; Oliver and Bewley, 1997; Pammenter and Berjak, 1999; Scott, 2000; Bartels and Salamini, 2001; Whittaker et al., 2001; Illing et al., 2005; Berjak, 2006; Peters et al., 2007). It is interesting to note that trehalose, a sugar which has been proposed to be an excellent membrane stabilizer (Kaushik and Bhat, 2003) and water replacement molecule in animal systems (Crowe et al., 1986; 1987; 1998) does not seem to accumulate to any significant extent in these plant systems (Bewley and Black, 1994; Farrant, 2007). Sucrose has been suggested to be the alternate trehalose and is thought to be a main player in the water replacement theory (Clegg, 1986) and also in the formation of intracellular glass (Leopold et al., 1994). However, it should be kept in mind that sugars may be interchangeable while affording similar protection and the type of sugar accumulated is associated with the metabolism occurring in the hydrated tissue prior to drying (Farrant, 2007).

Photosynthesis and starch breakdown are sources of carbon for the accumulation of hexose sugars, sucrose and amino acids observed in the initial phase of desiccation in the resurrection angiosperm, *Sporobolus stapfianus*. This phase is characterized by increases in sucrose phosphate synthase (SPS) and pyruvate kinase (PK) activities and maximal phosphoenolpyruvate carboxylase (PEPCase), NADP-dependent isocitrate dehydrogenase (ICDH), and NADH-dependent glutamate synthase (GOGAT) activities (Whittaker et al., 2007). SPS is involved in sucrose synthesis using glycolysis intermediates, fructose-6-phosphate and UDP-glucose, as substrates. PEPCase and PK are involved in providing glycolytic carbon for amino acid biosynthesis (Lunn and Furbank, 1999; Stitt et al., 2002; Foyer et al., 2003) while glutamate, the primary donor to amino acids, is produced mainly by chloroplast ferredoxin dependent-GOGAT and also in the plastids of vascular tissues by NADH-GOGAT) under stress conditions (Lancien et al., 2000). The increase in activity of these enzymes indicates the movement of carbon from sugar phosphate pools to sucrose and amino acid biosynthesis during the initial phases of dehydration in this plant. Subsequently, when starch is completely degraded and photosynthesis has ceased, hexose sugars are phosphorylated and the rate of sucrose accumulation increases. At this stage SPS protein levels are at a maximum while PK, PEPCase and GOGAT activities decrease. This indicates that SPS may have an increased capacity in carbon utilization during

desiccation as catabolic respiration has declined evidenced by the decrease in the enzymes involved in amino acid biosynthesis (Whittaker et al., 2007).

SPS and sucrose synthase transcript levels increase during re-establishment of desiccation tolerance in *Medicago trunculata* seedlings after incubation in PEG indicating a role in water deficit stress (Buitink et al., 2006). SPS activity also increases in *C. plantagineum* during dehydration but it is not correlated with an increase in transcript levels and the increased activity may therefore be a reflection of the activation state of the enzyme (Ingram et al., 1997). Additionally, the transcript levels of enzymes involved in the synthesis of other compounds that can act as compatible solutes are upregulated during drought in *C. plantagineum* and these include δΔ1-pyrroline-5-carboxylate synthetase which is involved in proline biosynthesis, and betaine aldehyde dehydrogenase which is involved in glycine betaine biosynthesis (Ingram and Bartels, 1996). In *X. viscosa* a galactinol synthase (GolS), a *myo*-inositol 1-phosphate synthase (MIPS) and an aldose reductase (AR) are upregulated during desiccation (Mundree et al., 2002; Peters et al., 2007; Lehner et al., 2008). GolS is involved in the first step in raffinose family oligosaccharide (RFOs) production (Mundree et al., 2002) and considerable raffinose is synthesized in response to drying in *X. viscosa*. MIPS catalyzes the first step in the formation of inositol (Ino), which although itself is not accumulated during desiccation of *X. viscosa*, Ino plays a central role in the synthesis of many compounds (e.g. D-glucuorini acid; glycosides, galactinol) that do play a role in protection against desiccation (Peters et al., 2007; Lehner et al., 2008). AR belongs to the aldo-keto reductase superfamily, that reduce sugars to their corresponding alcohols, and is known to be involved in the synthesis of sorbitol from glucose-6-phosphate (Bohren et al., 1989). The levels of sorbitol were shown to be elevated in the leaves during desiccation stress and it is likely that the aldose reductase functions in polyol biosynthesis to convert glucose-6-phosphate to sorbitol in *X. viscosa* (Mundree et al., 2000). The upregulation of genes together with changes in levels and activity of the proteins involved in sucrose and amino acid biosynthesis would suggest that osmoprotectants have an important role during desiccation in resurrection angiosperms. The similarity in the production of sugars during

desiccation in both seeds and vegetative tissue of resurrection angiosperms underscores another important mechanism shared by both tissues in response to water deficit.

1.4.2.3. Late Embryogenic Abundant (LEA) proteins

The expression of LEAs has been shown to be concomitant with the acquisition of desiccation tolerance in orthodox seeds and resurrection angiosperms. They were originally isolated from seeds (Baker et al., 1988; Galau et al., 1993) and are known to accumulate during the late maturation stage during seed development (Bartels et al., 1988; Baker et al., 1995; Blackman et al., 1995; Ingram and Bartels, 1996; Kermode, 1997; Oliver and Bewley, 1997; Cuming, 1999), in dehydrating vegetative tissues of desiccation tolerant plants (Ingram and Bartels, 1996; Oliver and Bewley, 1997; Blomstedt et al., 1998; Collett et al., 2004) and in sensitive species subjected to mild dehydration (Close, 1997). They are also induced by cold, salt, osmotic stress or exogenous abscisic acid treatment and a few of them are expressed constitutively (Illing et al., 2005).

The existence of LEA genes were initially demonstrated in developing cottonseeds (Cuming, 1999) and orthologues have subsequently been identified in many angiosperms. The grouping of these genes was originally based on their similarity to the first LEAs isolated from cotton and their expression patterns (Bray, 1993; 1997). However, a recent approach involving the analysis of peptide composition of the proteins rather than sequence similarity, based on Protein or Oligonucleotide Probability Profiles (POPP), has classified them into 9 groups, LEA-1 to LEA-9 (Table 1.1) (Wise, 2003; Wise and Tunnacliffe, 2004).

To date, the main assumption that LEAs are involved in desiccation tolerance is based on the appearance and accumulation of LEAs in both orthodox seed maturation and during stresses that cause water deficit in plants, particularly desiccation tolerant plants (Cuming, 1999). While there is relatively little evidence for a direct protective role for LEAs during desiccation, some plant LEA proteins have been shown to confer increased resistance to osmotic and freeze stresses when expressed in yeast (Imai et al., 1996; Swire-Clark and Marcotte, 1999; Zhang et al., 2000) and improved tolerance to water deficit in transgenic rice and wheat (Xu et al., 1996; Sivamani et

al., 2000). *In vitro* protection assays have shown that selected LEA proteins are able to prevent enzyme inactivation as a result of water depletion and freezing stress (Rinne *et al.*, 1999; Reyes *et al.*, 2005; Goyal *et al.*, 2005). Additionally, recent experiments have shown that two LEAs are able to prevent protein aggregation during drying (Goyal *et al.*, 2005), supporting the suggestion that LEAs play an important role in protecting tissues from the effects of water loss. It is possible that members of this group of proteins have different protective roles at different levels of hydration but these suggestions are purely based on conjecture with comparatively little supporting evidence (Boudet et al., 2006).

Table 1.1: Recent assignment of LEA proteins according to Interpro superfamilies and PFAM domains and common sequence motifs (Wise and Tunnacliffe, 2003, 2004).

LEA group	***Interpro Superfamily***	***PFAM domain***	***Sequence motif***
LEA-1	IPR000389	PF00477	GGQTRREQLGEEGYSQMGRK
LEA-2	IPR000167	PF00257	DEYGNP (Y domain) EEKK (K domain) Sn (S segment)
LEA-3	IPR004238	PF02987	TAQAAKEKAGE
LEA-4	IPR005513	PF03760	
LEA-5[1]	IPR004238	PF02987	
LEA-6	IPR007011	PF04927	
LEA-7	IPR004926	PF03242	
LEA-8	IPR004864	PF03168	
LEA-9[2]		PD68804	

[1] Group 5 LEA proteins contain the same domain as group 3 LEA proteins but with additional copies of the domains.

[2] Group 9 LEA proteins contain uncharacterized LEA proteins.

The precise function of LEAs in desiccation tolerance remains unclear as the majority of the work to date is based on DNA sequence characterization and predictive protein function rather than, with a few exceptions (Russouw et al., 1995; 1997; Mtwisha et al., 1998; Wolkers et al., 2001; Goyal et al., 2005), the actual protein function. LEA proteins appear to be natively unfolded in the hydrated state (Goyal et al., 2003) and secondary structure prediction programs show significant unstructured (loop) regions for Group 1 and 2 proteins while a high degree of folding mainly into amphipathic α-helices for group 3 and group 6 LEA proteins is suggested (Wise and Tunnacliffe, 2004). The high hydrophilic amino acid content, unstructured nature in the hydrated state and thermostabilility of these proteins has led to inference of possible functions including water retention, ion sequestration, chaperonin activity, prevention of protein or membrane aggregation and facilitation of glass formation together with sugars (Ingram and Bartels, 1996; Bray, 1997; Cuming, 1999; Wolkers et al., 2001; Mtwisha et al., 2006). LEA proteins belonging to groups 1 to 3 are the most well characterized to date.

Group 1 LEAs are believed to act as “water replacement” molecules. They are characterized by a 20-aa motif (Table 1.1) and contain a particularly high number of charged and uncharged polar residues indicating an extremely high potential for hydration (McCubbin et al., 1985). It is therefore possible that these proteins may bind intracellular macromolecules and coat them with a cohesive layer of water preventing their aggregation during drying (Close, 1996). Even with the eventual removal of their own hydration shell with continued drying, these proteins would still be able to contribute to macromolecular stabilization through the interaction of their own hydroxylated residues with the surface group of other proteins, in essence acting as a layer of “replacement water” (Cuming, 1999). These water retaining properties together with their structural flexibility has allowed the speculation that these LEAs could provide a hydration shell around macromolecules and membranes in desiccated tissues of resurrection plants and mature seeds (Cuming, 1999; Alpert and Oliver, 2002; Buitink et al., 2002).

The Group 2 LEAs, also known as dehydrins, are the most widespread and most studied of the LEAs. They are characterized by a lysine-rich K-segment, a stretch of serine residues (the S

segment) and a conserved motif, the Y segment, towards the N-terminus of the protein (Table 1.1) (Close, 1997). The K-segment has the potential to form amphipathic α-helices (Close, 1996), which in addition to further highly polar, unstructured repeating Φ regions are proposed to form intra- and intermolecular interactions thereby stabilizing and protecting various cellular constituents during desiccation (Baker et al., 1988).

The LEAs in Group 3 are also proposed to have the potential to form α-helices via their 11-mer repeat sequences (Table 1.1). It is suggested that these amphipathic helices, through hydrophobic interactions with each other, form bundles exposing highly charged surfaces which can bind ions (Dure, 1993). This ion sequestration may be essential to prevent irreversible damage to cellular proteins and structural components as a consequence of increasing ionic strength during drying. These proteins have also been shown to stabilize sucrose glass formation and it is possible that they may act as anchors in a structural network that stabilizes cytoplasmic constituents in the dry state (Wolkers, 1998b).

Current reports on LEAs in resurrection plants are based mainly on transcriptome analysis. A mini-microarray experiment showed the upregulation of 55 genes during desiccation in the leaves of the resurrection angiosperm, *X. humilis*, 16 of which were identified as LEAs (Collett et al., 2004) and a number of desiccation induced LEAs have also been reported from other resurrection species including *X. viscosa* (Mundree and Farrant, 2000; Ndima et al., 2001), *S. stapfianus* (Blomstedt et al., 1998) and *C. plantigineum* (Piatkowski et al., 1990). LEA homologues have also been identified in other desiccation tolerant species including nematodes and bacteria (Goyal et al., 2003; Browne et al., 2004; Wise and Tunnacliffe, 2004), yeast (Garay-Arroyo *et al.*, 2000; Sales *et al.*, 2000) and bryophytes (Alpert and Oliver, 2002). Northern blot analysis of the LEA transcripts identified in *X. humilis* show induction specifically at late stages of desiccation, suggesting they be a distinct set of desiccation-specific LEAs that deal with stresses associated with extreme desiccation (Illing et al., 2005).

A comparative analysis of LEA gene expression in abiotically stressed vegetative tissue and developing seeds of the desiccation-sensitive *Arabidopsis thaliana* showed significant expression of a number of LEAs during seed development but not in the stressed vegetative tissues (Illing et al., 2005). This is in sharp contrast to the observed LEA expression in the dry tissues of the desiccation-tolerant *X. humilis* (Collett et al., 2004). It was also noted that orthologues of the "seed specific" LEA 6 from *A. thaliana* were found to accumulate in the desiccated leaf of *X. humilis* indicating a possible role in protection against the severe water loss associated with seed development and desiccation in vegetative tissues of resurrection angiosperms (Illing et al., 2005).

Further support for the role of LEAs in seed development and associated desiccation stress comes from gene expression studies on the acquisition (or re-induction) of desiccation tolerance of *M. truncatula* seeds. A total of 18 genes encoding LEA proteins were identified among 187 genes which were expressed both during developmental acquisition of desiccation tolerance (seed maturation) and during the re-establishment of desiccation tolerance in seedlings after exposure to PEG (Buitink et al., 2006).

The expression level of LEAs or dehydrins in recalcitrant seeds is inconsistent as some species do not appear to express these proteins while others do to varying degrees. The transcripts for dehydrins have been reported to be absent during the late stages of development of the extremely desiccation sensitive seeds of *Avicennia marina* (Farrant et al., 1992) but they have been detected in seeds of *Zizania palustri* which exhibit characteristics intermediate between recalcitrance and orthodoxy (Bradford and Chandler, 1992). The seeds of *Z. palustri* are intolerant to dehydration at low temperature and this does not appear to be due to and inability to accumulate ABA or due to an absence of dehydrins (Bradford and Chandler, 1992), which would suggest that the presence of dehydrins alone is most likely insufficient to protect against desiccation injury (Blackman *et al.*, 1992; Bradford and Chandler, 1992). Dehydrins have been detected in mature seeds of five desiccation sensitive (recalcitrant) trees (all temperate species) (Finch-Savage et al., 1994) and in a study on mature seeds of two tropical tree species, *Castanospermum australe* and

Trichilia dregeana, dehydrin-type proteins were only found to accumulate in the former (Han et al., 1997). Additionally, the immature seeds and the seedlings of *C. australe* responded to dehydration, ABA application or exposure to cold with the production of these dehydrin-type proteins while those of *T. dregeana* did not (Han et al., 1997). These observations would suggest that the ability to express LEA's or dehydrin-like proteins are not an indication of the capacity of seeds of particular species to withstand dehydration. It implies rather that desiccation tolerance must be the result of the interactions of more than one mechanism. However, the variable expression of LEAs/dehydrins in the seeds of different recalcitrant species together with the presence or absence of other factors could account for the level of non-orthodox behaviour exhibited under certain conditions.

1.4.2.4. Heat shock proteins

There is increasing evidence for the role of heat shock proteins (HSPs) in cellular protection during desiccation. The small HSPs (sHSPs, 15-42 kDa) are the most prominent HSPs in plants (Waters et al., 1996) and they accumulate in the maturing seeds of many plants species prior to desiccation and persist in the dry state indicating a role in the acquisition of desiccation tolerance (Vierling, 1991; Coca et al., 1994; Wehmeyer et al., 1996; Kermode and Finch-Savage, 2002). *A. thaliana* mutants with reduced levels of HSP17 showed compromised tolerance to mild water loss (Yamagishi et al., 2005) and this observation together with their implication in seed longevity (Bettey and Finch-Savage, 1998) supports a role for sHSPs in desiccation tolerance.

HSPs are also expressed in the resurrection plant *C. plantagineum* where they are constitutively present in the vegetative tissues and are induced to higher levels in response to desiccation and heat shock. Additionally, exogenous application of the stress hormone, abscisic acid (ABA), induces the expression of sHSPs and the acquisition of desiccation tolerance in previously desiccation-sensitive *C. plantagineum* callus (Alamillo et al., 1995). In addition to the sHSPs, a member of the HSP90 family (Grp94) was found to be induced by desiccation and heat stress in *X. viscosa* (Walford et al., 2004). To date there is little experimental evidence that points to a specific role for HSPs in desiccation tolerance but they are thought to offer a general protective

role in the dry state based on their chaperone-like activity (Alpert and Oliver, 2002). Chaperones bind to proteins to minimize inappropriate interactions thereby maintaining protein structure under denaturing conditions. They may also facilitate appropriate refolding upon rehydration (Alpert and Oliver, 2002; Mtwisha et al., 2006).

1.4.3. Oxidative stress

In all plants reactive oxygen species (ROS) accumulate, mainly in mitochondria and chloroplasts, as a natural consequence of normal metabolic processes involving electron transport. Oxygen is an essential component of metabolism in aerobic life forms but it can form other oxygen radicals such as singlet oxygen ($_1O_2$), superoxide (O_2^-), the hydroxyl radical (OH) and nitric oxide (NO). The unpaired electrons of these free radicals are easily donated and thus highly reactive (Halliwell and Gutteridge, 1999) and can cause permanent damage to enzymes (Wolff et al., 1986; Dean et al., 1993; Halliwell and Gutteridge, 1999), membranes (Senaratna and McKersie, 1983, 1986; McKersie et al., 1988, 1989; Halliwell and Gutteridge, 1999; Leprince et al., 2000) and chromosomes (Dizdaroglu, 1994). Plants possess a multitude of free radical scavenging systems (Larson, 1988; Hendry, 1993, Appel and Hirt, 2004; Møller et al., 2007) including superoxide dismutase (SOD), ascorbate peroxidase (APX), glutathione peroxidase (GPX) and catalase (CAT) (Møller et al., 2007). SOD is the first line of defense against ROS dismutating superoxide to H_20_2, while APX, GPX and CAT subsequently detoxify the resulting H_2O_2 (Møller et al., 2007). Plants, unlike other organisms possess a multitude of genes encoding SOD and APX and different isoforms are targeted to the chloroplast, mitochondria, peroxisomes, cytosol and apoplast while GPX is cytosolic and CAT is found in peroxisomes (Møller et al., 2007). These, in addition to the non-enzymatic antioxidants including glutathione (GSH), ascorbic acid (Asc) (Noctor and Foyer, 1998), tocopherols and β-carotene (Munne-Bosch and Alegre, 2002) operate under hydrated conditions as "housekeeping" systems (Illing et al., 2005) to counteract ROS induced damage.

While the protective role of antioxidants in the survival of desiccation in seeds is not clear (McKersie, 1991; Leprince et al., 1993) it has been assumed that the antioxidant systems should

function maximally during initial stages of maturation drying of developing orthodox seeds, and to facilitate recovery during early germination (Arrigoni et al., 1992; Buitink et al., 2002; Berjak et al., 2007). At lower water contents, molecular antioxidants such as glutathione, ascorbate and tocopherol may quench remaining ROS to lessen the oxidative stress (Senaratna and McKersie, 1986). The thiol-requiring 1-cys peroxiredoxin, originally identified in the seeds of the desiccation sensitive angiosperms (Aalen, 1999), is thought to protect tissues from ROS damage during drying and early imbibition (Haslekas et al.1998, Stacy et al., 1999). The generation of free-radicals has been proposed to a major cause of metabolic damage during dehydration of recalcitrant seeds (Berjak and Pammenter 1997; Côme and Corbineau 1996a, 1996b; Smith and Berjak 1995) particularly because protective mechanisms appear to become impaired under conditions of water stress (Senaratna and McKersie 1986, Smith and Berjak 1995). Research has shown that free radicals are produced as a consequence of water stress in desiccation-sensitive seeds and also that the antioxidant systems are ineffective at controlling them and these factors can be considered to constitute on of the major causes of desiccation sensitivity (Hendry, 1992; Kermode and Finch-Savage, 2002).

In vegetative tissues abiotic stresses such as water deficit, disrupt the electron transport chain compromising the equilibrium between the production and scavenging of ROS (Hendry, 1992; Smirnoff, 1993; Kranner and Grill, 1996; Kranner and Birtic, 2005; Kranner et al., 2006; Møller et al., 2007). Resurrection angiosperms continue respiration to low relative water contents (RWC), albeit it at very low rates (Schwab et al., 1989; Hartung et al, 1998; Tuba et al., 1998; Farrant, 2000; Vander Willigen et al., 2001; Mundree et al., 2002). Continued respiration under these conditions produces chemical energy for the production of molecules that protect against adverse effects of water loss such as compatible solutes, LEAs and HSPs but it can also result in excess ROS production. Increased antioxidant capacity through the upregulation of "housekeeping" antioxidants and the preservation of antioxidant activity in the desiccated state have been suggested to reduce the desiccation-induced ROS production in these plants (Illing et al., 2005; Farrant, 2007).

In seeds ROS accrue mainly from respiratory metabolism (Hendry, 1993; Bailly, 2004), while vegetative tissues have an additional contribution from the disruption of photosynthesis, resulting in the transfer of excess excitation energy from chlorophyll to oxygen and leading to the rapid formation of singlet oxygen, superoxide and hydroxyl radicals (Halliwell, 1987; Seel et al, 1992a,b,; Smirnoff, 1993). As a result desiccation tolerant plants protect against photo-oxidative damage through the downregulation of photosynthesis together with the initiation of free radical scavenging systems. Photosynthesis can be regulated in one of two ways during drying. The first involves the reversible breakdown of chlorophyll and dismantling of thylakoid membranes (Tuba et al., 1993a,b; Sherwin and Farrant, 1998; Farrant, 2000; Mundree and Farrant, 2000). This strategy, termed poikilochlorophylly, efficiently minimizes photosynthetically induced ROS production and has been proposed to be key to the longevity of poikilochlorophyllous species in the dry state. Homoiochlorophylly is an alternative approach adopted by some species. In this case chlorophyll is retained and the thylakoid membranes remain intact in the dry state. Light-chlorophyll interactions are minimized by leaf folding or rolling (Dalla Vecchia et al., 1998; Sherwin and Farrant, 1998; Farrant et al., 1999; Farrant, 2000) and by anthocyanin pigment accumulation in surfaces that remain exposed. These pigments are believed to reflect photosynthetically active light shielding chlorophyll and hence act as antioxidants (Smirnoff, 1993; Sherwin and Farrant, 1998; Farrant, 2000; Farrant et al., 2003; Moore et al., 2007a,b). A reduction in metabolism at low RWC in desiccation tolerant angiosperms is supported by the increasing evidence that a number of genes linked with metabolism are downregulated during desiccation stress (Farrant, 2007). It has been proposed that a concerted downregulation of energy metabolism during desiccation is required to reduce the generation of ROS to ensure survival in the dried state (Leprince and Hoekstra, 1998; Leprince et al., 2000).

There is a growing body of evidence that reduction of metabolism coincides with desiccation tolerance (Leprince et al., 1999, Pammenter and Berjak, 1999) and it appears that a coordinated control of energy metabolism at the early stages of dehydration or during the acquisition if desiccation tolerance appears to be essential in avoiding oxidative stress (Hoekstra et al., 2001). Notably, down-regulation of metabolism appears to be an ancient regulatory mechanism that

allows a wide range of organisms to withstand severe environmental stresses such as anoxia, freezing and dehydration (Hand and Hardewig, 1996; Hardie et al., 1998).

A decline in the rate of respiration has been shown to be concomitant with the acquisition of desiccation tolerance in developing pea embryos and carrot somatic embryos (Rogerson and Matthews, 1997; Tetteroo et al., 1995). Leprince et al. (1999) have shown that increased sensitivity to desiccation is associated with an inability to repress metabolism and analysis of fermentation products, in conjunction with the measurement of membrane damage during dehydration has provided evidence that desiccation tolerance is associated with a tight control of metabolism during dehydration (Leprince et al., 2000). The precise triggers of metabolic arrest during drying are not known but electron microscopy studies have suggested that the decline in respiratory rate may be associated with a de-differentiation of organelles (Leprince et al., 1990; Farrant et al., 1997). A strategy like this would reduce the surface area of vulnerable membranous structures and reduce respiratory metabolism which has the potential for uncontrolled free-radical generation under conditions of water stress (Senaratna and McKersie, 1986). Hoekstra and Leprince (1998) have suggested that the rapid increase in viscosity during the onset of drying retards diffusion and that this can be regarded as a mechanism to slow metabolism and decrease O_2 concentration, thereby reducing the chance of generating ROS. A number of previous studies of mitochondria have shown that both the rates of electron transport and the mobility of the redox components within the lipid bilayer are more sensitive to changes in the medium bulk viscosity than in membrane microviscosity (Fato et al., 1993; Chazotte, 1994; Esmann et al., 1994).

Recalcitrant seeds are sensitive to damage in the water content range where free-radical mediated processes are thought to occur. This would suggest that desiccation sensitive seeds do not possess or express the mechanisms that permit orthodox seeds to pass through this water content range with little or no damage (Pammenter and Berjak, 1999). Additionally, respiration rates of mature recalcitrant seeds are relatively high (Poulsen and Eriksen, 1992; Farrant *et al.*, 1997;

Finch-Savage, 1996), which suggests that these seeds are metabolically active and this metabolic activity may be associated with desiccation sensitivity.

In desiccation tolerant systems, these avoidance mechanisms, which are initiated at relatively high water potentials (Sherwin and Farrant, 1998; Tuba et al., 1998; Farrant, 2000), act together with antioxidants (Bewley, 1979; Dhindsa, 1987; Smirnoff, 1993; Foyer et al., 1994; Kranner and Grill, 1997; McKersie et al., 1988; Sherwin and Farrant, 1998; Pammenter and Berjak, 1999; Farrant, 2000) to moderate any ROS associated damage occurring due to continued respiration at lower levels of hydration (Vertucci and Leopold, 1984; Vertucci and Roos, 1990; Salmen Espindola et al., 1994; Leprince and Hoekstra, 1998; Leprince et al, 1999; Farrant, 2000; Walters et al., 2001). While desiccation sensitive species also upregulate these "housekeeping" antioxidants, desiccation tolerant plants appear to have an improved antioxidant capacity through their ability to maintain the antioxidant activity of these molecules and to produce, *de novo*, "specialist" antioxidants during drying (Farrant, 2007). These "specialist" antioxidants refer to those genes whose transcripts have been shown to increase in mature, dry seed including 1- and 2-cys-peroxiredoxins, glyoxalase I family proteins, zinc metallothionein and metallothionein-like antioxidants (Blomstedt et al., 1998; Mowla et al., 2002; Collett et al., 2004; Illing et al., 2005). These genes have not been found to be upregulated in the vegetative tissues of desiccation sensitive plants (Aarlen, 1999; Stacey et al, 1999). A wide variety of amphiphilic compounds (such as phenolic acids and flavonoids) present in plants are known to be potent antioxidants (Larson, 1988; Wang et al., 1996; Kahkonen et al., 1999). They are present at high levels in dry seeds, pollen and resurrection plants and potentially play several roles in desiccation tolerance (Larson, 1988; Oliver et al., 1998; Buitink et al., 2000d; Moore et al., 2005b; Farrant, 2007). There is evidence for the movement of these molecules from the cytoplasm to membranes at the early stages of drying (Buitink et al., 2000d; Golovina and Hoekstra, 2001) and the partitioning of such antioxidants into membranes might prevent desiccation induced oxidative damage.

A comprehensive comparison of antioxidants across desiccation tolerant and sensitive species is lacking due to discrepancies in reporting and experimental conditions. The water content is seldom reported and the conditions under which the plants were dried are not presented and the

method of enzyme activity quantitation differs. However, a comparison between three antioxidant enzymes in desiccation tolerant and sensitive species was recently performed (Illing et al., 2005). Glutathione reductase (GR) was shown to be elevated in dry and hydrated tissue in most of the species tested suggesting it behaves as a housekeeping protectant rather than a having specific role in desiccation tolerance (Illing et al., 2005). APX did not show a consistent trend among tolerant tissues or between tolerant and sensitive plants. Additionally, APX activity appears to decrease during seed desiccation (Bailly, 2004) and it has been suggested that the ascorbate system is probably not involved in desiccation tolerance in seeds and resurrection angiosperms. SOD activity was down-regulated in desiccation sensitive species which would suggest a role specific to desiccation tolerance but a large body of literature suggests that the SOD enzymes are probably also housekeeping and may play a role in response to stress (Pammenter and Berjak, 1999; Bailly, 2004)

Investigation of enzyme activity showed initial increases in APX, GR and SOD in desiccation tolerant and sensitive species but the activity ceased in sensitive plants dried below their critical water contents and remained at elevated levels in the tolerant plants (Illing et al., 2005; Farrant, 2007). Mature orthodox seeds of both tolerant and sensitive species also retained significant antioxidant activity. Although this is an *in vitro* assay, only reflecting the potential for enzymatic activity, the fact that these enzymes remained active after extraction from dry tissue indicates that they were protected in the dry state. Hence, the evidence to date suggests that increased levels of antioxidants together with the preservation of the antioxidant activity may increase tolerance to desiccation. This is supported by a study in which maintenance of the antioxidant capability of glutathione in particular was shown to be crucial for the survival of desiccation in a number of desiccation tolerant systems (Kranner and Birtic, 2005; Kranner et al., 2006). These results correlated with longevity studies on *Myrothamnus flabellifolius* (Farrant and Kruger, 2001; Kranner et al., 2002). Additionally it has been shown that loss of GR, catalase and SOD activity was concomitant with loss in viability of dry stored *C. wilmsii* and *X. humilis* (Farrant, 2007). The ability to preserve antioxidant capacity together with the activation of “specialized”

seed-specific antioxidants offers tolerant species an advantage over sensitive tissues to survive severe desiccation.

The challenges of surviving desiccation are similar for seeds and resurrection plants. As shown above there is a large overlap in those genes (e.g. LEAs, sHSPs and antioxidants), that are expressed in desiccated seeds which leads one to consider the similarity in the regulation of expression of these genes in response to desiccation.

1.5. Regulation of protective measures during drying in vegetative tissue of resurrection plants and maturation drying in seeds

A great deal of what is known about desiccation tolerance to date is pieced together from independent analysis of individual genes induced during water deficit stress in resurrection plants together with inference from established seed development biology. Large scale transcript profiling, however, has mainly been limited to the responses of *A. thaliana*, barley and rice to abiotic stresses, including dehydration, high salinity and cold. These studies have revealed a significant overlap in the groups of genes activated in response to abiotic stress and ABA treatment (Seki et al., 2002; Kreps et al.,. 2002; Oztur et al., 2002; Dubouzet et al., 2003; Rabbani et al., 2003; Shinozaki et al., 2003).

ABA-signaling has been implicated as a mediator of changes in water availability, particularly desiccation stress, in vegetative tissues of both desiccation sensitive and tolerant plants and in the development of seed desiccation tolerance. Functional analysis of ABA-regulated gene promoters has identified the motifs involved in the ABA response. The most well known ABA responsive element (ABRE) is a sequence which resembles a G-Box with an ACGT core (Marcotte et al., 1989; Shen et al., 1993). A second element, the coupling element (CE) is required to form a functional ABA response complex (ABRC). This motif elicits a response to ABA when located adjacent to an ABRE (Shen and Ho, 1995). The RY/Sph motif is involved in seed specific gene expression (Dickinson et al., 1988; Hattori et al., 1992) and like the CE, the RY element also acts synergistically with ABRE (Ezcurra et al., 1999; Vasil et al., 1995). The

bZIP TFs and B3 domain proteins bind to the ABREs and RY elements respectively (Finkelstein et al., 2002).

A number of genes encoding products that regulate gene expression and signal transduction were found to be induced by abiotic stress in *A. thaliana* (Seki et al., 2004). These included a number of stress-inducible transcription factor (TF) families, such as the drought responsive element binding (DREB) proteins, the ethylene-responsive element binding factors (ERF), zinc finger family TFs, WRKY TFs, three members of the MYB family of TFs, two members of the basic helix-loop-helix TF family, four bZIP-containing TFs, five NAC TFs and 3 homeodomain-leucine zipper (HD-ZIP) containing TFs. Investigation of the promoters of genes upregulated by abiotic stress on the desiccation sensitive plants has identified potential cis-acting elements that bind these TFs. The well-known RD29A (responsive to dehydration 29A) promoter contains the ABA-responsive element (ABRE) in addition to a dehydration-responsive element DRE/C-repeat (CRT). The ABRE-binding protein (AREB)/ABRE-binding factor, which contains a bZIP DNA binding domain, recognizes the ABRE-responsive element and activates gene expression in an ABA-dependent manner. DREB1/CRT binding factor (CBF) and DREB2, belong to the AP2 TF family. These, TFs, however are induced by cold and dehydration in both *A. thaliana* and rice and bind to the CRE/CRT repeat A/GCCCGAC independently of ABA (Shinozaki et al., 2003). This would suggest the presence of an alternate ABA-independent pathway to regulate dehydration-induced gene expression which is evidenced by the large number of dehydration-inducible genes that do not respond to ABA treatment. ERD1, encoding a Clp protease regulatory subunit, ClpD is one of these genes and it is induced by dehydration and during senescence (Nakashima et al., 1997). Its promoter contains novel cis-acting elements involved in ABA-independent dehydration stress induction and senescence-activated gene expression (Simpson et al., 2003) and NAC transcription factors have been found to interact with these cis-elements (Tran et al., 2004).

1.5.1. Gene regulation during the desiccation stage of seed maturation.

Initial studies on the desiccation stage of seed maturation showed that high levels of ABA during mid development induced the accumulation of LEA proteins and HSPs to prepare the embryo for

desiccation (Kermode and Finch-Savage, 2002). Further investigation of the regulation of gene expression during maturation drying in *A. thaliana* and maize seeds initially focused on the regulation of the promoters of these LEA proteins and HSPs. These studies together with the subsequent characterization of mutants defective in ABA-signaling or that were developmentally impaired have identified ABI3/VP1, ABI4, ABI5, LEC1 and FUS3 as essential for the regulation of ABA- or seed specific gene expression.

Seeds of the ABA-insensitive mutants, *abi3*, *abi4* and *abi5* are unable to acquire desiccation tolerance and a similar phenotype was identified in maize mutants with a mutation in *VP1* (VIVIPAROUS1), an orthologue of *ABI3*. The *ABI3/VP1*, *ABI4* and *ABI5* genes encode B3, AP2 and bZIP-type TFs, respectively (Finkelstein and Lynch, 2000; Finkelstein et al., 1998; Giraudat et al., 1992; Lopez-Molina and Chua, 2000). Biochemical approaches revealed their target sequences as the RY element for ABI3/VP1, CE1-like sequences for ABI4 and ACGT-containing ABRE for ABI5 (Carles et al., 2002; Kim et al., 2002; Niu et al., 2002; Mönke et al., 2004). Although ABI4 is most closely related to the DREB family of TFs no DRE cis-elements have been identified in the promoters of ABI4-regulated genes. Studies with these mutants suggests that in response to ABA, ABI3 and ABI5 act in concert, via RY and CE motifs and ABREs, to regulate gene expression in the maturating seed (Nakabayashi et al., 2005).

Screening mutants with developmental defects resulted in the characterization of *Lec1* (LEAFY COTYLEDON1) and *fus3* (FUSCA3) mutants. The *LEC1* genes encodes a TF with homology to the HAP3 subunit of CCAAT binding factors while *FUS3* was also found to be a member of the B3 domain family of TFs and was shown to be able to bind to the RY motif *in vitro* (Finkelstein et al., 2002). *LEA* (*AtEM1* and *AtEM6*) and *HSP* (*HSP17*) genes are upregulated during the desiccation phase of seed maturation. The promoters of these genes contain ABA responsive elements and their expression has also been shown to be dependent on ABI3, ABI4, ABI5, LEC1 and FUS3, originally identified in the mutant studies described above (Wehmeyer and Vierling, 2000; Carles et al., 2002).

These studies have been extended to characterize gene expression studies during seed development in *A. thaliana* microarray analyses (Girke et al., 2000; Ruuska et al., 2002). Initially only the early stages of seed development were considered but recent studies to investigate the establishment and subsequent release of seed dormancy also considered gene expression during maturation drying (Cadman et al., 2006). A genome-wide analysis of *A. thaliana* seeds (accession Cvi) showed that ABA-response elements are overrepresented in genes with high transcript numbers in mature dry seeds. This agreed with previous observations of Nakabayashi et al (2005) who showed a high occurrence of ABREs in genes with high mRNA transcript levels in dry seeds but not in imbibed seeds of *A. thaliana* (accession col). This corresponds with the reduced endogenous levels of ABA in the imbibed seeds. It is believed that ABA accumulates in seeds through the regulation of 9-cis-epoxycarotenoid dioxygenase (*NCED*) gene family which encodes carotenoid cleavage enzymes that complete the first step in ABA synthesis (Nambara and Marion-Poll, 2005). Two members of this family, *NCED6* and *NCED9* were found to be expressed at high levels in mature seeds (Cadman et al., 2006). Collectively these results support an overlap between ABA, stress and maturation drying at the level of the transcriptome as previously suggested by Hilhorst (1995). Genes encoding proteins involved in signal transduction including a number of AP2 domain-containing TFs, WRKY TFs, CBF, MYB TFs, protein phosphatases, bZIP TFs, Heat Shock TF, AREBs, DREBs and basic helix-loop-helix TFs were also found to have increased levels of transcript in mature dry seeds (Cadman et al., 2006).

1.5.2. Gene regulation in desiccation tolerant plants in response to desiccation

Compared to the work on the desiccation phase of seed maturation and the response of desiccation sensitive plants to abiotic stresses, very little is known about the molecular mechanisms regulating gene expression in response to water deficit in resurrection plants. ABA signaling also appears to have a role in the regulation of gene expression in response to water deficit as ABA treatment was able to illicit responses similar to those observed during water loss in the dicotyledenous resurrection plant, *C. plantagineum*. ABA levels increase during dehydration and many dehydration-induced genes can also be induced by ABA and contain ABA response elements in their promoters (Bartels and Salamini, 2001).

Although adult *M. trunculata* are not true desiccation tolerant plants, incubation of seedlings in polyethylene glycol is able to re-induce desiccation tolerance in the sensitive radicles of germinated seedlings (after loss of the desiccation tolerant stage of the seed). The expression of 16086 *M. truncatula* genes were monitored in an attempt to identify genes involved in the re-induction of desiccation tolerance and additionally those involved in maturation drying in seeds (Buitink et al., 2006). A number of TFs including the DREB family members, the AP2-like EREBP (ethylene response binding protein) and AREB-like proteins were shown to be upregulated in PEG-treated radicles (Buitink et al., 2006). A comparison of the regulatory genes that overlapped in the re- establishment of desiccation tolerance and during maturation drying in seeds included a DREB family member, two MYB family members and an ERF (ethylene response factor) showing that these regulatory factors are involved in the developmentally regulated maturation drying in seeds as well as the re-establishment of desiccation tolerance in the radicles of this plant (Buitink et al., 2006).

The results thus far certainly imply common regulatory mechanisms in response to water deficit in both seeds and desiccation tolerant plants and during mild water loss in desiccation sensitive species. This shows an important regulatory mechanism in response to water loss appears to have been conserved through evolution. A more detailed analysis of the overlap between desiccation tolerant plants and seeds will reveal the degree of conservation of regulatory mechanisms in response to severe water deficit.

The observations to date, most of which has resulted from independent experiments on groups of or individual genes, have shown a significant overlap in induced genes and regulation of their expression particularly in response to severe water loss in desiccation tolerant plants and seeds. More recently, these results have been confirmed and our knowledge of responses to water loss expanded through large scale gene expression analysis. The majority of these transcriptomic studies have focused on desiccation sensitive species such as *A. thaliana* and recently attempts have been undertaken to investigate the responses of non-model plants to water deficit.

1.5.3. Transcriptomics of abiotic stress and seed development

The body of literature presented here points towards an overlap between the developmentally regulated desiccation tolerance program in orthodox seed development and that of the environmentally induced program in the vegetative tissues of desiccation tolerant plants. While this overlap is highlighted by the accumulation of sucrose, certain antioxidants and LEA proteins and regulation of gene expression through ABA pathways in both mature dry seeds and desiccated vegetative tissue, a direct comparison between gene expression in dry seeds and desiccation tolerant vegetative tissues is lacking.

Until recently much of what was known about desiccation tolerance was pieced together from independent analysis of individual genes upregulated during water deficit stress in resurrection plants together with inference from established seed development biology. More recently, large scale gene expression analyses of dehydrating *A. thaliana* and rice have identified a number of stress-inducible transcripts encoding proteins that function in abiotic stress tolerance. These include molecules such as chaperones, late embryogenesis abundant (LEA) proteins, osmotin, antifreeze proteins, mRNA-binding proteins, key enzymes for osmolyte biosynthesis, water channel proteins, sugar and proline transporters, detoxification enzymes, various proteases, transcription factors and phosphatases (Seki et al., 2002; Shinozaki et al., 2003; Rabbanni et al., 2003). Similarly, transcript profiling of mature dry seeds have revealed genes with highly abundant transcripts which encode proteins involved in storage protein synthesis, lipid metabolism, carbohydrate metabolism, transcriptional regulation, signal transduction, development, photosynthesis and abiotic stress response (Girke et al., 2000). The stress related genes included LEAs (including group 3 and 5 LEAs and Em1 and Em6), peroxiredoxins and sHSPs (Cadman et al., 2006). Additionally, the translation of many of these gene products have been confirmed by proteomic analysis in seed (Gallardo et al., 2002; Rajjou et al., 2004; Chibani et al., 2006).

While the vegetative tissues of these model plants are unable to survive extreme desiccation, desiccation tolerant tissues have to pass through hydration levels comparable to the milder

dehydration stresses tolerated by these sensitive species and there may be pathways that are similar between desiccation and dehydration tolerance. Profiling the response of desiccation-sensitive plants is therefore useful to understand the molecular features controlling the plant responses to water deficit and responses to mild water loss but they cannot provide sufficient insight into those additional traits affording resurrection plants the ability to tolerate extreme water loss. A more practical approach to understanding desiccation tolerance, therefore, is the study of non-model desiccation tolerant plants. In a recent study a library of 10 900 cDNAs was prepared from the resurrection plant, *X. humilis*, undergoing a cycle of desiccation and rehydration and an initial small scale microarray screen of this library identified 55 desiccation upregulated cDNAs and 79 downregulated (Collett et al., 2004). Genes related to photosynthesis and metabolism were downregulated while dehydration-upregulated cDNAs included previously identified stress-responsive genes encoding metallothioneins, galactinol synthases, an aldose reductase, a glyoxalase-I like protein, late embryonic abundant proteins (LEAs), dehydrins and desiccation-related proteins. Additionally, a number of genes not previously associated with desiccation responses were identified and included a putative chloroplast RNA-binding protein and a protein containing SNF2/helicase domains (Collett et al., 2004). Additional large scale analysis of this valuable genetic resource will allow further identification of desiccation related genes geared towards elucidation of the mechanisms involved in desiccation tolerance in this plant.

In another study, the expression of 16086 *M. truncatula* genes were monitored in an attempt to identify genes involved in the re-induction of desiccation tolerance in the sensitive radicles of germinated seeds incubated in polyethylene glycol and which, additionally, may be involved in the developmentally regulated acquisition of desiccation tolerance in maturing seeds (Buitink et al., 2006). Expression profiling revealed differential expression of more than 1300 genes and several clusters were identified as a function of time-dependant changes in expression. Genes belonging to the early response group were found to encode proteins known to be involved in initial protection and adaptation responses. Expression patterns at later time points were comparable with those involved in late seed maturation as a large number of genes related to

storage reserves were upregulated and those associated with active metabolism were downregulated. It thus appears that the re-establishment of desiccation tolerance in the germinated radicals is equivalent to the quiescent state prior to germination. This study also showed the rapid accumulation of sucrose with the re-induction of desiccation tolerance again indicating the significance of this sugar in desiccation tolerance (Buitink et al., 2006). An approach to identify proteins involved in desiccation tolerance involved the analysis of heat-stable proteins in *M. truncatula* seeds. Comparative analysis of imbibed seeds prior to loss of desiccation tolerance with desiccation-sensitive germinated seeds that could re-impose desiccation tolerance and those that could not, identified 15 polypeptides with a definite role in desiccation tolerance. Among these were 6 Group 1 LEAs, one Group 2 LEAs, three Group 3 LEAs and one Group 5 LEA. Expression of the genes for all these (except for the LEA 2) were shown to be seed-specific (Boudet et al., 2006).

1.6. Conclusion

Genetic and evolutionary evidence suggests that desiccation tolerance is a primitive characteristic that was lost as plants evolved mechanisms to prevent desiccation but was retained in seeds and spores. Comparisons of desiccation-induced genes in resurrection angiosperms show that they have counterparts in non-tolerant plant species which are expressed during early dehydration or are abundantly present during the seed maturation phase as the embryo acquires desiccation tolerance (Bartels, 2005, Illing et al., 2005). This suggests that all plants may have the genetic potential to tolerate desiccation but that relevant genes differ in their spatial and temporal expression patterns due to changes in regulatory control (Philips et al., 2002; Bartels, 2005).

An investigation of the genes expressed during seed desiccation and the re-induction of desiccation tolerance in *M. trunculata* radicles has revealed the existence of responses common to both conditions. However, this is not a truly desiccation tolerant system and there may still be other mechanisms employed by the extremely desiccation tolerant tissues of the resurrection plant, *X. humilis* also present in seed that were not identified. Similarly, analysis of gene

expression in dehydrating plants which are unable to survive extreme desiccation (such as *A. thaliana*) and their mature dry seeds have revealed the induction of signaling and protective factors related to abiotic stress in both tissues and suggests an overlap in pathways at the initial stages of water loss. Thus, while this approach has shown the overlap in the induction and regulation of genes during early desiccation, mechanisms induced during more extreme water loss will be excluded.

Therefore, we reasoned that a more directed approach to compare the gene expression in the desiccated vegetative tissue and mature dry seed of *X. humilis* with that in dehydrated vegetative tissue and mature dry seed of a desiccation sensitive species (*A. thaliana*) would allow identification of pathways common only to desiccated vegetative tissue of angiosperms and seed. This would also allow the identification of novel pathways common to extreme desiccation and potentially show a more significant overlap between the desiccated vegetative tissues of *X. humilis* with seed than that for *A. thaliana.* This would provide support for the hypothesis of the activation of seed-specific genes for the acquisition of vegetative desiccation tolerance in *X. humilis*. Additionally it will allow a large scale view of desiccation tolerance in *X. humilis* to investigate the range of mechanisms employed and potentially to discover new mechanisms as well as regulatory mechanisms within this plant.

The study presented here aims to test the hypothesis that survival of extreme water loss in the leaves and roots of the resurrection plant X. *humilis* (Farrant et al., 1999) evolved from the activation of genes that confer desiccation tolerance on seeds in vegetative tissue. Here we show, using large scale expression profiling, for the first time an extensive overlap between seed-specific gene expression and gene expression in desiccated leaf and root of *X. humilis*. Conversely there is little overlap between gene expression profiles of desiccated vegetative tissue and seeds of the desiccation-sensitive *A. thaliana*. These results support previous studies which show the activation of a limited number of seed-specific desiccation tolerance mechanisms in the vegetative tissues of desiccation tolerant plants (Ingram and Bartels, 1996; Ghasempour et al., 1998; Scott, 2000; Illing et al., 2005) and strengthens the hypothesis that a vegetative desiccation

tolerant mechanism present in resurrection angiosperms evolved from the activation of genes involved in DT during seed development, facilitating the survival of angiosperms in arid niches (Oliver et al., 2000).

Chapter 2

Microarray Data Exploration and Normalisation

2.1. Introduction

In this study we hypothesised that if seed-specific genes are activated in desiccated leaves and roots to confer desiccation tolerance, there should be a significant overlap between genes expressed in desiccated leaves, roots and seeds in *X. humilis*. Microarrays were employed to test this theory as the expression levels of large numbers of genes can be monitored under different conditions simultaneously and genes that are differentially expressed between these conditions can be identified (Schena et al., 1995). Cluster analysis can then group the genes according to similarities in their transcriptional profiles identified from the microarray data to reveal groups of genes with coordinated regulation (Eisen et al., 1998). This method therefore allows a comprehensive overview of the transcriptional response of a cell to the condition tested. We used a microarray printed with 3105 *X. humilis* cDNAs to investigate the relative mRNA transcript abundance of these *X. humilis* genes in desiccated leaf (DsL), desiccated root (DsR) and mature seed compared to hydrated leaf (HL) and root (HR). This was performed to find genes differentially expressed between desiccated and hydrated vegetative tissue and between seed and hydrated vegetative tissue and to identify those highly expressed in desiccated vegetative tissues and mature seed. Additionally cluster analysis was conducted to find genes with similar expression patterns and to determine the overlap in gene expression between the tissues.

Desiccation tolerant plants have to pass through hydration levels comparable to the milder dehydration stresses that can be tolerated by sensitive species as a result there will be similar responses in both desiccation tolerant and sensitive species at these earlier stages (Seki et al., 2002; Shinozaki et al., 2003; Rabbanni et al., 2003). Thus, it was decided to only investigate gene expression in extremely desiccated vegetative tissue and mature seed to identify mechanisms that are specific to extreme desiccation.

As a comparison, we performed a similar analysis on publicly available gene expression data sets from *A. thaliana* in which the seeds are desiccation tolerant but the vegetative tissues are desiccation-sensitive. The relative abundance of genes in dehydrated leaf (DhL), dehydrated root (DhR), mannitol treated leaf (OsL), mannitol treated root (OsR) and mature seed were to be compared to control leaf (CL) and control root (CR). This comparison was included to find genes that are highly expressed in water-stressed vegetative tissue and mature seed of a desiccation-sensitive plant. These genes were compared with those expressed in desiccated vegetative tissue and seed of *X. humilis* to identify those that are involved in extreme water deficit as opposed to the mild water loss tolerated by *A. thaliana* and to identify common responses in both species. The similarity in gene expression patterns between stressed vegetative tissue and seed was also determined and compared with that in the desiccation-tolerant *X. humilis*.

These *A. thaliana* expression data sets were chosen for this comparison as the raw microarray data is readily accessible and has been shown to be reliable and highly reproducible. In addition plant growth conditions, stress application, harvested tissues and the type of microarray used has been standardised for this set of expression data (Kilian et al., 2007). The expression data from both the *A. thaliana* dehydration (Dh) and osmotic stress (Os) treatments were compared with the expression data from the *X. humilis* desiccation treatment. The expression data from mature seed from the *A. thaliana* developmental series (Schmid et al., 2005) was compared to the expression data from the *X. humilis* mature seed.

Although the *A. thaliana* dehydration treatment was mild, resulting in a 10% loss of fresh weight (Kilian et al., 2007), this can be considered to be a significant stress for a desiccation-sensitive plant. Unfortunately, one of the drawbacks of this data set is that the Relative Water Content (RWC) values of the tissues were not reported for the stress treatments and as a result the severity of the stress imposed at the cellular level is difficult to assess. RWC is an important measure of plant water status which describes the actual content of water in the plant based on the maximal water content it can hold at full turgidity (Smart and Bingham, 1974). Water potential, which refers to the thermodynamic properties of water, has also been suggested as a

measure of plant water status. RWC, however, has been shown to have a closer correlation than water potential with various physiological and biochemical activities including photosynthesis, protein synthesis, leaf senescence as examples (reviewed in Huang, 2006). Under conditions of dehydration stress many plant species are able to retain water and maintain cellular turgor through osmotic adjustment. The osmotic adjustment in leaf tissue may lower the water potential at which a particular RWC is reached. As a result RWC, which is related to physiological activities and accounts for osmotic adjustment has been suggested to be a better indicator of water status and physiological activity particularly under stresses which cause cellular water deficit (reviewed in Huang, 2006).

As there were no RWC we looked at the study of Kilian et al. (2007) to identify the time point after of the dehydration stress that caused the biggest change in gene expression in response to this stress. We also included the expression data from the osmotic stress treatment as osmotic stress arises as a consequence of the loss of water from the cell. We therefore reasoned that a similar response to that observed for severe dehydration may occur after prolonged exposure to hyperosmotic conditions. Again the RWC values were not reported for this stress and so the study of Kilian et al. (2007) was consulted to identify the time point that resulted in the biggest change in gene expression.

Although the *A. thaliana* stress treatment data is reported as a time series experiment, the expression data from each time point is available for individual analysis. We chose the 1 hour time point for the dehydration stress and 24 hour time point for the osmotic stress as these points showed the biggest change in gene expression (Kilian et al., 2007). The expression data before the commencement of the stress was included as the hydrated or control samples. The expression data for the mature seed sample was retrieved from the “seed to silique” developmental series data set (Schmid et al., 2005).

However, before the question at hand can be addressed, basic issues related to microarray data, including the potential limitations of their use must be considered. Microarray experiments are

divided into a number of steps or phases (Fig. 2.1) which begins with the initial analysis comprising data generation, capture, pre-processing and normalisation (the focus of this chapter).

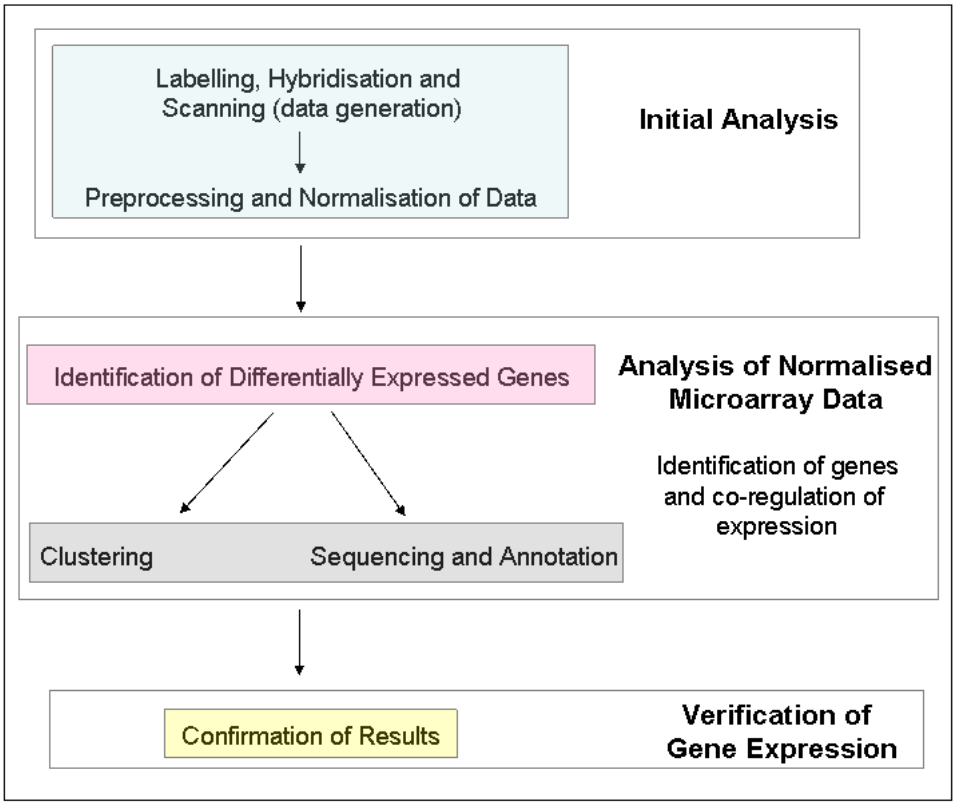


Figure 2.1: Flow diagram showing the main stages of a microarray experiment. Data is generated by labelling and hybridisation of samples to microarray slides. The fluorescence data is captured (scanned), processed (e.g. filtering to remove "bad" or "absent" spots) and normalised. The normalised data set is then ready for further processing including identification of differentially expressed genes and clustering of co-regulated genes. Functional interpretation of genes allows identification of pathways involved in the response to the condition tested. Sequence and annotation information is available for model organisms, like *A. thaliana*, but in the case of non-model organisms the genes must be sequenced first and then functionally annotated by comparison with data available for model organisms.

Higher end analysis involves the identification of differentially expressed genes and clustering and functional annotation (detailed in chapter 3) and finally, confirmation of microarray results (detailed in chapter 4). At each stage of a microarray experiment careful consideration must be given to factors which could negatively impact the final results. These are discussed below.

2.1.1. Microarrays

Microarray is a technology that allows for the quantification of expression levels of thousands to tens of thousands of genes under different conditions simultaneously (Schena et al., 1995). In this technique the probes are arrayed on an inert substrate (such as glass or silicone) and levels of gene expression in a biological sample are measured. RNA is extracted from tissues of interest, converted to cDNA and labelled with a fluorescent dye. The labelled cDNA is hybridized to the array where the cDNAs hybridize to complementary gene-specific probes and non-hybridised material is removed by washing. Images are generated by confocal laser scanning and the relative fluorescence intensity of each gene-specific probe is a measure of the level of expression of the particular gene. Microarray data can be generated by either a dual channel or a single channel approach. In the first case two samples of RNA are labelled with different dyes, and simultaneously hybridized to the array. The ratio between the two dyes indicates the relative abundance of a gene in the two samples tested. In single channel arrays, each sample is labelled and individually incubated on an array. The fluorescence value is a measure of the gene expression levels within that particular sample (Quackenbush, 2002). Two platforms exist for the preparation of microarray slides/chips and these include robotic spotting (Schena et al., 1995) and in situ synthesis by photolithography (Lipshutz et al., 1999). Generally cDNA clones and oligonucleotides (50 – 70mers) are deposited robotically while the photolithographically synthesized arrays (chips) contain shorter oligonucleotides (25mers) and each gene is represented by multiple oligonucleotides. The robotically printed slides can be used for both single and dual channel arrays whereas the oligo chips are only used for single channel arrays.

2.1.2. Experimental design

Microarrays provide information on the expression of thousands of genes but data from such high throughput experiments is known to be obscured by significant amount of variation. Random and systematic variations obscure the actual mRNA expression levels within the measured signal and must be removed to generate reliable high quality data (Leung et al., 2003). Random variation is mainly a consequence of measurement error and statistical tools are equipped to handle this. Systematic variation, however, is a dangerous bias that can be

introduced at many stages during a microarray experiment and there are many potential variables which must be addressed before the commencement of an experiment. Sources of variation in a typical microarray experiment arise from treatment of the biological samples, RNA extraction, the microarray slide production, the hybridization process (dye labelling and hybridization) and scanning. These variables must be addressed and a good statistical design aims to minimize the introduction of such bias. A sound experimental approach should consider a number of factors, including a clear hypothesis or biological question; careful planning of the treatments tested or samples collected; microarray experimental protocols free of systematic and experimental errors and, the statistical analysis of the data including the normalization strategy and comparisons of interest (Churchill, 2002; Yang and Speed, 2002).

2.1.3. Replication

Replication can be used to decrease the uncertainty introduced into an experiment by random and systematic variations and to establish the significance of these results (Kerr and Churchill, 2001). There are two types of replication namely technical and biological. Technical replication involves the application of the same RNA sample to different arrays and is used to measure the variation as a result of differences in RNA preparation, labelling, array irregularities, hybridization, washing, scanning and image analysis. Due to the expense of microarray slides the use of technical replicates at the cost of a new slide is not generally recommended but several single-channel systems such as Affymetrix GeneChip come with built in technical replicates. Unwanted effects such as spatial variation or spot inconsistencies can be averaged out with these replicates. In cDNA arrays the spots are often printed in replicates across each slide and although not independent replicates, they give an indication about the quality of hybridization. Additionally, in cDNA arrays a dye swap experiment provides a measure of bias in the labelling and hybridization of each dye. Biological replication involves the hybridization of RNA from independent sources to separate arrays (Wit and McClure, 2004). This allows quantification of the variation from the actual biological system being studied. In general multiple independent biological conditions should be included. Both technical and biological replication is required for statistically meaningful inference from the samples tested (Kerr and Churchill, 2001).

It is often necessary to pool RNA from several individuals within the population to generate enough to hybridize to a single microarray slide. Pooling RNA reduces the bias that each member of a population introduces and is recommended in cases where there are only a few slides available and the main focus is the mean expression levels within a population and not the variation of the expression within that population. This approach is a cheap way to maximize the accuracy of results and has been recommended for single channel analysis as pooling can aid inferential precision when the number of slides is limited (Churchill, 2002; Churchill and Oliver, 2001; Kendziorski et al., 2003). The inclusion of a number of pooled biological replicates still allows the estimation of biological variance (Wit and McClure, 2004). It must be remembered however that pooling samples will result in the loss of information from individual samples and some of the algorithms in the downstream analysis, may require the biological noise as a parameter.

The decision about the number of replicate slides to hybridise can be problematic as it is based on the estimation of the variance in signal intensity. This is a difficult value to estimate as the variance for each probe is dependent on sequence, expression level and tissue source. In general, a minimum of 3 to 5 replicates have been suggested but a larger number are recommended to increase statistical power (Lee et al., 2000; Pavlidis et al., 2003). However, in many cases so few replicates are performed because of the cost of the microarray slides and limitations in availability of RNA from different biological replicates. While these are valid experimental constraints it is still important to consider the amount of noise present in the system and the amount of certainty needed for the conclusion (Wit and McClure, 2004). Mathematical modelling has derived that a minimum of three microarray replicates suffice for most forms of inference (Lee et al., 2000) and while this has been widely accepted (Yang and Speed, 2002; Wit and McClure, 2004) it remains that increasing the sample size naturally reduces the amount of error introduced.

2.1.4. Controls

An important part of experimental design is the consideration of possible normalization strategies. Control spots should be included as part of the design to assist with normalization, particularly in the case of arrays with smaller numbers of genes or "boutique" arrays. These are generally synthetic genes (for example lucidea spikes), housekeeping genes or a set of genes whose expression levels do not change across the conditions tested. In some cases this may not be possible and a set of genes whose expression levels have been tested and verified beforehand can be included instead. These controls aim to provide a reference for the validation and normalization of the data. There should be enough control spots to allow statistically robust reference, they should preferably span the entire intensity range and have a random distribution across the slide (Drãghici, 2003).

2.1.5. Normalization

Differences in labelling, hybridization, and detection methods are difficult to prevent and may still introduce bias. These errors can be detected and identified by visualization and exploration of the raw data after collection and this assists with the choice of a normalization strategy to eliminate systematic artefacts from the data while preserving the biological data. The normalization strategy chosen must correspond with the nature of the specific bias detected. Local or Spatial bias affects subsets of data within an array and is introduced as a consequence of differences between microtitre plates or print tips, hybridization artefacts, slide surface and scanner inconsistencies (Yang et al., 2002; Stekel, 2003; Park et al., 2003). Global bias which affects the spread of data between arrays is usually the result of differences in scanner settings used to scan each array (Park et al., 2003), differences in mRNA concentrations from each sample or differences in the labelling efficiencies of the samples. Both biases require specific methods of normalization which must be applied carefully as they can have profound effects on the final data (Quackenbush, 2002; 2006).

This process of normalisation is conducted in a stepwise manner; firstly within-array normalization corrects local artefacts within an individual array before progressing to

normalization across all the arrays (Wit and McClure, 2004). Within-array normalization involves spatial correction, background correction and intensity dependent dye bias (for dual channel analysis) correcting biases as a result of uneven spotting and hybridization and differences in dye intensities. It can be performed locally (at pin-tip or block level) or globally (across the entire slide). Local normalization is generally applicable when inconsistencies are detected in specific areas of an array, such as those caused by differences in the printing pins, surface variation within slide, and differences in hybridization across the slide (Yang et al., 2002; Stekel, 2003; Park et al., 2003). A number of loess-based methods exist for correction of location spatial (location) bias including print-tip loess and 2D loess (Smyth and Speed, 2003; Stekel, 2003). An alternative spatial correction method, which is based on the median of log2 values of spots within a predefined area, also effectively smoothes any spatial trends affecting the data (Khojasteh et al., 2005). In dual channel analysis dye intensity bias is a result of the differences in the properties of the two dyes used. Simple linear regression was originally employed to remove this bias but as intensity-dependent effects are non-linear (Stekel, 2003) more robust non-linear methods have since been developed. These include loess regression based methods (Cleveland and Devlin, 1998; Yang et al., 2002; Drãghici, 2003; Stekel, 2003) and equivalent nonparametric, spline-based methods (Rupert et al., 2003).

Both single- and dual-channel arrays are affected by scale bias and methods to correct this systematic error are generally applicable to both platforms. Across-array normalizations which involves within-replicate scaling and between-condition scaling, accounts for scale differences in dye intensity across the slides and ensures the data is on the same scale to facilitate comparison across conditions (Wit and McClure, 2004). Several methods of normalization are available for each type of array and new methods are continually being developed. Most were originally developed for dual channel arrays but are also applicable to single channel data. These methods aim to centre the means of the data sets so they are equal and to adjust or scale the values to result in equivalent and thus comparable distributions of data. The simplest of these approaches is to adjust the means of all distributions to zero by subtracting either the $\log_2$ mean or median of a distribution from each feature's $\log_2$ values which results in the mean of all distributions being

equal to zero. This can also be extended by dividing by the standard deviation to bring all the standard deviations of the distributions to one (Stekel, 2003). More sophisticated approaches, including quantile normalisation, cyclic loess and Qsplines, aim to force the distribution of values in each array to be the same or to adjust scale differences between arrays (Bolstad et al., 2003; Smyth and Speed, 2003) and ANOVA based methods are able to distinguish between interesting and random variations after modelling systematic variation (Drãghici, 2003).

The choice of normalization is based on the nature of the data set itself. This is important as two main assumptions underlying many normalization methods (such as global loess) is that the majority of genes are not differentially regulated and that the number of genes upregulated genes is roughly equivalent to the number of downregulated genes (Quackenbush, 2002; Wang et al., 2005). Additionally the choice of normalization algorithms depends on the distribution of expression values. Parametric based methods, such as loess, are applicable to data sets with similar distributions, however non-parametric based methods, like quantile normalisation, are employed if this is assumption does not hold (Drãghici, 2003).

2.1.6. MIAME

In order to make microarray data usable to other researchers the Microarray Gene Expression Data Society (MGED) has developed criteria for reporting data that are known as “minimal information about a microarray experiment” (MIAME) (Brazma et al., 2001). It provides a standard annotation of the core information from most microarray experiments including; experimental design, sample extraction and labeling, hybridization procedures and parameters, raw and processed measurement data and array design. This standard information about microarray experiments should be made available through public repositories for microarray data such as BASE (BioArray Software Environment). This database is a free web based resource that allows the storage of large volumes of microarray data following the MIAME regulations (Lao et al., 2002).

2.2. Materials and methods

Plant treatments and seeds

X. humilis plants of were collected from Barakalalo National Park (Limpopo Province, South Africa). Plants of similar age and size were maintained in trays in a glasshouse and were watered once a week. Three trays of plants with four to five plants in each were allowed to dry naturally by withholding water for a total of three weeks until fully desiccated (~5% relative water content (RWC)) for a total of three weeks and an additional three trays of plants were maintained in a hydrated state (~90% RWC) with watering three times a week. Desiccated leaf and root (main root plus lateral roots) samples and hydrated leaf and root samples from each of the three to four plants in each tray were harvested simultaneously and flash frozen and stored at -80^{o}C. Three leaves and three roots were removed from each plant and pooled as a biological sample. After three weeks without water, three leaves and three roots were removed from each of the plants to calculate the RWC. The samples were weighed to obtain the fresh weight and then placed in a beaker of water for 24 hours and weighed to obtain the weight at full turgor. Subsequently the samples were placed at 80°C for 24 hours and weighed to obtain the dry weight. The RWC was determined by subtracting average dry weight from the average fresh weight, which is divided by the average fresh weight subtracted from the average weight at full turgor, multiplied by 100. Mature seeds were collected over several flowering intervals from different trays of *X. humilis* plants.

RNA extractions

Vegetative tissue

A total of 1 g of *X. humilis* leaf and root tissue each were homogenized to a powder in liquid nitrogen using a mortar pestle and thawed in 5M Na-perchlorate, 0.3M Tris-HCl (pH 8.3), 1% [v/v] SDS, 8.5% [w/v] PVPP, 2% [v/v] PEG-6000, 1% [v/v] β-ME. The samples were incubated at room temperature with shaking for 30 mins, and centrifuged at 10 000 x g for 10 mins at 4^{o}C. The supernatant was extracted with two rounds of 24:1:1 phenol:chloroform:isoamylalcohol (PCI) followed by an overnight isopropanol precipitation at -20^{o}C. The precipitated RNA was recovered by centrifugation at 10 000 x g for 10 mins at 4^{o}C, and the pellet resuspended in 1ml

dH_2O at 65^oC for 5 mins. The RNA was selectively precipitated overnight at 4^oC with 2.5M LiCl and recovered by centrifugation at 12 000 x g at 4^oC for 30 mins.

Seed tissue

Seeds were collected and allowed to dry at room temperature for one month. After this they were stored in eppendorf tubes at room temperature. Approximately 200 mg mature, dried, *X. humilis* seeds were homogenized to a powder in liquid nitrogen together with 700 µl XT buffer (0.2 M sodium borate decahydrate, 30 mM EGTA, 1% SDS, 1% sodium deoxycholate, 2% PVP, 10 mM DTT, 1% IGEPAL, pH 9.0) in a mortar and pestle. The powder was decanted to a 15 ml falcon tube, thawed at room temperature and an additional 700 µl XT buffer added. The liquid was divided into two eppendorf tubes, 56 µl 2M KCL added to each tube, mixed and incubated on ice for 60 min. The supernatant was recovered by centrifugation at 13 000 g for 20 min at 4ºC and the RNA selectively precipitated by the addition of LiCl to a final concentration of 2.5M and incubation at -20ºC overnight. The RNA was collected by centrifugation at 13 000 g for 20 min at 4ºC and the pellets washed in 70% ethanol and resuspended in 100 µl RNase-free water. The RNA was subsequently purified following the RNeasy kit protocol (Qiagen, Germany).

Amplification of cDNA Inserts and Microarray Slide Preparation

A total of 3,105 clones derived from the desiccation stress libraries from *X. humilis* described previously (Collett et al., 2004) were used for the array printing. Target DNAs for spotting on the microarray were amplified by PCR using the T3 and T7 promoter primers flanking the cDNA inserts. 1 µl overnight bacterial culture was added to a 100 µl PCR mixture containing a total of 200 mM of each nucleotide, 0.3 mM of each primer, 2.5 mM $MgCl_2$, 1 X Super-Therm reaction buffer, and 0.5 units of SuperTherm Taq polymerase (Supertherm, Southern Cross Biotechnology). Inserts were amplified by PCR as follows: 94^oC for 1 min; for 10 cycles of 94^oC for 20 sec, 56^oC for 20 sec, 72^oC for 2 min; for 30 cycles of 94^oC for 20 sec, 51^oC for 20 sec, 72^oC for 2 min; at 72^oC for 7 min and at 4^oC for 10 min. PCR products were purified using Multiscreen PCR plates (Millipore, Billerica, MA) according to the manufacturer's specifications, resuspended in 50 µl of 50% DMSO in Nunc 96 well V bottomed plates

(Amersham, Germany). Three microlitres of PCR products were electrophoresed on a 0.8% agarose gel containing 100 ng/ml ethidium bromide to confirm amplification quality and quantity before and after purification. PCR products were printed on Corning GAPS II coated slides (Corning, Netherlands) using a MicrogGrid II Arrayer (Biorobotics, USA). Target DNAs were printed with 16 10μm split pins in a 4x8 configuration and each target DNA was spotted in duplicate randomly. Lucidea ratio and dynamic range controls (Amersham, Germany) were included in the array. The slides were allowed to dry for 10 hours after completion of printing and then crosslinked (900mJ cm^2) and baked at 80^{o}C for 2 hours.

Fluorescent Probe Preparation

RNA samples extracted from *X humilis* leaf, root and seed were assessed by visualisation on denaturing formamide gels and spectrophotometrically with a Nanodrop ND-1000 (Nanodrop, USA). The ratios of $A_{260\ nm}/A_{280\ nm}$ and $A_{260\ nm}/A_{230\ nm}$ were recorded. Fifteen micrograms of total RNA was fluorescently labeled by reverse transcription in the presence 10 mM of 5-(3-aminoallyl)-dUTP (Sigma-Aldrich, St. Louis), 15 nM dTTP, 25 mM each of dATP, dCTP and dGTP and 500ng oligodT primer using Superscript III (Invitrogen) overnight at 42ºC. The RNA was hydrolysed with 500 nM NaOH and 250 nM EDTA at 65°C for 15 mins and the cDNA was purified through QIAquick PCR Purification columns (Qiagen, Germany) according to the manufacturers recommendations with the replacement of wash buffer and elution buffer by phosphate wash buffer (5 mM KPO_4, pH 8.0, 80% ethanol [v/v]) and phosphate elution buffer (4 mM M KPO_4, pH 8.5), respectively. Cy3 monofunctional dye (Amersham, Germany) was linked to the aminomodified nucleotides of the purified cDNA in 0.2 M sodium carbonate buffer (pH 9.0) for 1 hour at room temperature in a separate coupling step to generate fluorescently labeled cDNA. The labelled cDNA was separated from uncoupled Cy3 by purification through QIAquick PCR Purification columns (Qiagen, Germany) following the manufacturers protocol and measured at 260 nm and 550 nm on a Nanodrop ND-1000 (Nanodrop, USA) to quantitate both cDNA concentration and Cy dye incorporation.

Hybridization and Washing

The *X. humilis* cDNA microarrays were hybridized with Cy3 fluorescently labeled probes of desiccated leaf, hydrated leaf, desiccated root, hydrated root and seed. Three biological replicates were examined for all experimental conditions apart from the hydrated root condition. The biological replicates constituted labelled extract prepared from pooled RNA from three leaves or roots from individual plants. For the hydrated root sample, one biological replicate was hybridised to three separate slides. The statistical analysis was based on a total of three replicates per experimental condition (three biological replicates for all conditions except the hydrated root condition which has three technical replicates). The processed microarray slide was prehybridized with 100 µl of hybridization buffer (50% formamide [v/v], 1% BSA [v/v], 0.2% SDS [w/v]) for 2 h at 42°C to reduce background. Slides were washed in dH_20 and dried by centrifugation at 1000 g for 5 mins. A total of 100 µl denatured probe mix (80 pmol labelled cDNA, 25% formamide [v/v], 5X SSC, 0.1% SDS [w/v]) was injected under the lifterslip (Erie Scientific, USA) covering the array and the slide was incubated horizontally in a Telechem hybridisation chamber (Telechem International Inc, USA) at 42°C for 16 h in the dark. Slides were subsequently transferred to a fresh dish and washed in a series of wash buffers preheated to 55°C beginning with 2X SSC, 0.5% SDS [w/v], followed by 0.5.X SSC and finally 0.05X SDS for 5min each with agitation at room temperature. They were dipped in water and then in 100% ethanol and dried by centrifugation at 1000 g for 5 mins.

Scanning and Data Capture

Axon confocal scanning system (Genepix 4000B) and GenePix Pro 6.0 data acquisition software (Molecular Devices Products, USA) were used according to the manufacturer's instructions to capture the *X. humilis* data. Initial normalisation of the Cy3 fluorescent dye emission intensities across all the slides was achieved by adjusting the level of the photomuliplier gains so that there were an equal number of saturated spots on each slide. To extract the expression data from the microarray slide, spots had to be identified and differentiated from background. Gridding determined the location of the centre of each spot on the array. A fixed grid was placed over the array and semi-manual adjustments were performed to finalise this step. As a result the array was

partitioned into areas, each of which contained one spot and some surrounding background pixels. Segmentation was then employed to assign pixels in this area to represent either spot or background and in this case fixed circle segmentation was performed. A circle of fixed size was placed around the centre of the spot. The area inside the circle was used to calculate the intensity of the spot, and the area outside the circle was used to calculate the background associated with the spot. The signal intensities for all 7200 spots were quantified and the mean fluorescence values were extracted. The microarray experiment details and raw data were entered into BASE and are available at http://cbio.uct.ac.za.

Arabidopsis thaliana expression data

The publicly available Arabidopsis affymetrix AtGenExpress stress treatments (Kilian et al., 2007) and the silique to seed developmental series (Schmid et al., 2005) data sets were downloaded from http://affymetrix.arabidopsis.info. The treatments and time points selected are presented in table 2.1. The expression profiles of these conditions were generated with the Affymetrix ATH1 gene chip containing 21 425 unique transcripts and each treatment included two biological repeats except for seed which had three. The osmotic stress was included as the dehydration stress was relatively mild (Kilian et al., 2007).

Table 2.1: Selected stress treatments, developmental series and time points from the A. thaliana affymetrix AtGenExpress database.

Treatment	*Stress Imposed*	*AtGenexpress accession number*
Dehydrated Leaf (DhL)	dehydration 1 hour	NASCARRAY-141
Dehydrated Root (DhR)	dehydration 1 hour	NASCARRAY-141
Control Leaf (CL)	hydrated 1 hour	NASCARRAY-141
Control Root (CR)	hydrated 1 hour	NASCARRAY-141
Osmotic Leaf (OsL)	300 mM mannitol 24 hours	NASCARRAY-139
Osmotic Root (OsR)	300 mM mannitol 24 hours	NASCARRAY-139
Control Leaf (CL)	hydrated 24 hours	NASCARRAY-139
Control Root (CR)	hydrated 24 hours	NASCARRAY-139
Seed	stage 10	NASCARRAY-154

Prior to the dehydration and osmotic stress treatments *A. thaliana* plants were cultivated in rafts in liquid media for 18 days. For the dehydration stress, the plants were exposed to a stream of air for 15 mins during which time they lost 10% of their fresh weight. The plants were placed back in the phytochamber until harvest. For the osmotic stress, mannitol was added to the medium to a final concentration of 300 mM and plants were maintained in this medium in the phytochamber until harvest. In both cases the plants were maintained in under long-day conditions (16 h light/8 h dark) at a light intensity of 150 lmol photons m^{-2} sec^{-1} and a relative humidity of 50% in a standard phytochamber. RNA was extracted from pooled samples in duplicate at various time points including 0 min, 30 min, 1h, 3h, 6h, 12h and 24 h after the onset of the treatment (Kilian et al., 2007). We used the expression data from the 1 hour time point and 24 hour time for the dehydration and osmotic stress treatment, respectively. The mature seeds (stage 10) were collected from 8 week old *A. thaliana* plants grown in soil under long days (16 hour light/8 hour dark). RNA was extracted from pooled samples in triplicate (Schmid et al., 2005). The fluorescence values were extracted from the raw .CEL files downloaded from the above URL. All the spreadsheets presented in the database have been trivially normalized using the Affymetrix standard procedure. A "Scaling Factor" was calculated by removing the top and bottom 2% of signal values, and a value calculated to adjust the mean of the remaining 96% to 100. All the results were multiplied by this factor to give normalised results. This allowed the different experiments to be comparable.

Data Visualisation

Both the *X. humilis* and *A. thaliana* expression sets were imported into R, a language and environment for statistical computing (R development core team, 2008), to visualise the raw data. Spatial analysis was performed for the *X humilis* data using the plot.spatial function from the sma package in R. Box plots and histograms were plotted to visualise the distribution of expression values across the treatments for the two data sets, using the boxplot function in the graphics package (Chambers et al., 1983; Becker et al., 1988) and the plotDensity function from the affy package in R, respectively (A.2.1, A.2.2 and A.2.3).

Normalisation

Spatial correction of the *X. humilis* data was conducted using the spat.norm function followed by background correction (bkg.norm function) using the "deterministic" approach in smida (Wit and McClure, 2004). The median of the lowest 200 expression values was subtracted from all the expression values and any resulting negative values were replaced by zero. Both the *X. humilis* and *A. thaliana* data sets were normalised using the normalize.quantiles function in the affy package (Bolstad et al., 2003) (A.2.4, A.2.5, A.2.6).

2.3. Results and Discussion

Labelling and Hybridisation of X. humilis Micorarray Slides

The 3501 cDNAs were each printed in duplicate next to each other and cDNAs from the desiccating leaf and desiccating root libraries were printed in each block (Fig. 2.2). As the cDNA library had not been sequenced and was prepared in such a manner as to include mostly genes whose expression levels change during desiccation, it was difficult to choose a control set genes, such as an invariant set or housekeeping genes. It was decided to include a set of 424 cDNAs from a previous microarray experiment. The expression levels of some of these cDNAs have also been previously verified by northern blot analysis (Collett et al., 2004; Illing et al., 2005). Lucidea control spots were also printed in replicate in each block and both sets of controls were included in each block to validate the chosen normalization strategy.

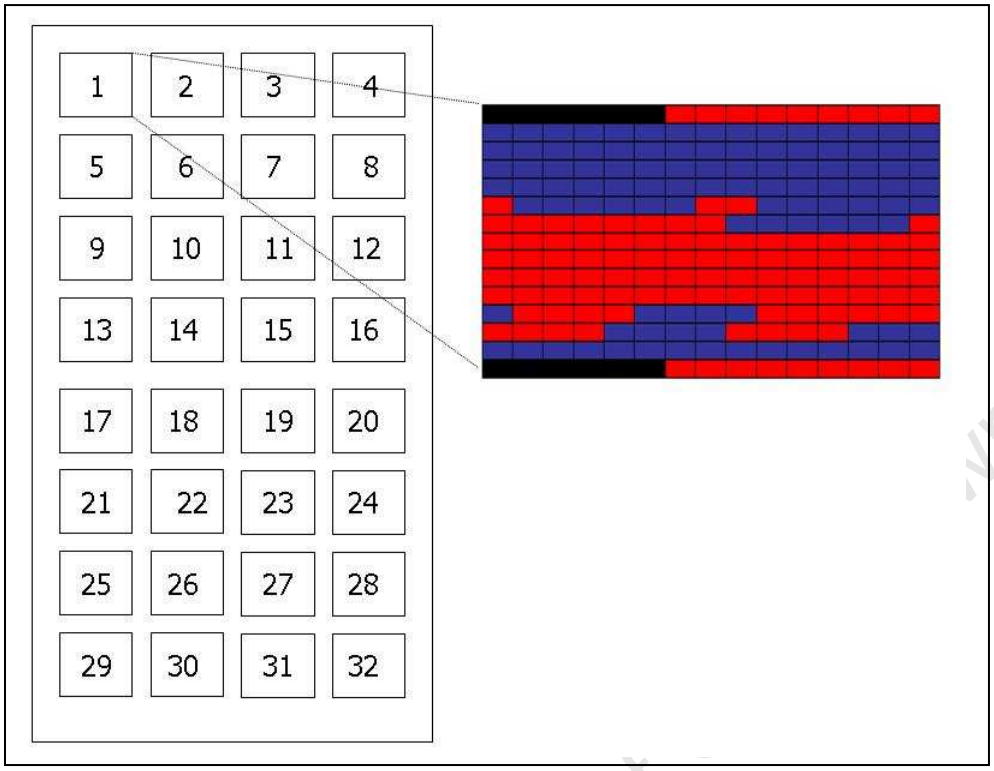


Figure 2.2. Schematic representation of the printed *X. humilis* microarray slides. Black = lucidea controls, Blue= root desiccating library cDNAs, Red= leaf desiccating library.

The most important step in the microarray is good quality RNA. The RNA samples hybridized against the *X. humilis* slides are shown in Fig. 2.3 and the corresponding $A_{260\ nm}/A_{230\ nm}$ and $A_{260\ nm}/A_{280\ nm}$ ratios are in Table 2.2. As can be seen in Fig.2.3 the RNA is intact and of good quality and in all cases, except the seed, the $A_{260\ nm}/A_{230\ nm}$ and $A_{260\ nm}/A_{280\ nm}$ ratios are greater than 2 (Table 2.2) indicating low levels of polysaccharide and protein contamination. It is recommended that RNA samples contain as low levels of protein and polysaccharides as possible as these contaminants can decrease the efficiency of labelling. The lower $A_{260\ nm}/A_{230\ nm}$ ratio for the seed samples is most likely due to the high levels of polysaccharides and storage proteins present in seeds which are difficult to remove during extraction.

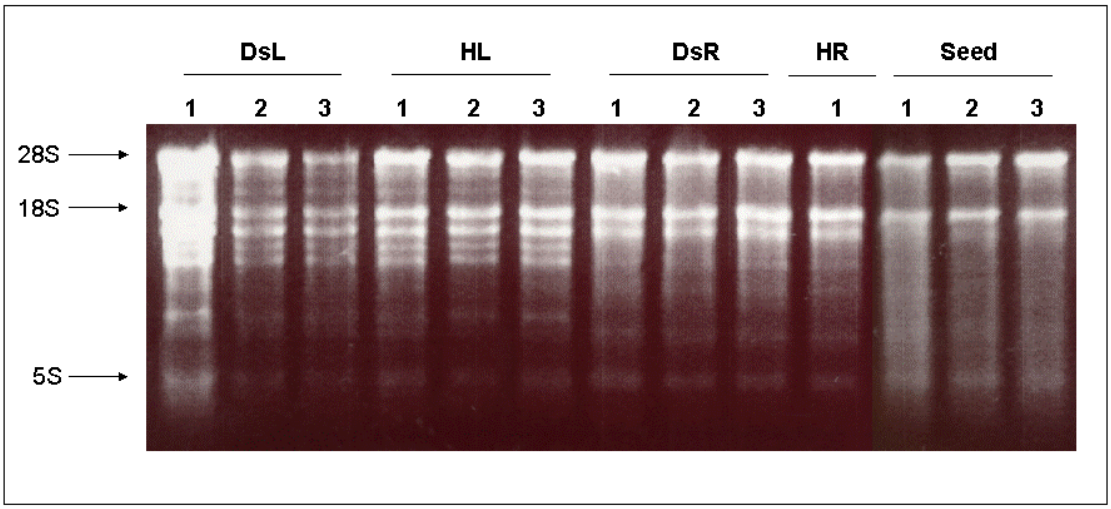


Figure 2.3. 1 µg of total *X. humilis* RNA from each pooled sample used for the microarray analysis. Each biological replicate is indicated below the sample name. Ribosomal sizes are indicated on the left. DsL – Desiccated Leaf, HL – Hydrated Leaf, DsR- Desiccated Root, HR – Hydrated Root.

In all cases the RNA samples were converted to cDNA and coupled to Cy3 and the frequency of incorporation (FOI) and amount of labelled cDNA are reported for each sample in Table 2.2. The $A_{260\ nm}/A_{230\ nm}$ ratio was below 2 for the seed samples and this is also reflected in the lower labelling efficiencies of these samples. Additionally, some of the samples with ratios above two also showed lower efficiency in labelling (e.g. DsL3, HL3, and DsR3). As this appeared to affect all the samples performed together (biological repeat number 3) it is thought to most likely be due either to a problem with the batch of Cy dye used. Alternatively, the coupling buffer should be buffered to pH 9.0 and any deviations from this point could reduce the efficiency of the coupling reaction. Each of the DsL, HL, DsR and seed RNA samples were labelled and hybridised in separate experiments such that DsL1, HL1, DsR1 and seed1 were performed together and so forth. The hydrated root RNA, however, was labelled, divided and hybridised against 3 slides in a separate experiment.

Table 2.2: Absorbance readings ($A_{260\ nm}/A_{230\ nm}$) and ($A_{260\ nm}/A_{230\ nm}$) of samples of total RNA used to prepare labelled cDNA and the frequency of incorporation (FOI, pmol dye incorporated x 324.5 (average molecular weight of one kb of DNA in g/mol)/ ng cDNA) and amount of labelled cDNA for each *X. humilis* sample used in the microarray analysis. The RWC of each sample is indicated in brackets next to the sample name.

Sample	*A260 nm/A230nm ratio*	*A260 nm/A280 nm ratio*	*FOI*	*pmol labelled cDNA*
HR (96%)	2.12	2.12	23.12	64
DsL1 (5%)	2.05	2.02	15.56	104
DsL2 (6%)	2.29	2.17	20.74	110
DsL3 (4%)	2.28	2.14	9.74	56
HL1 (97%)	2.13	2.14	17.95	69
HL2 (92%)	2.18	2.32	11.91	87
HL3 (93%)	2.3	2.13	3.73	46
DsR1 (4%)	2.16	2.15	25.3	90
DsR2 (7%)	2.23	2.15	22.79	93
DsR3 (6%)	2.25	2.04	7.15	45
Seed1	0.98	2.14	17.46	86
Seed2	1.35	2.13	6.8	38
Seed3	0.89	2.11	9.26	30

A total of 30 pmol of each labelled cDNA was hybridised to a microarray slide and incubated at 42°C for 16 hours. The slides were washed and scanned and a representative of each biological sample is presented in Fig. 2.4. Signal was detected on all slides including those slides hybridised with RNA from seed tissue which shows that although the slides contain cDNAs prepared from desiccating leaf and root tissue there is a good representation of genes expressed in seeds too. A variation in hybridisation is evident on some of the slides (Fig. 2.4), for example the slide hybridised with DsL1 RNA shows a brighter signal at the bottom of the slide versus the top and the slide hybridised with HR1 RNA shows lower signal in the middle of the slide than on the edges. This may be as a result of uneven hybridisation kinetics when the probe was applied to the slide. The level of background on the slides is low and a closer look at a small section of each slide show uniform spot morphology and even hybridisation across individual spots. These initial observations indicate good hybridisation levels with slight variation in signal due to uneven hybridisation on some slides which must be considered during normalisation. Additionally, the background levels appear fairly low and should not obscure the true signal.

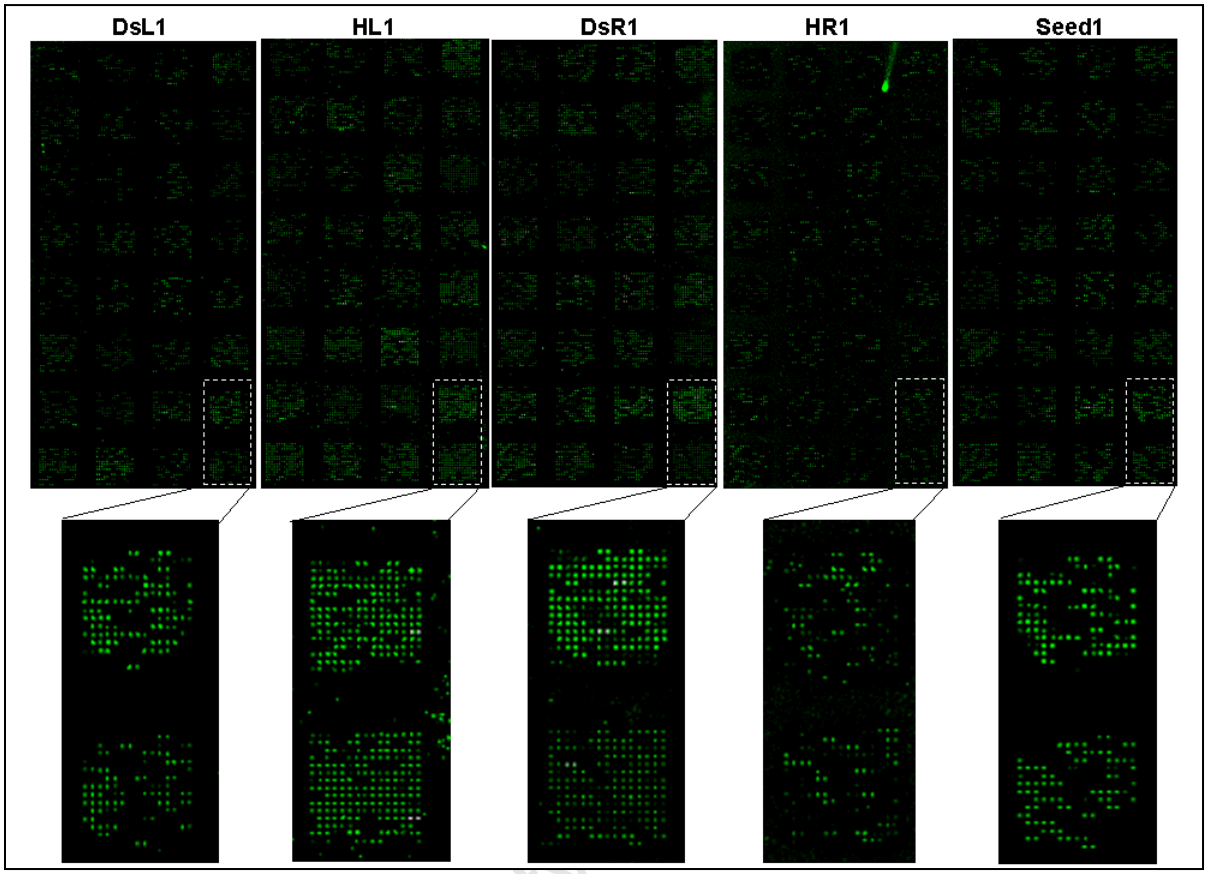


Figure 2.4. Microarray slides of *X. humilis* desiccated leaf, hydrated leaf, desiccated root, hydrated root and seed (biological and technical replicate number one). The two blocks on the bottom hand side of each slide are magnified to show spot morphology and uniformity of hybridisation across each spot.

Data Exploration and Normalisation

The first step in deciding upon an appropriate normalization method is visualization of the patterns of variation in the raw data. The spatial plot in Fig. 2.5 was generated by plotting the log transformed intensity values as a function of their x and y coordinates. It is a false colour plot and represents a red and yellow colour image of the log expression values of each array in the *X. humilis* microarray experiment.

Spatial bias can be introduced by the scanning process or the spatial tendency of the hybridisation sample can lead to uneven brightness across the array. All the slides hybridised with DsL and HL RNA (Fig 2.5) show a gradient of brighter signal from the bottom of the slide decreasing slightly towards the top. This is also apparent when you look at the slides represented

in Fig. 2.4 (DsL1 and HL1). The slides hybridised with DsR, HR and seed RNA show uneven "patches" of higher signal across them which is also evident but not as clearly on the scanned slide images (Fig. 2.4). For example DsR1 appears to have brighter signal along the bottom and right hand edge of the slide (Fig. 2.4) and this is reflected by the higher amount of yellow shown on the bottom and right side of the slide in the spatial plot (Fig. 2.5). These random effects, most likely artefacts of uneven hybridisation must be accounted for and smoothed during normalisation. An additional consideration is the spatial distribution of features across the slides. The library used to print the slides in this study was made from RNA extracted from a series of desiccating leaf and root samples, and cDNAs corresponding to desiccating leaf and root samples were grouped together accordingly in microtitre plates. This could result in localised clusters of genes which are differentially regulated under the various treatments tested. However, in Fig. 2.5 no obvious patterns of clusters of high signal were evident which indicates an even hybridisation of library specific genes across the slides.

Spatial normalisation was performed on the *X. humilis* data using the spat.norm function of Wit and Mclure (2004). This function, based on the 2-D loess principle, performed local loess on the data to find any spatial trend affecting local average expression values. This trend was then removed from the original data, followed by scale normalisation to remove second order trends across the entire array. In more detail, a smooth spatial fit was calculated by a series of local linear regressions using a significant fraction of the original data. In this case a total of 50% of all the spots on the slide closest to the spot (~3600 data points) were included for each regression. The smooth surface (also referred to a loess of degree one) was then subtracted from the original data thereby removing a smooth spatial trend from the data. Additionally, regions of variation across an array may result in location-dependent scale differences between spots (Yang et al., 2002). This was dealt with again by smoothing the absolute difference between the observed expressions and in this case the second smoothed surface was an estimate of the location-dependent scale parameter, calculated from 75% of the total number of spots closest to each location (~5400 data points) for each regression. Division of the location smoothed surface by the smoothed scale surface removed any first and second-order spatial effects.

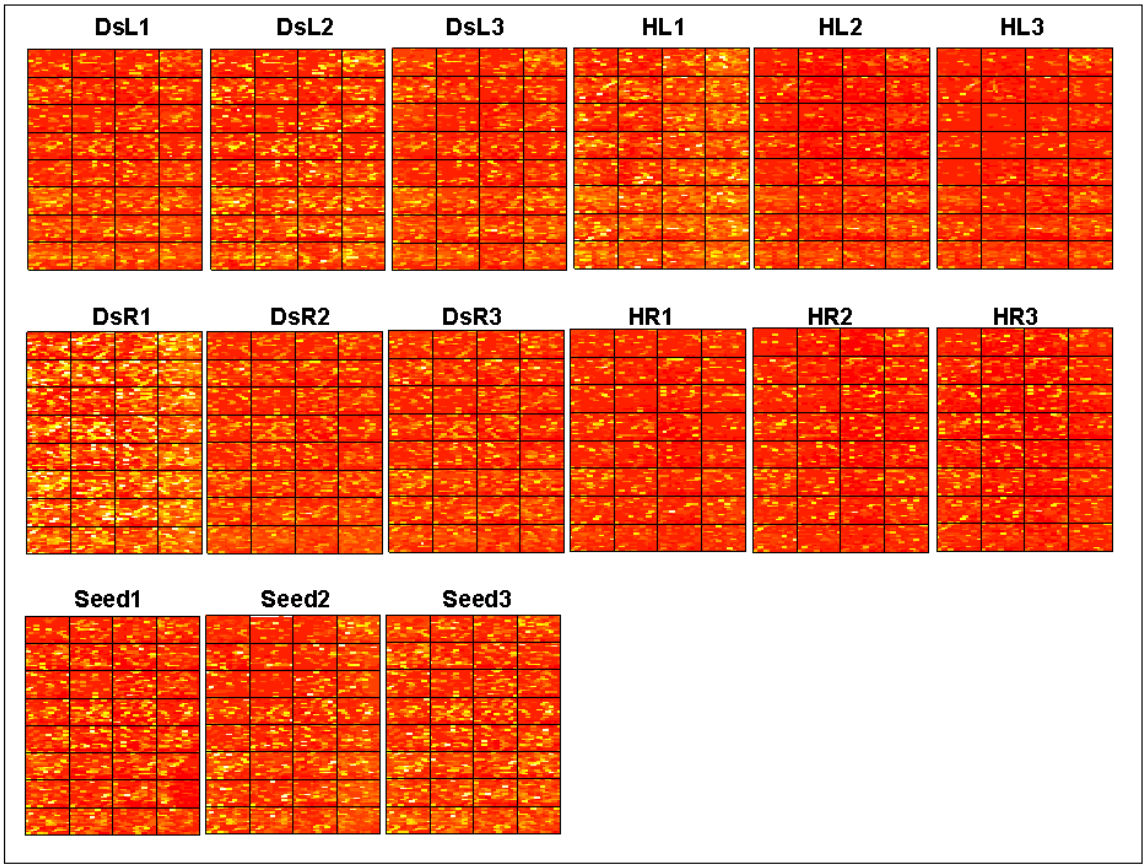


Figure 2.5. False colour plots of raw (non-normalised) *X. humilis* data. Larger values are represented with yellow of increasing intensity, and smaller values are represented with reds of increasing intensity. Each plot represents a microarray slide and blocks represent spots printed by individual pin tips.

Transforming the data changed their scale and to remedy this problem, the smoothed values were multiplied by the median of the scale smoothed surface median, before adding the median of the original data (Wit and McClure, 2004). The x and y coordinates and the intensity for each spot are required for this function and the normalized data is returned automatically on the original scale as in this case. Lucidea spike controls were ignored for this step as control spots in a concentrated region of the array may result in inaccurate adjustments by the smoothing method. Figure 2.6 shows the 15 *X. humilis* slides after spatial normalisation and a smoothing of the spatial effects is evident. The gradient in the DsL and HL and DsR slides has been smoothed and the signal is more even across all the slides.

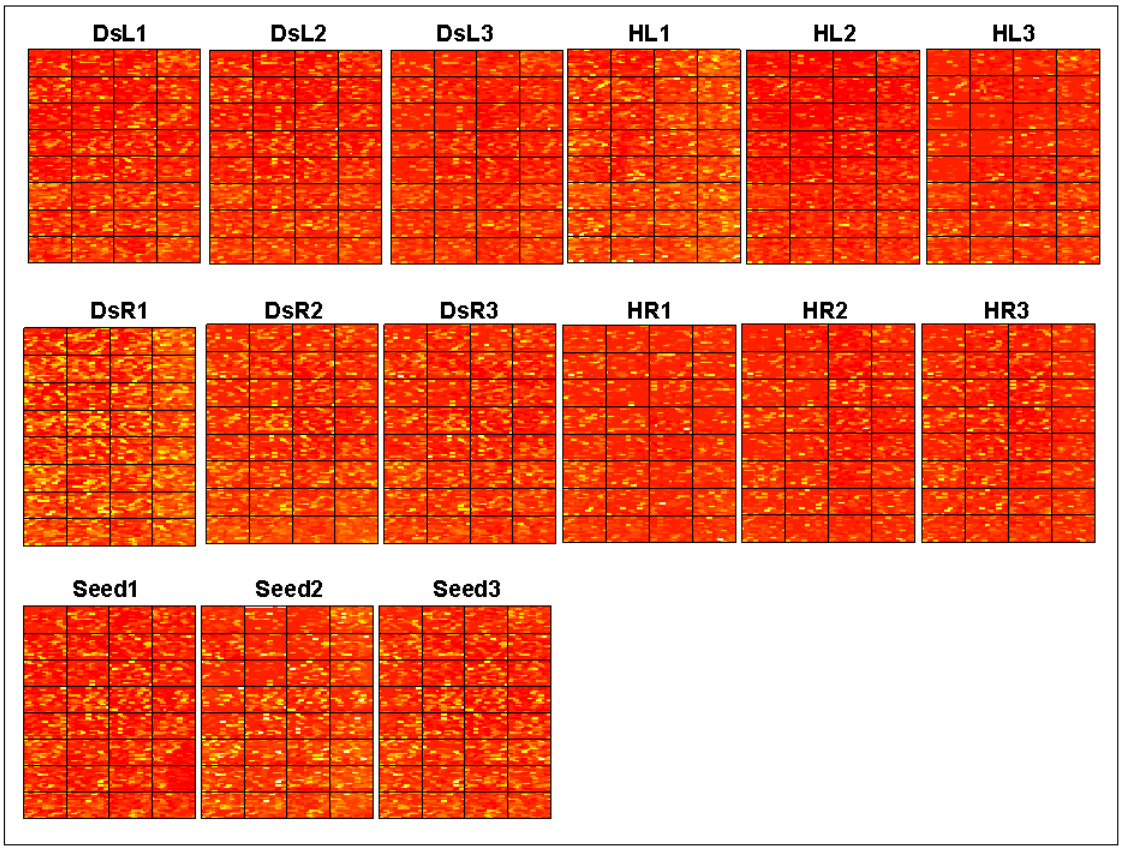


Figure 2.6. False colour plots of spatially normalised *X. humilis* data. Larger values are represented with yellow of increasing intensity, and smaller values are represented with reds of increasing intensity. Each plot represents a microarray slide and blocks represent spots printed by individual pin tips.

Background is a consequence of non-specific binding of targets to the slide and fluorescence by the glass slide itself (Stekel, 2003) and can contribute to the overall spot intensity resulting in overestimation of target abundance for specific features. After spatial normalisation background correction was performed using the bkg.norm function following the "deterministic" approach in the smida package in R (Wit and McClure, 2004). This simple, deterministic method is an adequate method for background normalization of a single channel arrays, but it may not be advisable for dual channels. This is because the noise at the lower level may be exaggerated by simply setting the lowest values (or negative values) to zero in each channel. Subsequently this may influence the dye normalization, which uses the scatterplot of Cy5 vs Cy3 as a reference for correction.

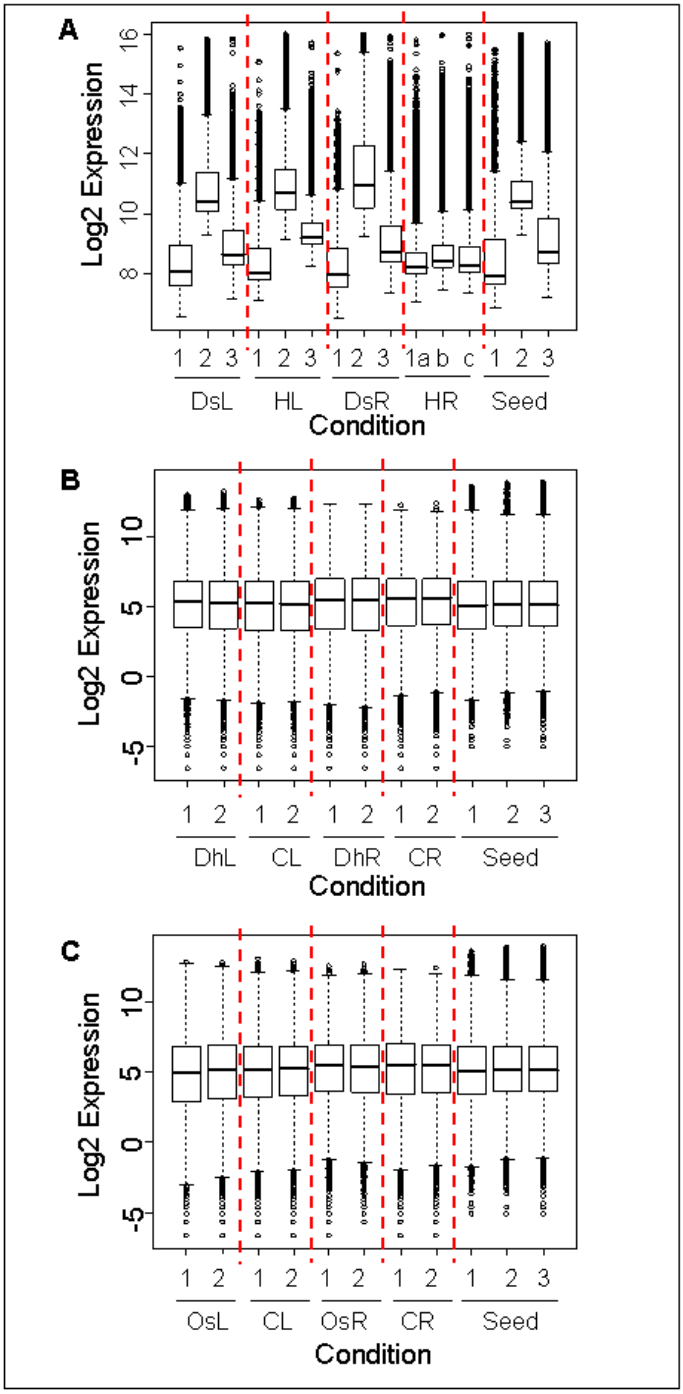


Figure 2.7. Box plot of (A) raw (non-normalised) *X. humilis* desiccation treatment data and mean-scaled (B) *A. thaliana* dehydration treatment data (C) *A. thaliana* osmotic stress treatment data. Each biological replicate is indicated above the sample name, except for the HR sample which indicates three technical replicates (1a, 1b, 1c).

To normalise the expression values between conditions or treatments, the spread of the data between the arrays must be compared and evaluated. Fig 2.7 shows box plots of the *X. humilis* and *A. thaliana* non-normalised expression data across biological replicates and conditions tested. A large variation in the spread of data between the fifteen arrays for *X. humilis* is apparent (Fig. 2.7A) as evidenced by the differing means and ranges of intensities. This is a consequence of hybridisation against the microarray slides with one sample per treatment as there is no common reference sample to normalise the PMT against for all the slides. Additionally, the variation may also arise as a consequence of differences in labelling of the biological samples. While the spread of data in the *A. thaliana* experiment is more consistent (Fig. 2.7B and C) there is slight variation between replicates and conditions which would need to be corrected.

Fig. 2.8 shows that the distribution of the expression data sets for *X. humilis.* These histograms show the frequency of occurrence of log intensity values for each of the condition tested. The biological replicates of each condition are represented individually in the first 5 panels and the last panel shows all the conditions (and biological replicates) together on one plot. The distributional spread of each of the biological replicates (e.g. DsL1, DsL2 and DsL3) for each condition are similar, with the exception of DsR2 and HL2 (Fig. 2.8 panels 1-4) The mean log intensity values are different for all the biological replicates, denoting variation in overall log intensity levels across the slides. In each of these cases the first and third replicate are most similar to each other while the second replicate is consistently brighter. This is also reflected in the box plots in Fig. 2.7A. Notably, the experiments performed on the same day (e.g. DsL1, HL1, DsR1 and seed1) are more similar to each other as indicated by the equivalent distributions and mean intensity values of these samples. The HR samples (Fig. 2.8 panel 5), which were labeled and hybridized together in an independent experiment, showed equivalent distributions and mean log intensities. This is not unexpected as these samples are technical replicates from the same RNA source which were labeled and hybridized under the same conditions. Significantly, the slides hybridized with seed RNA (Fig. 2.8 panel 3) show a robust signal even though the cDNA probes on the slides originate from a leaf/root desiccation library (Collett et al., 2004) and these distributions are similar to those of the leaf and root samples indicating a significant overlap in the hybridization of genes between seed and vegetative tissue. These

histograms show that the spread of data for all the conditions and replicates vary as reflected by the difference in mean intensities and extreme values, and they do not appear to follow a normal distribution. The variation between data sets is most likely due to differences in labeling, hybridization and data capture and must be accounted for during normalization.

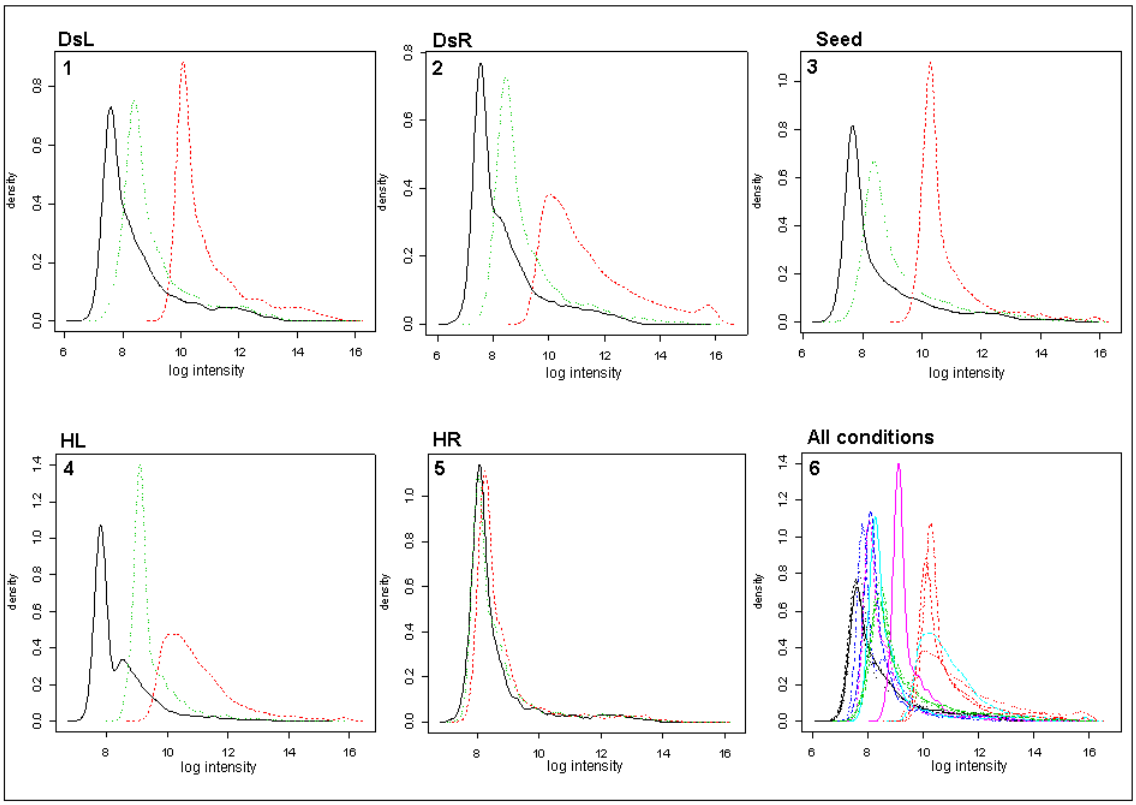


Figure 2.8. Histograms of raw (non-normalised) *X. humilis* desiccation treatment and seed. Panel 1, Desiccated Leaf biological replicates 1, 2 and 3. Panel 2, Desiccated Root biological replicates 1, 2 and 3. Panel 3, Seed biological replicates 1, 2 and 3. Panel 4, Hydrated Leaf biological replicate 1, 2 and 3. Panel 5, Hydrated Root technical replicate 1, 2 and 3. Panel 6, all conditions and replicates. — Biological replicate 1, ----Biological replicate 2, ---- Biological replicate 3. The density (frequency of occurrence) of raw log2 intensity values is plotted on the y axis and the raw log2 intensity values are plotted on the x axis for each distribution.

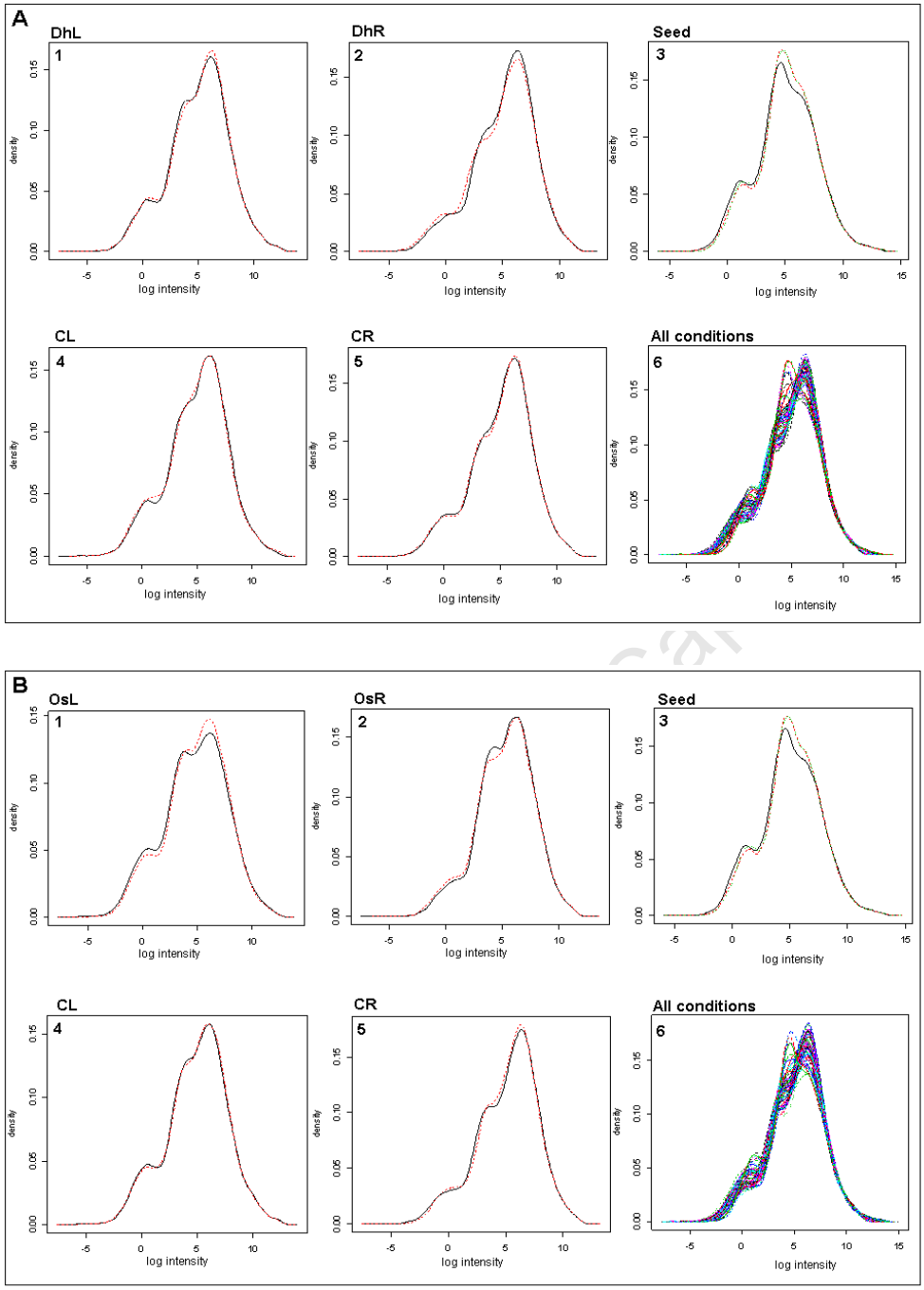


Figure 2.9. Histograms of mean-scaled *A. thaliana* A) dehydration stress treatment and seed and B) osmotic stress treatment and seed. Panel 1, Dehydrated/Osmotic Leaf biological replicates 1 and 2. Panel 2, Dehydrated/Osmotic Root biological replicates 1 and 2. Panel 3, Seed biological replicates 1, 2 and 3. Panel 4, Control Leaf biological replicate 1 and 2. Panel 5, Control Root technical replicate 1 and 2. Panel 6, all conditions and replicates. — Biological replicate 1, ----Biological replicate 2, ----Biological replicate 3. The density (frequency of occurrence) of raw log2 intensity values is plotted on the y axis and the raw log2 intensity values are plotted on the x axis for each distribution.

The distributions and mean intensity values of all the conditions and biological replicates are very similar for both the *A. thaliana* dehydration stress and seed (Fig. 2.9) and osmotic stress and seed (Fig. 2.9B) data. This is expected as the data had been subjected to a coarse mean-scale adjustment. The x axis shows log intensity values as low as -5 and this is from genes which had very low or non detectable levels of signal. Intensity values below one will result in a negative value when log transformed. There are quite a high number of these genes with low intensity levels as this data set includes genes representing the majority of the *A. thaliana* genome and not all the genes will be expressed under the each condition tested. In all cases, including the seed data, the spread of intensity values are very similar with the same mean intensity values (as reflected in the box plots in Fig2.7 B and C) but the expression values do not follow a normal distribution.

The range of density values on the y-axis of the plots is different between the two species as a consequence of the difference in the number of genes. The *A. thaliana* data comprises a far larger number of genes (21 425 versus the 3500 of *X. humilis*) and as a result the relative frequency of genes with certain log intensity values will be diluted in the *A. thaliana* data compared to that for *X. humilis*. Also, *X. humilis* is a "boutique" array, containing genes which are known to change in response to the desiccation (Collett et al., 2004). This will affect the relative abundance of transcripts at various log intensity ranges. The log intensity values (on the y-axis) differ for the two species and this is a consequence of the different labeling techniques employed (biotin labeled Affymetrix samples versus fluorescent labeled cDNA microarray samples) by the two platforms to generate the intensity data.

For both the *X. humilis* and *A. thaliana* data sets the distribution of the intensity values must be adjusted onto the same scale to be able to compare the conditions for each species. The values must be adjusted so that the distributions are comparable and the mean intensity values are the same for each data set. The normalization strategy chosen is based on assumptions regarding the distribution of the data sets. The normalize.quantiles function in the affy package (Bolstad et al., 2003) was applied to both the *X. humilis* and *A. thaliana* data sets separately. This method, based on the quantile-quantile plot concept, ensures an equivalent distribution of intensity values

between data sets and is effective for dealing with non-linear data that does not follow a normal distribution (Bolstad et al., 2003). Non-linearity arises as brighter arrays tend to have values compressed near the top of the intensity range while darker arrays are inclined to exhibit the opposite trend and this can be clearly seen in the *X. humilis* data set (Fig. 2.8). Quantile normalisation involves the ranking of features based on their intensity values, the mean of which is then calculated across the rank distributions and is ideal for data with non-equivalent distributions. The calculated mean replaces the original value and finally the normalised data is rearranged back to the original ordering effectively smoothing out the non-linearity (Fig 2.10A). Quantile normalization is also recommended for dealing with array-wide changes that may occur due to different conditions (Wit and McClure, 2004). This was expected with the *X. humilis* data in particular as it was a focused array including genes prepared from a library of desiccating vegetative tissue and as a result a large percentage of the genes on the array were expected to show changes in expression levels across the conditions tested. Other methods, such as global normalisation, assume that the variations in brightness across the arrays can be resolved by a linear shift in the data to adjust the levels of each of the arrays to an equivalent overall level, a suboptimal solution. Large complete genome arrays are not generally affected by this due to the large number of data points but for small custom arrays with a few hundred genes, important biological information may be lost (Wit and McClure, 2004).

The *A. thaliana* data set was also normalised using quantile normalisation to be consistent with the *X. humilis* data set and as it has been shown to be the best approach for the normalisation of high density oligonucleotide array data based on variance and bias (Bolstad et al., 2003). After normalisation it can be seen that the distributions for all the treatments conducted for *X. humilis* (Fig. 2.10A and Fig. 2.11A) and those of *A. thaliana* are similar and that they have been centred similarly for each species (Fig 2.10 B,C and Fig2.11B,C). The histograms (Fig. 2.10) show the same mean value and spread of data for each species which will allow comparisons between the conditions tested. The small bump indicated by the arrow on the histogram of *X. humilis* (Fig. 2.10A) is a consequence of the background subtraction method which as mentioned earlier can result in a number of low values being set to zero (Wit and McClure, 2004).

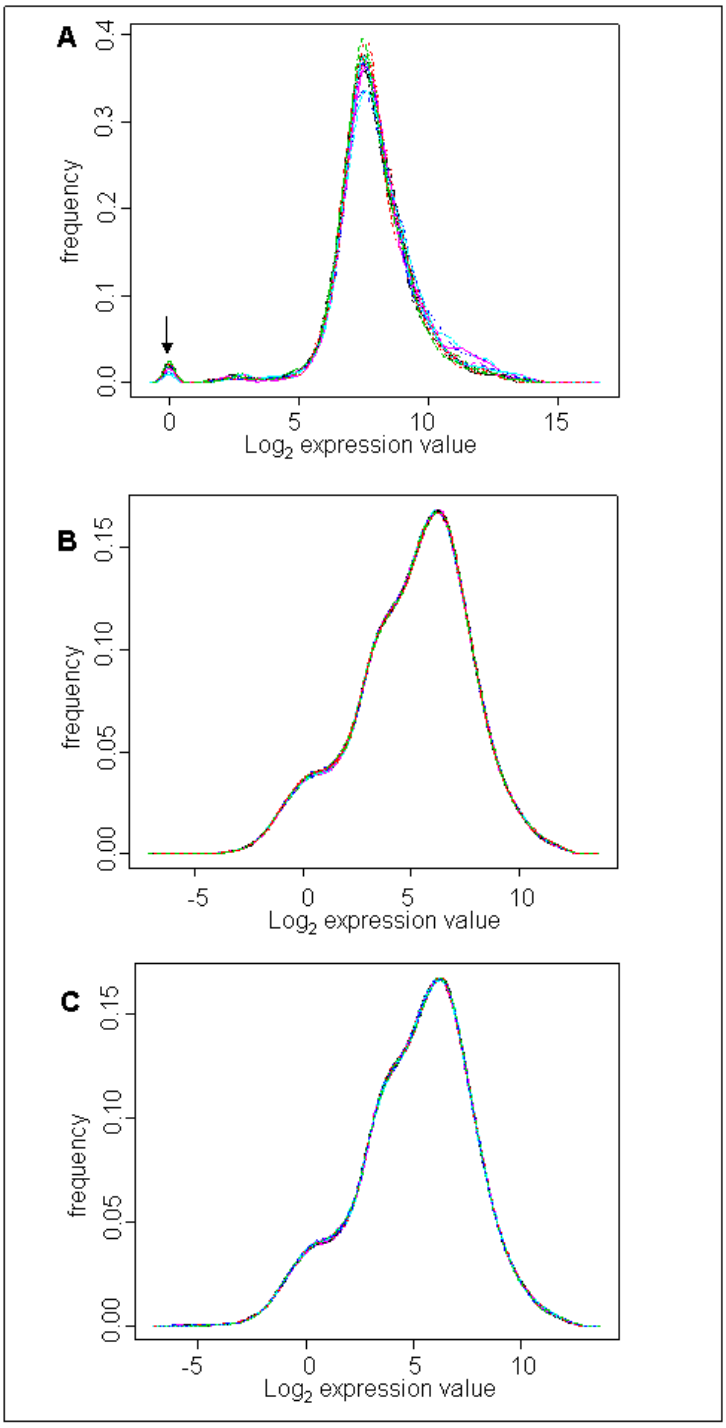


Figure 2.10. Histograms of normalized A) *X. humilis* desiccation stress treatment and seed, B) *A. thaliana* dehydration stress treatment and seed and C) *A. thaliana* osmotic stress treatment and seed. The density (frequency of occurrence) of normalized log2 intensity values is plotted on the y axis and the normalized log intensity values are plotted on the x axis for each distribution.

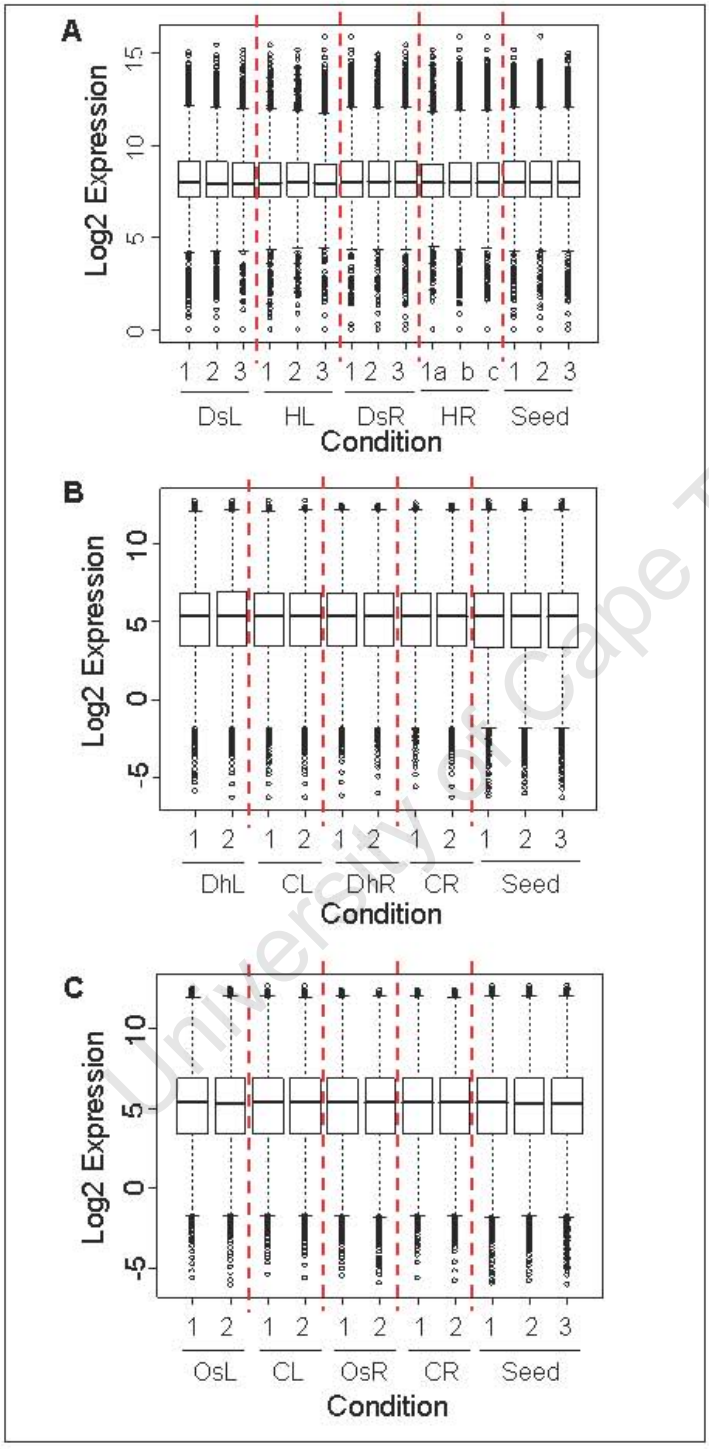


Figure 2.11. Box plot of each normalised data set. (A) *X. humilis* desiccation treatment (B) *A. thaliana* dehydration treatment (C) *A. thaliana* osmotic stress treatment. Each biological replicate is indicated above the sample name, except for the HR sample which indicates three technical replicates (1a, 1b, 1c).

Verification of Normalisation Strategy

It must be noted at this point that quantile normalisation is generally recommended as a means to normalise replicate arrays while normalisation across conditions where a lot of genes are expected to change in expression level should be based on quantile normalisation of an invariant set of genes (Wit and McClure, 2004). This involves ranking of set of genes whose expression levels remain constant across the tested conditions and then the rest of the data is normalised by extrapolation of the ranked invariant set.

However, this approach was not possible for the *X. humilis* data as a consequence of the focused library used to prepare the microarray slides. These cDNAs were expected to show changes in expression level across the conditions (Collett et al., 2004) and this meant there were insufficient genes with constant expression levels across the conditions to create a significant invariant set. The lucidea controls initially included to verify the normalisation approach were not usable due to inconsistent hybridization (data not shown). Therefore, the normalisation approach was verified by comparison of the normalised gene expression values for the *X. humilis* DsL and DsR samples with the expression levels of genes whose expression has been verified by microarray and Northern blot analysis (Fig. 2.12) (Collett et al., 2004; Illing et al., 2005). For each of these genes the average log2 intensity level was plotted of the HL and DsL samples and these plots were compared with the gene expression levels at 100% and 5% RWC on the *Northern* blots generated previously (Fig. 2.12). In each case the increase or decrease in gene expression observed in this microarray study correlated with the expression pattern in the *Northern* blot. Additionally, the data correlated for low and moderately expressed genes (Fig. 2.12 top graph) and for highly expressed genes (Fig. 2.12 bottom graph). While data is only available for comparison of the HL and DsL samples it showed a correlation in the trend of the expression patterns for these genes between the two methods which supports the choice of normalisation strategy.

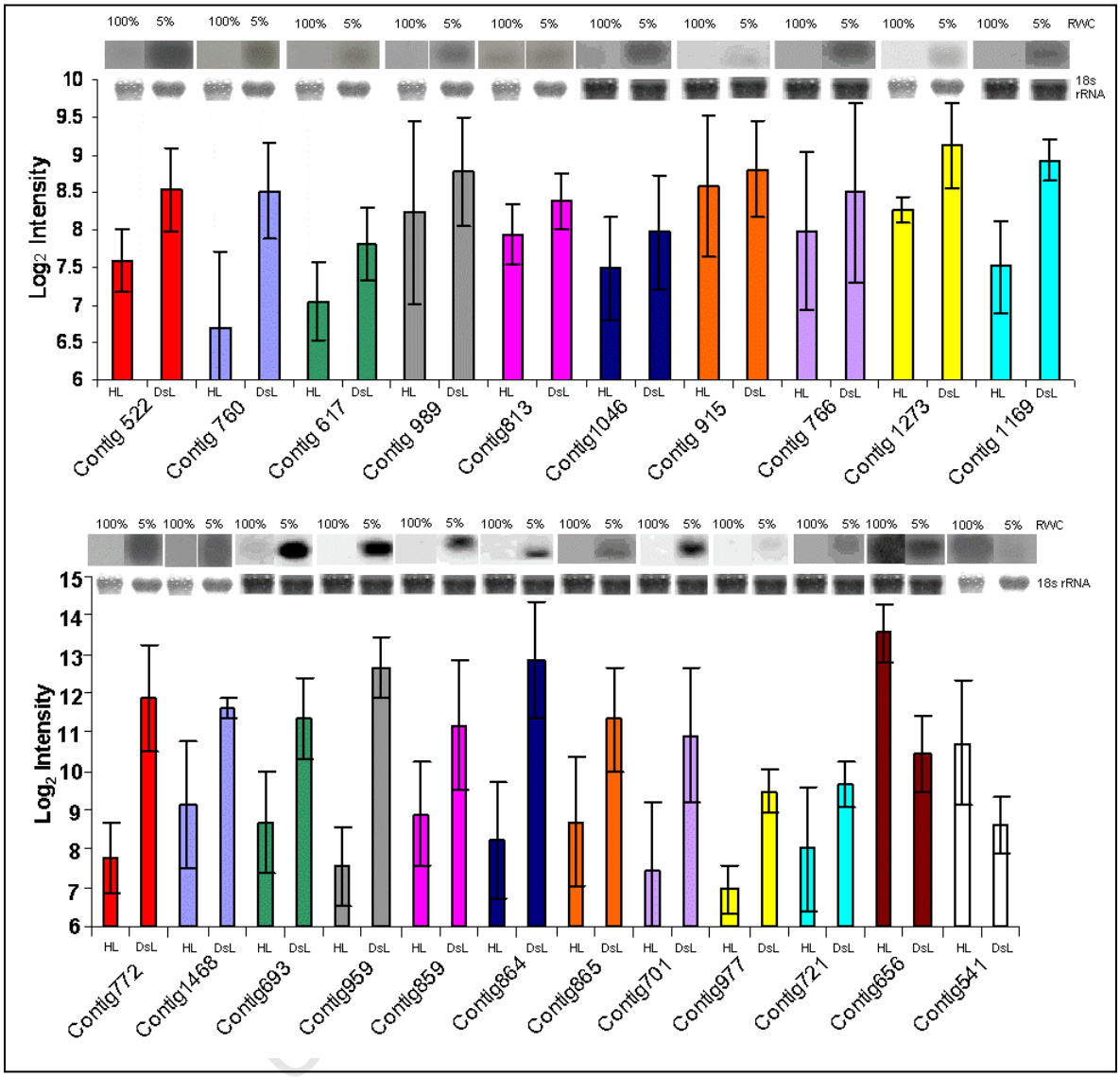


Figure 2.12. Verification of normalisation approach for *X. humilis*. Average normalised log2 intensity values for the HL and DsL sample for each contig are plotted. Error bars represent ± standard deviation. The corresponding *Northern* blot is presented above each barchart. These *Northern* blots were conducted by Helen Collett and Athur Shen (Collett et al., 2004; Illing et al., 2005).

A set of housekeeping or reference genes was selected to verify the normalization approach for the *A. thaliana* microarray data. These genes were recently shown to be stably expressed across a number of developmental and environmental conditions as well as in different organs (Czechowski et al., 2005). A subset of these reference genes, which were present on the Affymetrix ATH1 chip, showed stable expression levels across all the conditions tested in this

study (Fig. 2.13). These expression levels were uniform in control and dehydrated vegetative tissue and mature seed (Fig. 2.13A) and also in control and osmotic stressed vegetative tissue and mature seed (Fig. 2.13B). They also covered a wide range of absolute expression levels showing that genes expressed at high, moderate or low levels all responded well to the normalization strategy.

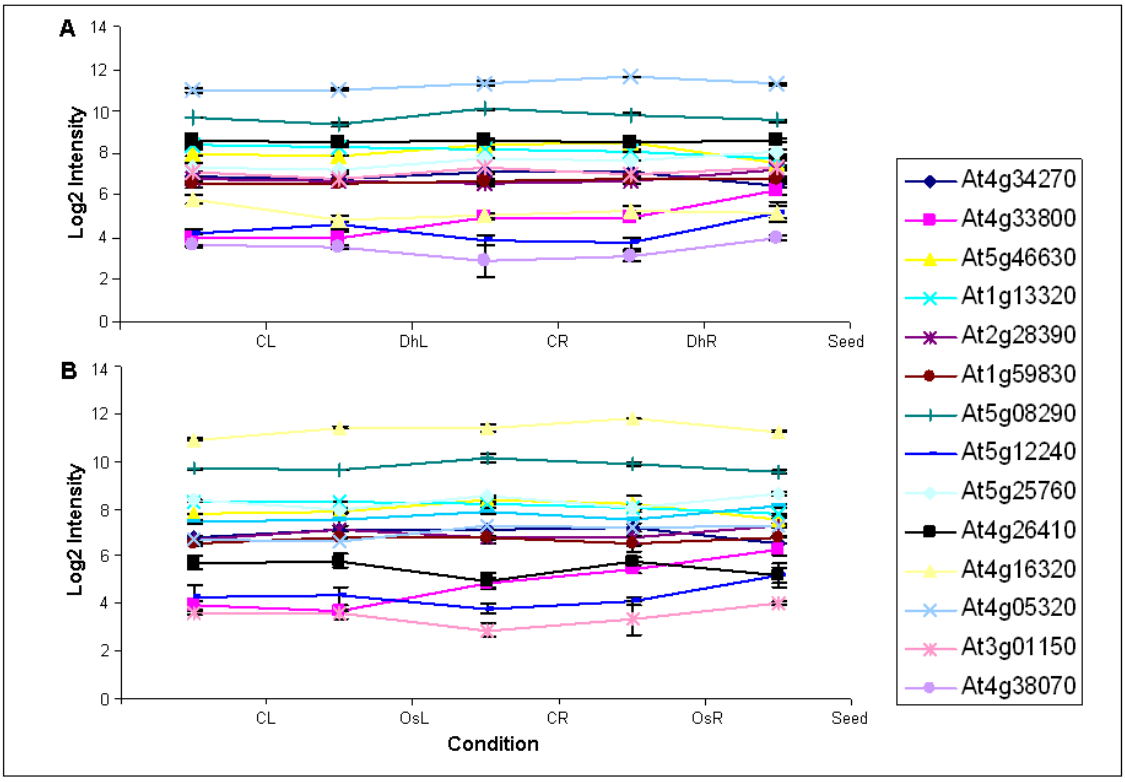


Figure 2.13. Verification of normalisation approach for *A. thaliana* A) Dehydration stress treatment and seed. B) Osmotic stress treatment and seed. Average normalised log2 intensity values for each condition are plotted. Error bars represent ± standard deviation.

2.4. Conclusion

Microarray experiments comprise a number of levels from the initial data capture through to the final analysis. Careful consideration during the initial stages of data capture and normalisation is crucial as this can impact negatively on the final outcome. The choice of normalisation is empirical and data exploration is important to assess the bias and to assist with the choice of normalisation strategy. A set of control genes should be included to assist with or at least to

validate the chosen method of normalisation. The *X. humilis* and *A. thaliana* data sets were processed after detection and identification of systematic variations. The *X. humilis* data was spatially normalised, background corrected and subjected to quantile normalisation and the *A. thaliana* data was normalised by quantile normalisation. These normalisation strategies were validated through comparison of normalised gene expression values with genes whose expression levels have been verified beforehand for *X. humilis* and through a set of reference genes of *A. thaliana.* In both cases the correlation between the microarray data and the validation sets indicate that the choice of normalisation strategy was suitable. The next step involves the identification of differentially expressed genes and clustering of genes with similar expression patterns across the conditions for both the *X. humilis* and *A. thaliana* normalised datasets.

Chapter 3

Microarray data analysis

3.1. Introduction

We hypothesize that DT in vegetative tissues in *X. humilis* is a consequence of the activation of genes that increase in transcript abundance in mature orthodox seeds, in leaves and roots in response to extreme water loss. To test this theory we investigated the overlap between genes expressed in desiccated leaves, roots and seeds in *X. humilis*. To do this the mRNA transcript levels of 3105 *X. humilis* cDNAs in desiccated leaf (DsL), desiccated root (DsR) and mature seed were compared with those in hydrated leaf (HL) and hydrated root (HR). A similar analysis was performed on gene expression data from *A. thaliana* in which the levels of mRNA transcripts in leaves and roots in response to abiotic stress was compared to mature seed. Our hypothesis predicted that there would be a significant overlap of genes that were up-regulated in desiccated leaves, roots and seeds in *X. humilis*, but that this would not be the case in *A. thaliana* where most genes activated in response to abiotic stress would be different from those up-regulated during the latter stages of seed maturation. Additionally it was predicted that the expression of homologues of genes that are specifically expressed in desiccated vegetative tissue of *X. humilis* would be restricted to seed in *A. thaliana*. In order to perform this analysis, differentially expressed genes were identified and their expression profiles across the conditions investigated. This was to assess the similarity between tissues with respect to their gene expression profiles and to identify groups of genes that were regulated in a coordinated fashion. Differentially expressed genes in *X. humilis* were sequenced and annotated and those genes overlapping between seed and desiccated vegetative tissue were identified. The expression patterns of these homologues were investigated in *A. thaliana*.

An overview of methods to identify differentially expressed genes from microarray datasets, and the clustering and functional interpretation of these genes is reviewed below with particular emphasis on those methods used for this study.

3.1.1. Identifying differentially expressed genes

The majority of microarray experiments aim to find genes whose expression change across different conditions as it is assumed that these changes in mRNA transcript abundance contribute to the change in phenotype under investigation. In the early days of microarray analysis a fixed fold change (usually 2-fold) was used to identify differentially expressed genes. Although a logical approach, this does not take biological variation into account and genes which show smaller, but reproducible changes in mRNA transcript abundance are ignored. It has since been decided that fold change should be accompanied by statistical testing (Quackenbush, 2002). Tests such as the Student's t-test and its variants (Lönnstedt and Speed, 2002; Storey and Tibshirani, 2003), empirical Bayesian methods (Long et al., 2001; Baldi and Long, 2001; Lönnstedt and Speed, 2002), Mann-Whitney test (Wu, 2001), CLEAR-test, data adaptive methods, regression analysis (Montaner et al., 2006) and ANOVA (Kerr et al., 2000; Long et al., 2001) have subsequently been applied to microarray datasets to identify genes that show statistically significant differences in mRNA transcript abundance. These tests allow statistically significant interpretation of the output and genes are identified as differentially regulated or not based on the provided p-value. An accurate estimation of the variability associated with the expression of each gene is essential for correctly identifying differentially expressed genes. However, the number of biological repeats which can be afforded is often limited due to budget constraints and the cost of microarray experiments. It is thus difficult to get an accurate estimation of variance associated with each gene and parametric-based tests such as T-tests and ANOVA are not robust in this scenario. In this case, empirical Bayesian hypothesis testing is a powerful alternative (Efron et al., 2000, 2001) as it combines the estimate of variance across all genes to calculate a "moderated" variance for each gene, weighted averages of the gene specific sample variances and the pooled estimate of variance. Using the pooled information of variance from multiple genes has been shown to increase statistical power, especially when a small number of arrays are considered (Lönnstedt and Speed, 2002; Smyth, 2004; Cui et al., 2005). Limma is one of the packages available in R, which uses empirical Bayes methods for assessing differential expression in microarray experiments (Smyth, 2004). In this method a linear model is fitted to the microarray data which models the systematic part of the data and distinguishes it

from random variation. After this step a series of contrasts can be considered which will allow any combination of comparisons between the initial samples to be made. Empirical Bayesian methods are used to moderate the standard errors (estimate of the error in the measurement) of the estimated log fold changes and to identify cDNAs which show statistically robust changes in mRNA transcript abundance.

In all types of testing the p-values must be adjusted for the number of comparisons conducted to reduce the false discovery rate (Storey and Tibshirani, 2003). The large number of comparisons in microarray experiments, often with a small number of biological replicates, increase the likelihood that putative differentially expressed genes are identified as differentially expressed when they are in fact not. This problem of multiple testing can be dealt with by either controlling the FWER (family wise error rate) (Holm, 1979) or the FDR (False Discovery Rate) (Benjamin and Hochberg, 1995). FWER is the probability of having at least one false positive (Type I error) and FDR is the expected proportion of false positives among the number of rejected (Type II error). The Bonferroni correction is a classical method to control FWER, but this approach can be too stringent often limiting the power to identify significantly expressed genes (Dudoit et al., 2002). A more practical approach is to control the FDR and accept a few false positives if the majority of true positives are chosen (Reiner et al., 2003). The empirical Bayesian approach allows the user to control FWER (Holm, 1979; Hochberg, 1998; Hommel, 1998) and also less stringent methods to control FDR (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).

3.1.2. Clustering

Differential expression analysis produces long lists of genes with associated expression values from which it is difficult to interpret biological function and relevance. These lists can be further mined by cluster analysis and functional classification (the focus of this chapter). Cluster analysis identifies patterns of gene expression and groups genes with similar patterns of expression into classes which allows greater insight into functional pathways or common regulatory mechanisms (Quackenbush, 2002). Several techniques have been used for the

clustering of gene expression data including hierarchical clustering (Eisen, 1998), principal component analysis (PCA; Raychaudhuri, 2000), k-means clustering (McQueen, 1967), self-organizing maps (Tamayo, 1999) and SOTA (Herrero et al., 2001). Clustering algorithms calculate a distance measure between gene expression profiles and sort the genes into groups based on distance from each other (Dopazo, 2007). Experiments can also be clustered based on gene expression profiles across each experiment to analyze and group these experiments to look for similarities between treatments or conditions (Fig. 3.1) (Quackenbush, 2001).

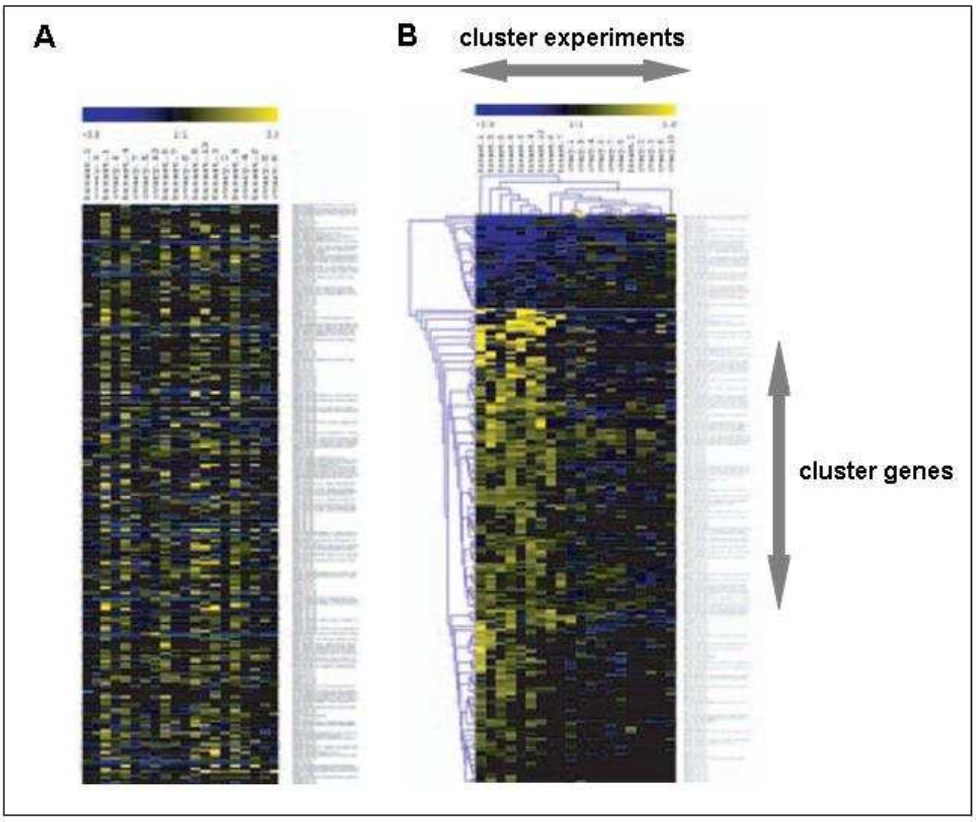


Figure 3.1: Cluster analysis of microarray gene expression. A) Unordered microarray data. B) Microarray data subjected to clustering. Microarray data can be clustered across conditions and across genes based on gene expression profiles (Adapted from Quackenbush, 2006).

Hierarchical clustering is the most commonly used method in microarray experiments. This is an unsupervised method (meaning we have no a priori knowledge about the expected clusters) that produces an easy to visualize linear ordering of the objects. Single expression profiles are joined to form groups and these are further joined in an agglomerative manner until the process

terminates with a single hierarchical tree. To achieve this, a pairwise distance matrix is calculated for all the genes to be clustered. The matrix is searched to find the two most similar genes and these are merged to produce a new cluster, the distance is then calculated between this new cluster and all the other clusters (or genes). These steps are repeated until all the objects are in one cluster. There are several variations of hierarchical clustering that differ in the how the distances are measured between clusters including single-, average- and complete- linkage. In general average-linkage clustering is suitable for gene expression data (Quackenbush, 2001). While hierarchical clustering algorithms tends to be mainly agglomerative (Dopazo, 2007), a divisive version of this clustering method has also been developed (Kaufman and Rousseeuw, 1990). The 'DIANA' algorithm constructs a hierarchy of clustering beginning with one large cluster containing all the gene expression profiles (Fig.3.2). The clusters are divided repeatedly until each cluster contains a single gene. At each stage, the cluster with the largest diameter is selected which is represented by the largest dissimilarity between any two of its gene expression profiles. To divide a cluster, the genes which are the most dissimilar from the other genes within that cluster are identified and assigned to a "splinter group". In subsequent steps the algorithm reassigns genes that are closer to the splinter group than the original cluster. This results in the division of the selected cluster into two new clusters. This process continues until to one gene remains per cluster or to a user defined end. The output is visualized as a dendrogram (Fig. 3.2) which represents the clustered groups as a tree. The groups are represented as the leaves of the tree and conventionally all the leaves are shown on the same level. The heights of the internodes are related to the metric information used to form the clustering.

Unfortunately hierarchical clustering may impose a hierarchical structure on the data even if there is not any. An alternative clustering approach is to represent the data in a two-dimensional plot without imposing a tree structure and the most common approach is PCA (Wit and McClure, 2004). This technique reduces large sets of differentially expressed genes into a few representative numbers for each sample using factor analysis and principal coordinate analysis. The result is a summary of the most variable parts of the data from a sample represented as principle components. These principle components are a measure of the distance between the

samples, and the information on specific genes is forfeited to allow a global view of patterns within the data. As a result, the closer two samples are in a projection of the principle components (Fig. 3.3), the more similar the transcriptional expression between these two samples (Raychaudhuri, 2000). This reduction technique is suitable for genes and experiments as a means of classification (Quackenbush, 2001).

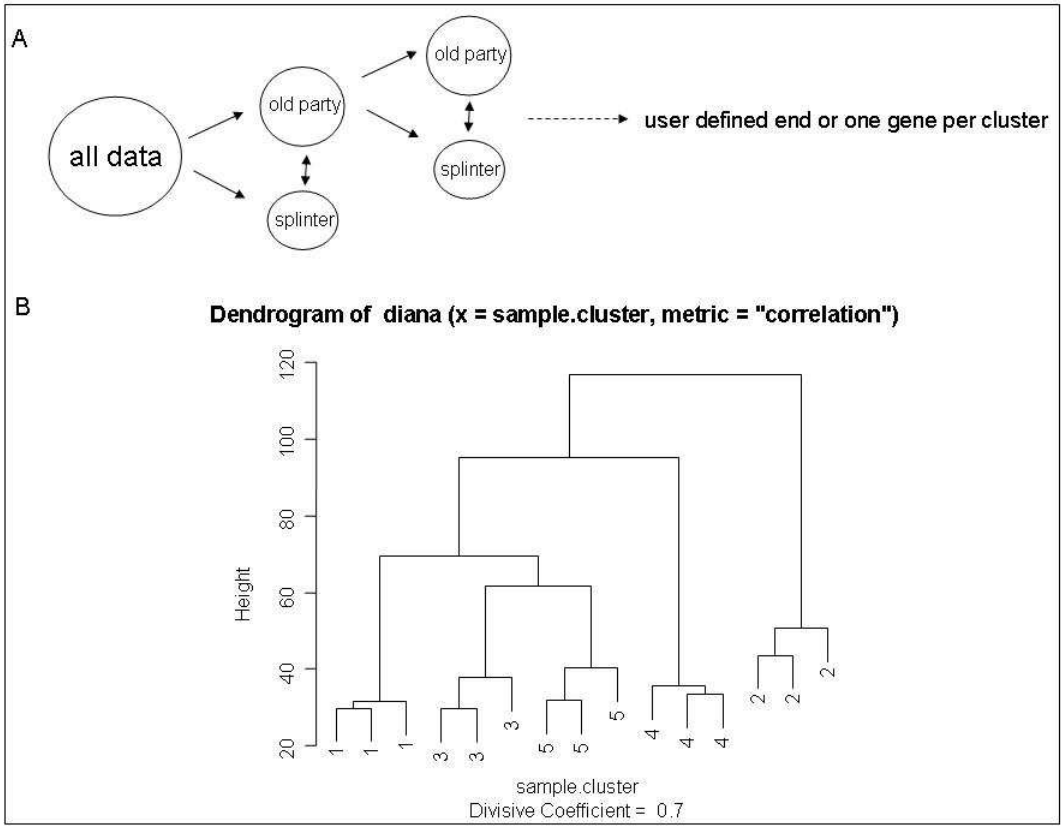


Figure 3.2: 'DIANA' clustering. A) The hierarchy of clustering begins with all gene expression profiles in one cluster. Genes that are most dissimilar from other genes within that cluster are assigned to a splinter group. Genes which are closer to the splinter group are reassigned and the process continues to a user-defined end number of clusters or until one gene remains in each cluster. B) Dendogram of 'DIANA' clustering. The observations are listed in order found by the algorithm. The height represents the diameter of the cluster being split and the divisive coefficient indicates the strength of the cluster structure. The closer this value is to 1 the better the structure.

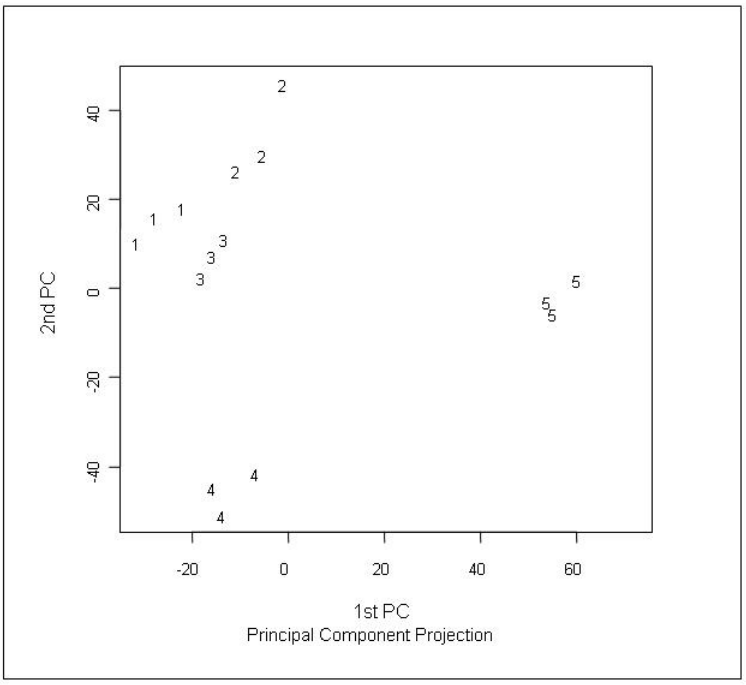


Figure 3.3: An example of a principle component projection. The projection represents the distance between five samples (with three biological repeats each). The biological replicates are very similar and therefore close to each other. Samples 1, 2 and 3 are closely related, while 4 and 5 are more dissimilar.

If there is no prior knowledge regarding the number of clusters that should be present in the data, *k*-means clustering is a good alternative to hierarchical methods (McQueen, 1967). This is a divisive clustering approach that partitions the genes or experiments into groups that have similar expression patterns. The data is partitioned into a fixed number (*k*) of clusters that are similar to each other without consideration for relatedness with other clusters. It partitions the genes into these groups by minimizing the within group dissimilarity for a specified number of groups and a number of user-defined *k* clusters must be considered. While the process is theoretically simple it can be computationally intensive. It begins with a random assignment of genes to one of the user-defined *k* clusters and an average expression vector is calculated for each cluster, this is used to work out the distances between the clusters. Then through a series of iterations genes are moved between clusters and intra- and inter-cluster distances are measured and expression vectors for each cluster are recalculated with each move. Genes will remain in the new cluster if they are closer to it than the previous cluster. The shuffling will cease once moving genes makes

the clusters more variable increasing intra-cluster distance and decreasing inter-cluster dissimilarity (Fig. 3.4) (Quackenbush, 2001; Dopazo, 2007). It is advisable to try and confirm that the data does actually partition into this number of groups and *k*-means clustering can be used in conjunction with other clustering and visualization techniques. An example of such a technique is PAMSAM.

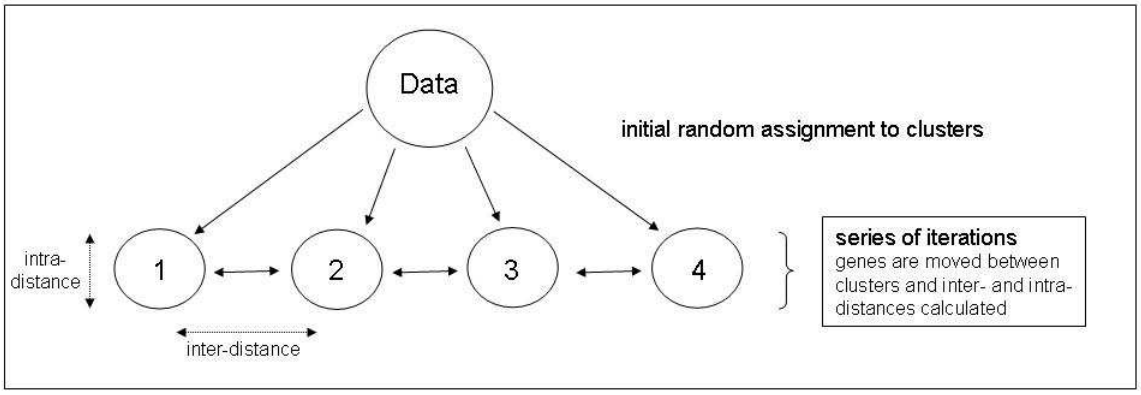


Figure 3.4: An example of *k*-means clustering into 4 groups. Initially the data is randomly assigned to groups. Genes are moved between the 4 groups and either remain in the assigned group or are reassigned to one of the other groups depending on the intra- and inter-cluster distances calculated at each step.

PAMSAM is a powerful technique to identify clusters of genes with similar expression profiles and to visualize the average expression profile (mediod) of each cluster and the level of similarity between clusters (Wit and McClure, 2004). It is a partitioning method that combines a k-means clustering method with a means for visualizing the similarity between the clusters via a Sammon plot. Partitioning methods, such as PAM (Partitioning Around Mediods) are different to hierarchical methods as they try to partition the data into groups in such a way as to optimize a particular score function. In this case the aim is to maximize the average silhouette width (a.s.w) of each cluster (Kaufman and Rousseeuw, 1990). This measure has been shown to be good at distinguishing clusters and large values of a.s.w correspond to clear differences between clusters. PAM selects gene profiles (mediods) from the original data set based on the number of clusters *k* specified by the user. Gene expression profiles are then assigned to the closest mediod and the a.s.w. is calculated for each proposed cluster. PAMSAM tries to increase the a.s.w by reassigning genes from one cluster to another through a series of expand-collapse steps and the final PAMSAM clustering is the one that results in the highest a.s.w. Each time a cluster is

chosen at random and further partitioning or collapsing of that cluster is attempted. A partitioning operation is attempted and if this does not increase the a.s.w. then a collapsing operation will be performed. In a partitioning step, the chosen cluster is clustered itself in the same manner as the initial clustering stage. If splitting this cluster in this way increases the a.s.w. the step is accepted and a new clustering is created with the genes from this cluster now moved into the clusters found in the partitioning step. In a collapsing step, the chosen cluster is combined with another cluster. If the best of these possible collapse operations improves the overall a.s.w. then the collapse is accepted. This process continues until maximal a.s.w. is achieved for the number of clusters *k* specified. Once clustering is completed the medoids of the clusters are visualized. The medoids are plotted in a 2-D Sammon map showing the level of similarity between the clusters (Fig. 3.5). Sammon mapping is a form of multi dimensional scaling that allows visualization of objects in lower-dimensional form. It represents the distances between objects based on distance measures calculated for each cluster rather than for individual genes within the data set. Each medoid is represented by a small graph that shows the medoids' value across the different variables (Wit and McClure, 2004).

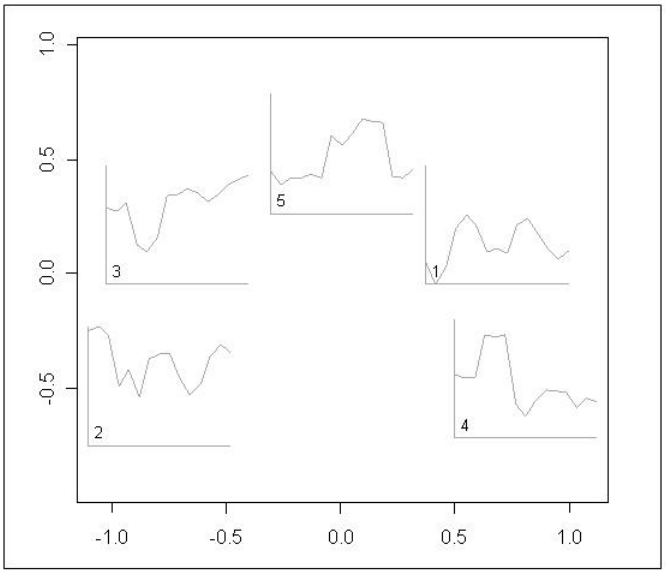


Figure 3.5: An example of PAMSAM clustering. The medoids of 5 gene clusters selected via PAMSAM and the level of similarity between them is shown by a Sammon plot. Each cluster of genes is represented by a small graph that shows the gene expression profile for that group of genes. The y-axis of the cluster graphs are the gene expression values. The x and y axis of the main graph represent distances between the medoids.

3.1.3. Functional Interpretation

Genes operate in sophisticated networks within a cell and those that are co-expressed together generally have common roles in the cell (Stuart, et al., 2003; Lee, et al., 2004). Clustering identifies groups of co-expressed genes and as a result the last but crucial step of clustering analysis is functional interpretation (Al-Shahrour and Dopazo, 2005). There are a number of tools, such as FatiGO (Al-Shahrour et al., 2004), Onto-Express (Khatri et al., 2002), MAPPFinder (Doniger et al., 2003) or FunSpec (Robinson et al., 2002), designed to investigate significant enrichment of biological terms within a cluster of genes. These algorithms compare the enrichment of terms, such as gene ontology (GO) terms (Ashburner et al., 2000), in a cluster of genes against the distribution of that term in the rest of the genes (Al-Shahrour et al., 2004) allowing inference of the functional roles played by the genes in this cluster. Many of these tools are stand-alone applications that are not capable of processing large numbers of genes and the issue of multiple testing is not well addressed. FatiGO, however, is specifically designed to deal with thousands of genes from many different organisms. It extracts relevant GO terms for a group of genes with respect the rest of the genes and a Fisher's exact test for 2x2 contingency tables determines the significance of the terms for the chosen level. The p-values are adjusted to account for multiple testing by controlling either the FWER (Westfall and Young, 1993) or the FDR (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). The results are returned in order of decreasing adjusted p-value and show the GO terms with the most significant differences (i.e. GO terms that are significantly over- and under-represented in a set of genes with respect to the reference group). Tools for the functional interpretation of gene lists are limited to organisms for which detailed gene ontologies have been built up. Currently these include *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae* and *Arabidopsis thaliana* (Al-Shahrour et al., 2005). In order use these software tools for a non-model organisms, such as *X. humilis*, use has to be made of identifying *A. thaliana* orthologues for each of the differentially expressed *X. humilis* genes.

3.2. Materials and Methods

Selection of X. humilis genes for sequencing

A flow diagram (Fig. 3.6) shows the overall outline of the strategies of analysis to identify and cluster differentially expressed genes in *X. humilis* and *A. thaliana*. Differentially expressed cDNAs were identified across the normalised microarray datasets (Fig. 3.6 step 1) for biological replicates using linear modelling in limma in R v2.2.1 (Smyth, 2004) using contrasts summarized in Table 3.1, and correcting for multiple testing (Benjamini and Hochberg, 1995). The code is provided in the appendix (a.3.1).

Table 3.1: Contrasts used in linear modelling to identify differentially expressed genes during desiccation stress and seed maturation in *X. humilis* (n=3)

Contrasts to identify differentially expressed genes
DsL/HL
DsR/HR
Seed/HL
Seed/HR

Sequencing and annotation of differentially expressed X. humilis cDNAs

A total of 2702 cDNAs were selected from the cDNA library which was stored as glycerol stocks in round bottom 96 well microtitre plates (Amersham, Germany). Selected clones were inoculated and incubated in 100µl Luria broth supplemented with 100 mg/ml ampicillin overnight at 37°C with shaking in a 96 well plate. Sequencing was conducted with both M13F and M13R primers at The High-Throughput Genomics Unit (HTGU), Department of Genomic Sciences, University of Washington, Seattle, WA, USA. The raw sequence file (in standard chromatogram format .abd format) were converted to text files in FASTA format using Trace2dbest. During this stage quality values for the bases were written to FASTA format files using phred (Ewing et al., 1998), which can be used by the phrap sequence assembly program (http://www.phrap.org) at a later stage to increase the accuracy of the assembled sequence. The trace2dbest fasta files were downloaded into a sequence directory and vector sequence, adapter sequence and polyA tails were removed and sequences shorter than 100 bp in length were also removed. These fasta files were passed to Partigene, which uses CLOBB2.pl to cluster redundant and overlapping sequences into a set of putative gene objects (Parkinson et al., 2002). The clustering in this software is based on an all against all megablast similarity searches and allows

for incremental updates and the default settings were used to produce a nonredundant set of sequence objects. The sequences in each cluster were then assembled into a consensus sequence (contig) using the program phrap. The resultant nucleotide sequences were translated with Prot4EST (Wasmuth and Blaxter, 2004) using the *A. thaliana* genetic codon usage table. These software packages are publicly available and were accessed from http://zeldia.cap.ed.ac.uk/PartiGene/index.html. A blastp search was performed where each translated sequence was individually queried against the non-redundant GenBank protein sequence database to assign gene identity (Altschul et al., 1997). The best *A. thaliana* match was recorded for each contig and the corresponding At locus tag, gene annotation and best match e-value were retrieved. An e-value of $>1x10^{-2}$ was used as a cut off to assign an *A. thaliana* best match hit and potential homologues with an e-value higher than this was recorded as a "no match" (Fig.3.6 step 1). LEA proteins contain regions of low sequence complexity (Wise and Tunnacliffe, 2004) and these regions are usually masked out in BLAST. BLAST also typically scores matches over longer oligopeptides and introduces penalties for non-contiguous matches, which could arise due to the filtered regions. As a result the algorithm might miss similarities between the query sequence and a LEA protein sequence or return a low e-value. We therefore used a lower e-value as a cut off to ensure that we identified as many *LEA* homologues as possible.

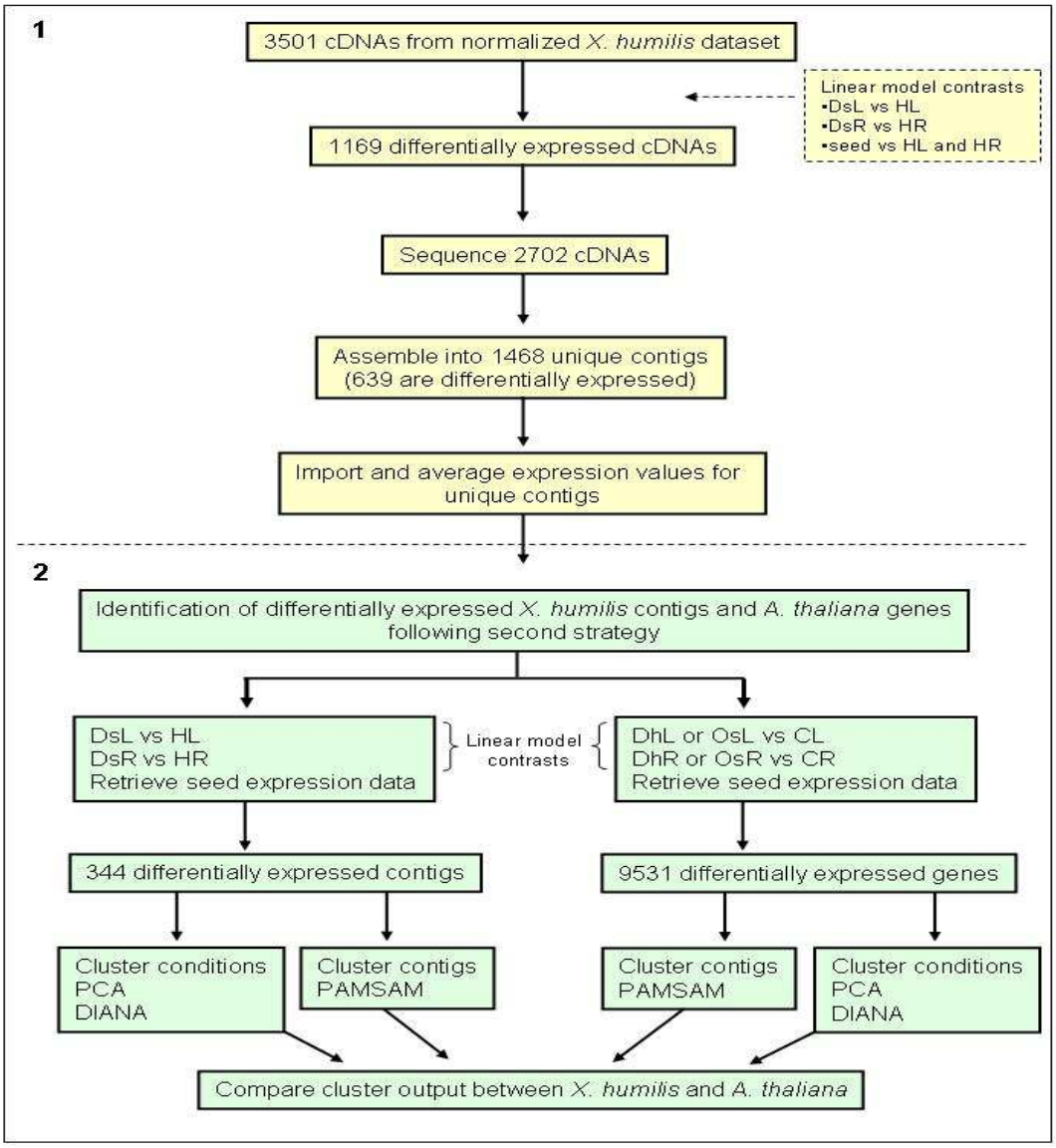


Figure 3.6: Flow diagram depicting the strategy of analysis employed. Step 1: Identification and sequencing of *X. humilis* cDNAs differentially expressed between desiccated vegetative tissue and hydrated tissue and between seed and hydrated tissue. Step 2: Identification of contigs differentially expressed between desiccated and hydrated tissue in *X. humilis* and of genes differentially expressed between osmotically stressed and control vegetative tissue in *A. thaliana*. Gene expression profiles were clustered and compared between the two plants. Step 3: Analysis of gene expression profiles of *X. humilis* homologues in osmotically stressed vegetative tissue and seed of *A. thaliana*. Functional annotation of *X. humilis* gene clusters.

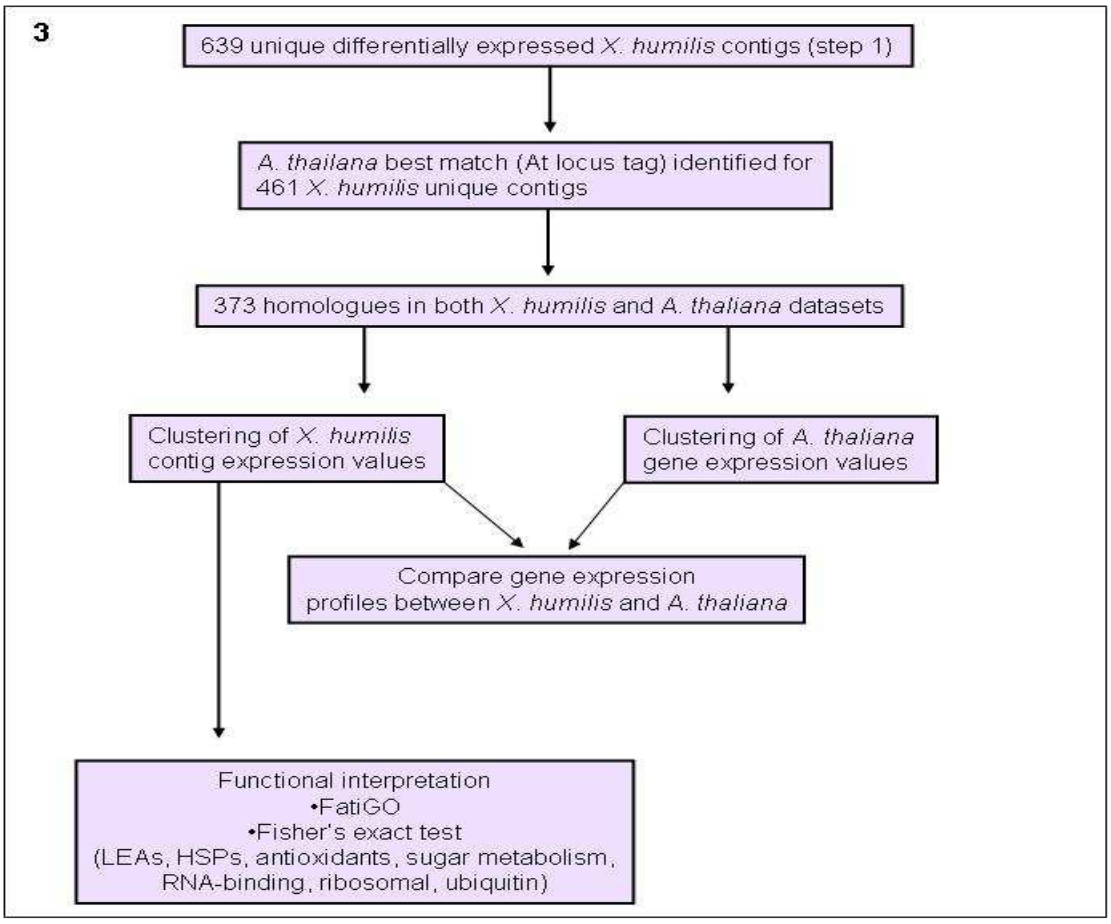


Figure 3.6 continued**.**

The At locus tags were matched with the contig numbers in the corresponding *X. humilis* clusters using vlookup in excel and the corresponding e-value and annotated gene name were retrieved (Fig. 3.6 step 3). The genes in the *X. humilis* clusters were compared with the genes in the *A. thaliana* osmotic treatment and seed clusters using vlookup in excel.

Gene selection and clustering of gene expression values for X. humilis and A. thaliana

The expression values in the original list of differentially expressed genes for *X. humilis* were exported to the unique contigs. The expression value of the contig for each experimental sample was taken as an average value of all cDNAs belonging to that contig for *X. humilis* (Fig. 3.6 step 1). The remaining cDNAs that were not sequenced (or sequencing failed) were excluded from

further analysis. Contigs that were differentially expressed between desiccated leaf and hydrated leaf and between desiccated root and hydrated root were identified for *X. humilis* using linear modelling in Limma (Fig. 3.6 step 2) with the contrasts summarized in Table 3.2. In *A. thaliana* differentially expressed genes were identified between stressed and control tissue in both the dehydration and mannitol stress treatments (Fig. 3.6 step 2; Table 3.2) and in all cases correction for multiple testing was implemented (Benjamini and Hochberg, 1995). For both *X. humilis* and *A. thaliana* the corresponding seed expression values were retrieved for the genes identified as differentially expressed between the vegetative tissues under the various stress conditions.

Table 3.2: Identification of differentially expressed genes during desiccation stress in *X. humilis* and during dehydration stress and osmotic stress in *A. thaliana*

X. humilis desiccation stress (n=3)	**A. thaliana dehydration stress 1hr (n=2)**	***A. thaliana osmotic stress 24 hr (n=2)***
DsL/HL	DhL/CL	OsL/CL
DsR/HR	DhR/CR	OsR/CR

The expression values of these genes were subjected to clustering based on condition (appendix a.3.2), using PCA (Wit and McClure, 2004) and DIANA (Fig. 3.6 step 2) with correlation as a measure of similarity (Kaufman and Rousseeuw, 1990). The biological replicates of the differentially expressed genes were averaged for each condition and the average gene expression patterns were clustered using PAMSAM (Fig. 3.6 step 2). The number of PAMSAM clusters was chosen based on the average silhouette width (a.s.w) of the resultant clusters for a particular value of K. K was chosen to maximize the a.s.w (Wit and McClure, 2004).

Functional interpretation of clusters

Contigs with an *A. thaliana* homologue that were present in both *X. humilis* and *A. thaliana* sets of data were included for functional interpretation of gene expression clusters (Fig. 3.6 step 3). FatiGO (http://fatigo.bioinfo.cipf.es/) was used to test for functional enrichment within clusters using At locus tags querying the *A. thaliana* database. Additionally, the expression values for all the conditions for these genes were retrieved and clustered using PAMSAM for both *X. humilis* and *A. thaliana*. The genes within the resulting clusters were compared between the species (Fig. 3.6 step 3). The number of genes encoding LEA family proteins, antioxidants, HSPs, ribosomal-

related proteins, RNA-binding proteins, ubiquitin-related proteins, sugar synthesis related proteins, photosynthesis related proteins was counted in both the *X. humilis* and *A. thaliana* clusters and a Fisher's exact test was applied to the count data (Fisher, 1962).

3.3. Results and discussion

Sequencing of differentially expressed cDNAs and annotation of contigs

Of the 3501 *X. humilis* cDNAs printed on the slide, only 424 had been sequenced and annotated previously (Collett et al., 2004). A total of 1619 cDNAs were identified as differentially expressed between DsL and HL, DsR and HR and between seed and hydrated vegetative tissue in *X. humilis* (Fig. 3.6 step 1). We selected these 1619 cDNAs for sequencing and annotation. It was found that this set of cDNAs included 164 of the 424 previously sequenced cDNAs (Collett et al., 2004) and therefore only 1455 cDNAs were sequenced de novo (from this study). An additional 823 cDNAs which were found to be constitutively expressed in this study were randomly selected for sequencing. As a result a total of 2278 cDNAs identified in this study were sequenced and 424 were sequenced by Collett et al. (2004) to give a total of 2702 sequences (Table 3.3). Of all these sequence reactions together, 643 cDNAs failed to produce sequence and 35 sequences were removed during the sequence processing stage to leave a total of 2024 sequences which clustered into 1468 unique contigs.

Table 3.3: Sequencing of *X. humilis* cDNA clones

	Differentially expressed cDNAs	**Constitutively expressed cDNAs**	*Total*
Printed cDNAs	1169	2332	3501
Sequenced previously (Collett et al., 2004)	164	260	424
Sequenced this study	1455	823	2278
Total sequenced			2702
Passed sequence			2024
Unique contigs			1468

The number of sequences per contig ranged from 1 to 58 but the majority of the contigs contained between 1 and 10 sequences each (Table 3.4) indicating a low level of redundancy in

the cDNA clones printed on the slides. This reflected the efficiency of the normalization procedure used to construct the original libraries (Collet et al., 2004).

Table 3.4: Number of sequences representing unique contigs.

No. of sequences	**<10**	**10-30**	*>30*
No. of unique contigs	2007	15	2

The 639 unique contigs representing the 1619 differentially expressed *X. humilis* cDNAs (Fig. 3.6 step 1) were preferentially annotated (Fig. 3.6 step 3). Individual blast searches were performed against the *A. thaliana* database for each translated sequence to identify the best match with an e-value cut off of $1x10^{-2}$. Contigs were annotated as a "no-match" if blastp did not identify an *A. thaliana* homologue or if the e-value was above the cut-off. The data is presented in the appendix (Table a.3.1). BlastP identified *A. thaliana* homologues with an e-value> $1x10^{-2}$ for 461 of the 639 differentially expressed *X. humilis* contigs. The e-values ranged from $9x10^{-3}$ to $2x10^{-168}$ with the majority of the identified homologues having an e-value greater than $1x10^{-10}$ which increased our confidence in the identity of the *A. thaliana* homologue. A number of the homologues with an e-value $<1x10^{-10}$ were putatively identified as LEA family members which are known to have regions of low complexity making it difficult to align these sequences with similar or related sequences. A number of stress-related proteins, LEA proteins, seed-specific and general antioxidants, ribosomal proteins and photosynthetic proteins were identified (Fig. 3.6 step 3; Table A.3.1).

Gene selection and clustering of differentially expressed genes for both X. humilis and A. thaliana

Cluster analysis was performed to investigate the similarity of gene expression profiles between the samples and to identify groups of co-expressed genes for each of the species. A different approach was implemented to identify differentially expressed genes to that performed previously for selecting genes to sequence (Fig. 3.6 step 2). This was because the differences between the design of the *X. humilis* and *A. thaliana* arrays had to be taken into account in setting up the screen for statistical analysis of differential gene expression. The first strategy (Table 3.1) was not applicable to allow a direct comparison between *A. thaliana* and *X. humilis*

gene expression profiles. This is because one could argue that the large number of *A. thaliana* genes that were differentially expressed between dehydrated tissue and mature seed could be a consequence of the full genome representation of the arrays, while the *X. humilis* arrays were constructed from cDNAs from normalized root and leaf libraries, and thus many seed cDNAs may have been missed in the comparison. Thus a comparison such as this between the *X. humilis* and *A. thaliana* datasets would not be balanced. A second strategy (Table 3.2) was devised to allow a more consistent comparison between the *X. humilis* and *A. thaliana* microarray datasets. Genes that were differentially expressed within vegetative tissues were first identified, and subsequently seed expression data associated with these genes was extracted (Fig. 3.6. step 2). This strategy revealed 344 unique contigs differentially expressed across *X. humilis* leaf and root samples, while in *A. thaliana* 87 genes and 9531 genes were identified as being differentially expressed between dehydrated tissues and mannitol treated tissues respectively. Cluster analysis of the differentially expressed genes explored the degree of overlap between DsL, DsR and seed samples in *X. humilis* and between OsL, OsR and seed samples in *A. thaliana*, based on gene expression values. The *A. thaliana* dehydration treatment was excluded from further analysis as only a small number of genes (87) were shown to change. This is most likely a consequence of the mild dehydration stress imposed (Killian et al., 2007).

Principal component analysis (PCA, Wit and McClure, 2004) was used to visualise the level of similarity between gene expression profiles across the different conditions (Fig. 3.6 step 2 and Fig. 3.7). The first two principal components indicate a remarkable similarity between desiccated vegetative tissues and seed in *X. humilis* (Fig. 3.7A). The x and y axis represent the distances between the sample matrices and as observed, the DsL, DsR and seed samples are considerably closer to each other than to the HL and HR samples, indicating a significant overlap in gene expression values among the desiccated tissues in *X. humilis*. However, the patterns observed for *A. thaliana* (Fig. 3.7B) was very different with a clear overlap of corresponding tissue types (e.g. CL and OsL, OsR and CR) instead of an overlap between the stressed vegetative tissue samples and seed samples (e.g. OsL and seed). Additionally, the seed expression profile appears to be significantly different from those of leaf and root in general. This indicates that gene

expression profiles of mannitol-stressed vegetative tissue were more similar to their corresponding hydrated counterparts than to mature seed in *A. thaliana*.

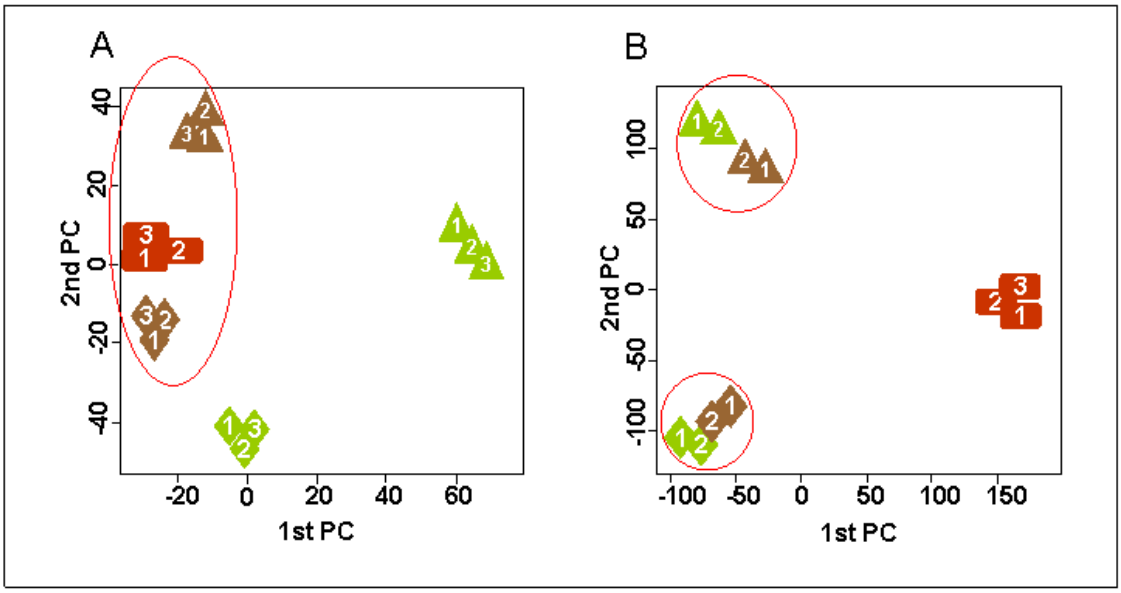


Figure. 3.7: Principal component analysis of differentially expressed genes from *X. humilis* and *A. thaliana* hydrated leaf and root, dehydrated/desiccated leaf and root, and seed. The axes represent the first and second principal components accounting for the most variation among the expression profiles. (A) The relationship between desiccated and hydrated vegetative tissue (DsL, DsR, HR, HL) and mature seed in *X humilis* based on 344 differentially expressed genes. Expression profiles of DsL, DR and are more similar to seed than to other samples. The first dimension explains 26% and the second dimension describes 18% of the observed variation. (B) The relationship between mannitol-stressed vegetative tissue and hydrated tissue (OsL, OsR, CL, CR) and seed in *A. thaliana* based on 9531 differentially expressed genes. The first dimension explains 20% and the second dimension describes 18% of the observed variation. In both *A. thaliana* treatments tissue types, rather than dehydrated samples, group together. DsL, OsL = ▲, HL, CL= ▲, DsR, OsR = ♦, HR, CR = ♦, Seed = ▬. The numbers represent the sample number for each treatment.

These results were verified with DIANA, an alternative, divisive hierarchical clustering algorithm (Kaufman and Rousseeuw, 1990). Dendrograms generated using DIANA (Fig. 3.8) show three clusters for both *X. humilis*, and *A. thaliana*. In. *X. humilis*, DsL, DsR and seed samples are most similar (Fig. 3.8A). Interestingly, HR samples are more similar to this desiccation cluster than the HL samples. In *A. thaliana* the pattern of clustering indicates a greater similarity between corresponding tissue types than between water-stressed leaf, root and seed (Fig. 3.8B) in agreement with the PCA results. These results indicate a significant overlap

in gene expression between DsL, DsR and seed in *X. humilis* but not between osmotically stressed vegetative tissues and seed in *A. thaliana*.

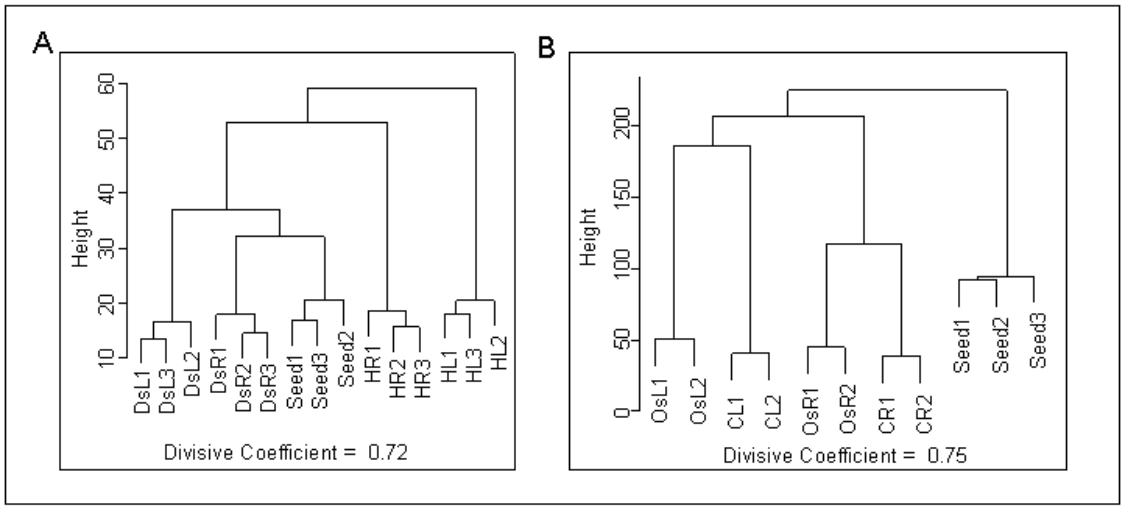


Figure 3.8: DIANA clustering results showing the relationships between differentially expressed genes from *X. humilis* and *A. thaliana* hydrated leaf and root, dehydrated/desiccated leaf and root, and seed. The stem represents the entire data set and the vertical coordinates, where a branch splits in two, equals the diameter of that cluster before splitting. The divisive coefficient (DC) for all three treatments indicates a strong clustering structure within the data sets. (A) For the *X. humilis* 344 differentially expressed genes, desiccated leaf and root samples (DsL, DsR) cluster more closely to mature seed than hydrated leaf or root (HL, HR). (B) The 9531 differentially expressed genes in the *A. thaliana*, mannitol-stressed leaves (OsL) and roots (OsR) cluster more closely to their respective controls (CL, CR) than to seed.

The patterns of gene expression was compared across the conditions to determine whether there are a significantly more genes expressed in DsL, DsR and seeds in *X. humilis* compared to dehydrated or osmotically stressed tissues and seeds in *A. thaliana* (Fig. 3.6. step 2). Gene expression profiles were clustered using PAMSAM which is a partitioning method that combines a k-means-type clustering method with a method to visualise the levels of similarity between the medoids of the clusters (Wit and McClure, 2004). Mclust (Fraley and Raftery, 2002a; 2002b) was used as a guide to choose the number of clusters. The number of clusters were chosen to minimise the a.s.w of the cluster. The smaller the a.s.w the more likely the data is to cluster into this number of groups.

Biological replicates were averaged and differentially expressed genes were clustered according to their expression profiles into 5 groups for the *X. humilis* desiccation treatment (Fig. 3.9A) and

7 groups for the *A. thaliana* mannitol treatment (Fig. 3.9B). In *X. humilis* (Fig. 3.9A) a cluster containing 82 genes showing an increase in gene expression in DsL, DsR and seed was clearly evident (cluster 3) and a second group contains 48 genes showing up-regulation in DsL and seed and constitutive expression in root tissue (cluster 4). This cluster indicates a possible constitutive protective mechanism present in root that is only activated in leaf during desiccation. Remaining clusters included a set of 59 genes that are specifically expressed in HR and HL (cluster 2), a group of 41 genes expressed in HR (cluster1) and 115 genes constitutively expressed in leaves (cluster 5). Notably there are no clusters of genes showing similar expression levels in hydrated vegetative tissues and seed.

A total of 7 clusters represents the *A. thaliana* mannitol/mature seed gene expression profiles (Fig. 3.9B) one of which contains a fairly large group of 1078 genes showing seed-specific expression (cluster 6). A large group of 1680 genes with increased levels of transcripts in vegetative tissue but not seed was also identified (cluster 4) while cluster 3 contains 1314 genes that are constitutively expressed in all tissues with a slight increase in transcript level in control vegetative tissues. Cluster 2 contains 1307 genes which show an increase in transcript abundance in OsL and root and a decrease in control leaf and seed. Cluster 5 (1468 genes) and cluster 7 (817) show a similar pattern of high levels of transcript across all conditions with a slight increase in OsL, OsR and seed in 5 and in OsL and OsR in 7. The expression patterns in these clusters are similar to those in *X. humilis* cluster 3 (Fig. 3.9A) but the increase in transcript abundance in the stressed vegetative tissue and seed is not as remarkable as observed for *X. humilis* and do not show as strong an overlap in between the stressed vegetative tissue and seed. Cluster 1 contains 1867 genes with a similar expression pattern to those in *X. humilis* cluster 5 showing an increase in transcript abundance in control leaf.

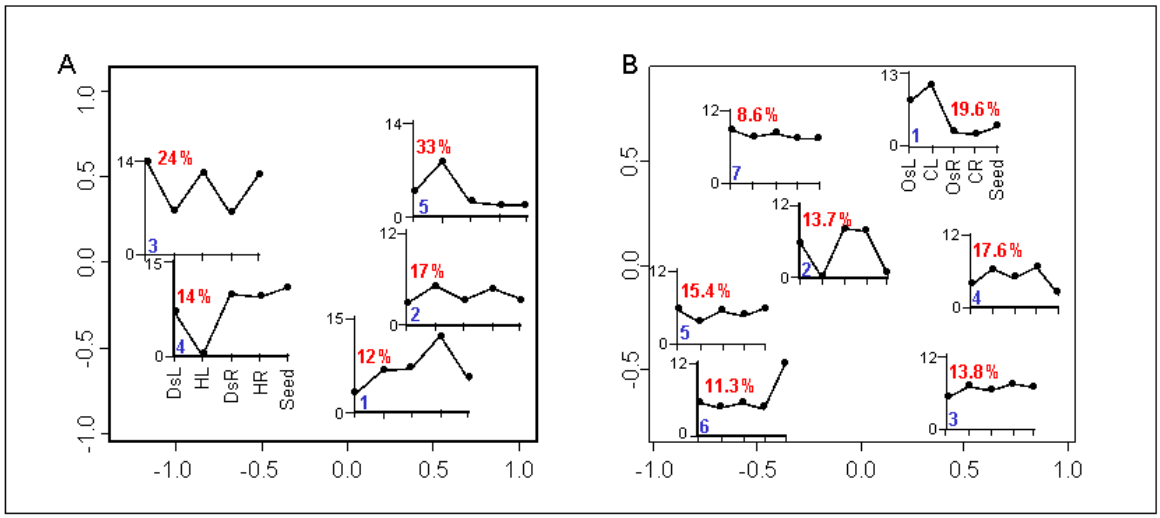


Figure 3.9: PAMSAM clustering of (A) 344 unique *X. humilis* genes which are differentially expressed between desiccated and hydrated vegetative tissue; and (B) 9531 *A. thaliana* genes which are differentially expressed between mannitol treated and untreated vegetative tissue. Each cluster of genes is represented by a small graph that shows the average value of log2 gene expression across the different conditions which are represented by a black circle. DsL=Desiccated leaf, HL=Hydrated Leaf, DsR=Desiccated Root, HR=Hydrated Root, OsL=Mannitol treated Leaf, CL=Control Leaf, OsR=Mannitol treated Root, CR=Control Root. The y-axis of the cluster graphs are the Log2 fluorescent intensity values. The x and y axis of the main graph represent distances between the medoids. Cluster number is shown in blue to the right on each graph and the percentage of genes represented in each cluster is indicated in red.

Significantly, *X. humilis* has a higher percentage of genes which show an overlap between stressed vegetative tissue and seed (cluster 3 and 4) compared to the corresponding *A. thaliana* gene expression clusters (clusters 7 and 8). *X. humilis* cluster 3 and 4 represents a total of 38% showing this overlap compared to the *A. thaliana* cluster 7 and 8 which represents 24%. Additionally, these *A. thaliana* gene expression clusters did not show a clear overlap in expression between OsL, OsR and seed, indicating that this may be a phenomenon unique to *X. humilis*. The *A. thaliana* mannitol/mature seed data set contains a cluster (cluster 2) that is expressed in OsL and constitutively in root and downregulated in seed. There are no clusters exhibiting this pattern of increased transcript levels in stressed vegetative tissue which are not also expressed in seeds, in *X. humilis*. Significantly, the *A. thaliana* osmotic stress data showed more of an overlap between all vegetative tissue (treated and stressed) and seed than *X. humilis* which showed a higher overlap with desiccated vegetative tissue and seed. Additionally, it even though seed-specific genes were not printed on the *X. humilis* microarray slides there are a

significant number of genes showing increased expression in seed tissue and that this overlap is only with the desiccated vegetative tissue and not the hydrated tissue.

Functional interpretation

Annotation of contigs that are differentially expressed

The hypothesis that DT in *X. humilis* arose from the activation of seed-specific genes in leaves and roots can be further explored by directly comparing the expression profiles of individual *X. humilis* and *A. thaliana* homologues. The question arises, is a gene that is expressed in seed and desiccated tissues in *X. humilis* only expressed in seed in *A. thaliana*? The group of 639 differentially expressed *X. humilis* genes identified via the first strategy (Table 3.1) were annotated and the At locus tag numbers, e-values and gene descriptions were recorded for those showing a significant match to an *A. thaliana* homologue (Fig. 3.6 step 3). Contigs which did not find a significant match (178 in total) were removed and a final set of 461 genes remained which included a number of *X. humilis* paralogues (contigs with the same At locus tags) (Table a.3.1). Of these only 373 were present in both the *X. humilis* and *A. thaliana* data sets and the expression values for these genes were averaged across biological replicates and clustered separately for each plant using PAMSAM. The expression profiles for these genes are shown for both *X humilis* (Fig. 3.10A) and *A. thaliana* (Fig. 3.10B) and in both cases the gene expression values clustered into 7 groups.

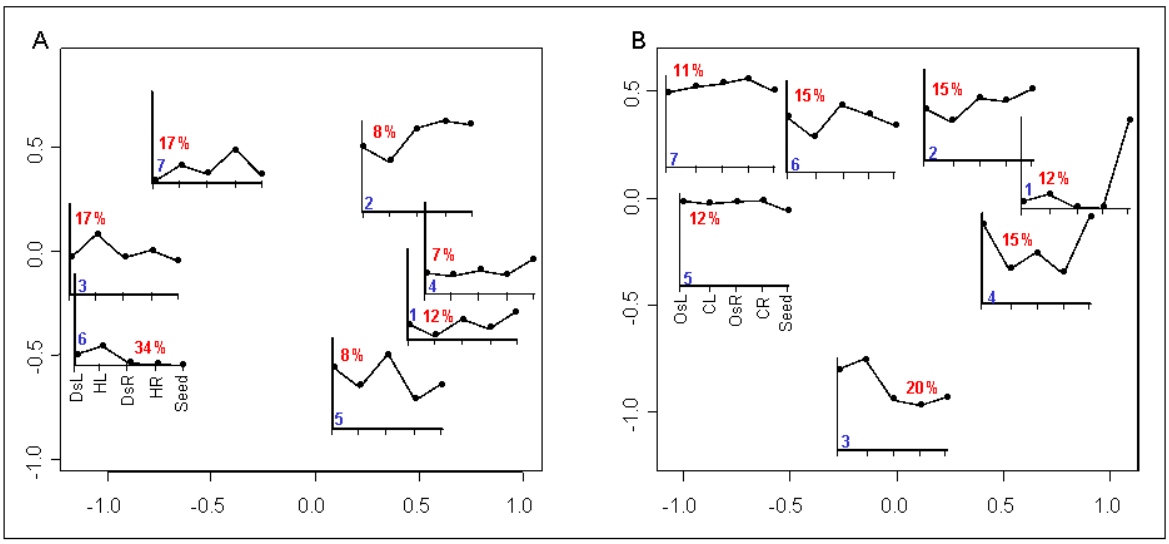


Figure 3.10: PAMSAM clustering of 373 genes which are differentially expressed between (A) desiccated and hydrated vegetative tissue and between hydrated tissue and mature seed in *X. humilis*; and (B) mannitol treated and untreated vegetative tissue and between hydrated tissue and mature seed in *A. thaliana*. Each cluster of genes is represented by a small graph that shows the average value of log2 gene expression across the different conditions which are represented by a black circle. DsL=Desiccated leaf, HL=Hydrated Leaf, DsR=Desiccated Root, HR=Hydrated Root, OsL=Mannitol treated Leaf, CL=Control Leaf, OsR=Mannitol treated Root, CR=Control Root. The y-axis of the cluster graphs are the Log2 fluorescent intensity values. The x and y axis of the main graph represent distances between the medoids. Cluster number is shown in blue to the right on each graph and the percentage of genes represented in each cluster is indicated in red.

A total of 4 clusters of genes (clusters 1, 2, 4 and 5) show an increase in transcript abundance in stressed vegetative tissue and seed in *X. humilis* (Fig. 3.10A) and they contain 44, 31, 27 and 31 genes, respectively. Cluster 3 and cluster 7, contain 63 and 51 genes, which show decreased levels of transcript in desiccated vegetative tissue and seed while cluster 6 includes 126 genes which have relatively higher amounts of transcript levels in hydrated leaf only. The cluster expression profiles are different in *A. thaliana* (Fig. 3.10B). While there are two groups, cluster 2 and 4 (56 and 55 genes) with increased levels of transcript abundance in osmotically stressed vegetative tissue and seed, there is an additional group of 57 genes with relatively higher amounts of transcript in stressed vegetative tissue and lower levels in seed (cluster 6). This pattern of gene expression is not observed in *X. humilis*. Another difference between the two species is the presence of 2 clusters (cluster 5 and 7) containing 45 and 40 genes that are constitutively expressed in vegetative tissue with a small decrease in transcript levels in seed in

A. thaliana which is not observed in *X. humilis*. Cluster 3, which includes 74 genes with increased transcript abundance in control leaf, is similar to cluster 6 in *X. humilis*. However, a striking difference between these two plants is the presence of a large number of seed-specific genes in *A. thaliana* (cluster 1) which is not in *X. humilis*. Significantly there is a higher representation of these seed-specific genes in clusters 1, 2, 4 and 5 in *X. humilis* than the remaining clusters (Fisher's exact test, 2 tail p=0.0006). These clusters contain genes that show an increase in transcript abundance in response to desiccation in the vegetative tissues and seed and this observed overlap supports the hypothesis that seed-specific genes are activated in response to desiccation in the vegetative tissues of *X. humilis*. These genes are listed in Table 3.5. This table includes 2 LEA family protein genes, namely *LEA3* (At3g53040) and *LEA6* (At3g22490). The *LEA6* gene has previously been shown to be seed-specific in *A. thaliana* (Illing et al., 2005) and both of these are examples of LEAs whose expression is restricted to the seed of the desiccation sensitive *A. thaliana* but are activated in the desiccated vegetative tissue of *X. humilis*. A 1-cys peroxiredoxin (At3g52960) is also included in this group and this seed-specific antioxidant (Aalen, 1999) was recently shown to be expressed in stressed vegetative tissue of another desiccation-tolerant plant, *X. viscosa* (Mowla et al., 2002). The transcript levels of this gene increases during the reacquisition of desiccation tolerance in *M. truncatula* seedlings and in desiccation tolerant embryos during seed maturation (Buitink et al., 2006). This antioxidant has also recently been suggested to have a role in sensing and/or reacting to seed environmental conditions to prevent germination under unfavourable conditions (Haslekas et al., 2003). The orthologues of two additional antioxidants (At4g10490 and At3g03070) which are only expressed in seeds of *A. thaliana* are also upregulated in response to desiccation in the vegetative tissues and seeds of *X. humilis*.

Table 3.5: *A. thaliana* seed-specific genes that are expressed in stressed vegetative tissue and seed of *X. humilis*

***X. humilis* cluster**	***Contig* number**	***At locus* tag**	***Gene annotated name***	***Best match e-value***
2	Contig831	At1g22600	unknown protein	2.00E-17
1	Contig1067	At1g27990	unknown protein	7.00E-65
1	Contig595	At1g47980	Desiccation related protein	3.00E-98
5	Contig628	At1g47980	Desiccation related protein	5.00E-100
2	Contig1064	At1g48130	1-cysteine peroxiredoxin; antioxidant	4.00E-78
2	Contig527	At1g48130	1-cysteine peroxiredoxin; antioxidant	5.00E-86
1	Contig690	At1g54870	Glucose and ribitol dehydrogenase	1.00E-113
1	Contig1309	At1g80090	SNF4 protein, CBS domain-containing protein,	3.00E-10
5	Contig862	At3g01570	glycine-rich protein / oleosin /oleosin5	4.00E-33
4	Contig1147	At3g02555	unknown protein	2.00E-07
4	Contig856	At3g03070	NADH-ubiquinone oxidoreductase-related	3.00E-33
5	Contig983	At3g07250	nuclear transport factor 2 (NTF2) family protein / RNA recognition motif (RRM)-containing protein	5.00E-17
1	Contig721	At3g22490	LEA6	2.00E-21
2	Contig734	At3g22490	LEA6	4.00E-54
2	Contig1007	At3g53040	LEA3	2.00E-06
2	Contig740	At3g53040	LEA3	1.00E-06
4	Contig1029	At4g10490	oxidoreductase, 2OG-Fe(II) oxygenase family protein, putative flavanone 3-beta-hydroxylase	1.00E-59
1	Contig522	At4g16160	protein translocase	5.00E-48
2	Contig989	At4g25140	OLEO1 (OLEOSIN1)	2.00E-28
2	Contig1136	At4g26740	XhS1 (Xerophyta humilis SEED GENE 1); calcium ion binding , embryo-specific protein 1 (XhS1); caleosin-related family protein	1.00E-38
5	Contig683	At4g27350	unknown protein	4.00E-79
2	Contig551	At5g24130	unknown protein	5.00E-18
1	Contig1001	At5g37770	XhTCH2 (TOUCH 2); CALMODULIN-RELATED PROTEIN (potential Ca(2+) sensor that is responsive to ABA signalling)	2.00E-13
5	Contig1296	At5g65570	putative pentatricopeptide	5.00E-79

Two genes encoding seed-specific oleosin proteins (At4g25140 and At3g01570) are also activated in the desiccated leaves and roots of *X. humilis*. Oleosins are found in oil bodies and are believed to be stored from earlier stages of seed development (seed filling). They are required to stabilize lipid bodies during extreme desiccation and they protect lipids from premature degradation (Kai and Huang, 2005). The expression of the genes which encode these proteins in the desiccated vegetative tissues of *X. humilis* suggest that lipid body stabilization may be necessary during extreme water deficit in these tissues. In general, the overlap in expression of

these seed-specific genes in the desiccated vegetative tissues of *X. humilis* points towards the acquisition of vegetative desiccation tolerance through the activation of seed-specific genes.

The At locus numbers were used to retrieve functional annotations of the genes from each of the clusters for both *X. humilis* and *A. thaliana* using FatiGO (Al-Shahrour, 2006, 2005, 2004). However, no significant terms were retrieved for any of the biological processes levels. This was assumed to be because the incomplete nature of the GO databases. For example it was noted that the majority of the LEA genes identified in our study were not assigned a GO term and as a result would not be included in the analysis. Therefore as an alternative approach to further our understanding of the expression patterns and the potential role of these homologues during desiccation and osmotic stress in *X. humilis* and *A. thaliana* respectively, the genes encoding LEA family proteins, HSPs, antioxidants, sugar metabolism related proteins, RNA-binding proteins, ribosomal-related and ubiquitin-related proteins, were compared between both species. The LEA, HSP, antioxidant and sugar metabolism related genes were investigated as they are known to have an important role in water deficit stress in seeds and vegetative tissue of both desiccation tolerant and sensitive species.

Genes encoding RNA-binding proteins and ribosomal-related proteins were included as it has been suggested that some desiccation tolerant systems, like the moss *Tortula ruralis*, are able to survive due to their ability to store a pool of mRNAs during drying that can be translated immediately following rehydration. It is thought that these mRNAs are stored as stable messenger ribonucleoprotein particles (mRNPs) (Wood and Oliver, 1999). This is a mechanism that has also been proposed for *X. humilis* as it has been shown that the transcripts for some of the genes encoding proteins involved in photosystem II (PSII), namely *PsbA* and *PsbP*, are stably stored at low levels during severe desiccation in *X. humilis* (Collett et al., 2003). Additionally, PSII function is able to recover partially without transcription (Dace et al., 1998) implying that transcripts are stored for immediate translation within the initial recovery phase in this plant.

The main function of ubiquitin, a small highly conserved protein found in all eukaryotes as either a free monomer or covalently linked to a variety of proteins, is to tag proteins for selective degradation by the 26S proteasome (Hasselgren and Fischer, 1997). Protein degradation is a normal cellular activity which increases during stress to remove damaged proteins from the cell in order to maintain cellular function (Ferguson et al., 1990).

The numbers of these genes present in *X. humilis* clusters 1, 2, 4 and 5 ("desiccated vegetative tissue and seed overlap" clusters) were compared with those in clusters 3, 6 and 7 ("no overlap" clusters). They were also counted and compared between *A. thaliana* clusters 2 and 4 ("osmotic stressed vegetative tissue and seed overlap" clusters) and clusters 3, 5, 6 and 7 ("no overlap" clusters) and in both cases a two-tailed Fisher's exact test was employed to test for significance (Fisher, 1950) (Table 3.6)

Table 3.6: Enrichment of transcripts encoding various protein families in stressed vegetative tissue and seed in *X. humilis* and *A. thaliana*. + indicates a significant representation of the genes within the "stressed vegetative tissue and seed overlap" clusters. The p-value is indicated next to it.

	X. humilis stress clusters	***A. thaliana stress clusters***
LEA proteins	+ ($p=3.7 \times 10^{-8}$)	nd
antioxidants	nd	nd
HSPs	+ (p=0.015)	nd
ribosomal-related proteins	nd	nd
RNA-binding proteins	nd	nd
ubiquitin-related proteins	nd	+ (p=0.02)
sugar metabolism related proteins	nd	nd

nd – no statistical significant difference

This table shows a significant enrichment of genes encoding LEA family proteins and HSPs in *X. humilis* in response to desiccation in vegetative tissue and seed. *A. thaliana*, however, only showed a significant enrichment of genes encoding ubiquitin related proteins in response to osmotic stress in vegetative tissue and mature seed. The expression of LEA and HSP genes have been shown to occur in response to drying during seed maturation (refer to chapter 1) and the enrichment of these genes in desiccated vegetative tissue and mature seed of *X. humilis*

underscores the activation of a seed-specific desiccation program in the vegetative tissue of resurrection plants. While some of the LEA genes are expressed in stressed *A. thaliana* vegetative tissue and seed (clusters 2 and 4) they were not significantly enriched within these clusters. This is as a result of those LEA genes which are only expressed in seed in this plant and thus the overlap in expression of these genes between stressed vegetative tissue and seed is not significant. The genes encoding 2 HSP family members were upregulated in response to desiccation while the remaining three in this data set were shown to be constitutively expressed or downregulated during stress in *A. thaliana*. This would suggest that while HSPs are involved in general stress in *A. thaliana* they appear to be important during desiccation in the vegetative tissue and seeds of *X. humilis*.

A significant enrichment of antioxidants was not observed for either *X. humilis* or *A. thaliana*. This is most likely due to the requirement of antioxidant activity during normal cellular activity and as a result they are expressed under a variety of conditions. It has been proposed that antioxidants are housekeeping enzymes and that they are protected from damage in desiccation tolerant species during extreme water deficit (Illing et al., 2005). This was suggested after the investigation of AP, GR and SOD enzymatic activity in desiccation tolerant *Eragrostis nindensis* compared to the desiccation sensitive *E. teff* and *E. curvula* during desiccation. While an initial increase in activity was observed for all three enzymes in all the species, the activity ceased in the sensitive plants below 50 and 40%, respectively. It would therefore seem that protection of antioxidant enzymes together with the expression of "specialized" seed-specific antioxidant genes (Table 3.5) may afford increased tolerance against water loss in desiccation tolerant angiosperms. This proposed protection of enzymes is supported by the increase in transcript abundance of genes encoding LEAs and HSPs (Table 3.6), which are known to protect protein from damage during stress. Also, the enrichment of genes encoding ubiquitin-related proteins in the desiccation sensitive *A. thaliana* suggests inadequate protection of proteins during water deficit stress in this species. These genes are not significantly represented as being up-regulated in *X. humilis* in response to stress.

Genes encoding proteins related to sugar metabolism were induced during stress in the vegetative tissue and seed in both species. However, they were not significantly enriched in response to water stress as a number of them were expressed constitutively in both plants. A significant difference between the two was the expression of a seed-specific sugar metabolism related gene (At1g54870) in the vegetative tissue and seed of *X. humilis* during desiccation (Table 3.5). These genes may show a more significant response during initial stages of drying as opposed to the later stages we are investigating. This is because sugar is thought to accumulate and function as a compatible solute early during drying and finally as a replacement for water when the bulk water is removed (Hoekstra et al., 2001). A measurement of the amount of accumulated sugars would give a more accurate assessment of this response in both species and their seeds to water loss.

It has been postulated that transcripts which are maintained in the dried state accumulate in the polysomal fraction and are associated with proteins as messenger ribonucleoproteins (mRNPs). Polysomal retention has already been shown for a number of genes including, *Tr288* (Wood and Oliver, 1999), *Rps3a* (Duff et al., 1999), *Rps14*, *Rps16* and *Rpl23* (Wood et al., 2000) in *T. ruralis*. Investigation into the genetic regulation of seed germination (Nakabayashi et al., 2005), seed after-ripening (Carrera et al., 2008) and seed dormancy release (Finch-Savage et al., 2007) has revealed a large number of stored mRNA species in mature, dry *A. thaliana* seed. These pools of mRNAs are believed to be reservoirs from embryogenesis and seed maturation and also provide for germination (Nakayabashi et al., 2005). This mechanism of RNA storage is proposed to allow a means of storing transcripts necessary for quick response upon rehydration. Bearing this in mind it was thought that there may be a higher representation of genes encoding RNA-binding proteins and ribosomal related proteins in response to desiccation in the seeds and vegetative tissue of *X. humilis*. However, this was not observed, possibly due to the housekeeping nature of these proteins. They are highly expressed under healthy hydrated conditions in both species and therefore do not appear to be significantly over represented during desiccation. But, as transcription and translation have slowed down in desiccated tissues the number of transcripts present to encode RNA-binding proteins and ribosomal-related proteins

could be sufficient for the formation of mRNPs to stably store mRNA species whose products are required early during the rehydration process. Analysis of these polysomal fractions from desiccated vegetative tissue and seeds of *X. humilis* would reveal if transcripts are stored during desiccation and if so, what proteins they encode.

Table 3.5 shows that we have identified a number of genes whose expression is restricted to seed in *A. thaliana* that are activated in response to desiccation in the vegetative tissue of *X. humilis*. This supports the proposed hypothesis that vegetative desiccation tolerance evolved from the adaptation of a developmentally regulated seed program in the vegetative tissue of resurrection angiosperms. In addition, Table 3.6 confirms that genes encoding protective proteins are expressed in the desiccated vegetative tissue and seed of *X. humilis* and this is not the case for the desiccation-sensitive *A. thaliana*. This suggests that even though the genes for other proteins required for desiccation tolerance may not increase in transcript abundance, the gene product itself may be protected allowing it to function or remain stable at lower RWC in *X. humilis*.

3.4. Conclusion

The expression levels of a total of 3501 cDNAs were analyzed in the microarray experiment, 2702 of which were sequenced and found to represent 1468 unique contigs. The level of similarity between gene expression profiles across the different conditions showed that DsL, DsR and seed are most similar to each other in *X. humilis* while a greater similarity between tissue types rather than between water-stressed leaf, root and seed is apparent in *A. thaliana*. These results suggests a significant overlap in gene expression between desiccated vegetative tissue and seed in *X. humilis* but not between dehydrated or osmotic stressed vegetative tissues and seed in *A. thaliana*. This was supported by comparison of the gene expression profiles which showed that *X. humilis* had two groups which contained genes whose expression was upregulated in desiccated vegetative tissue and seed. The *A. thaliana* clusters did not show a clear overlap in gene expression between stressed vegetative tissue and seed suggesting that the response to desiccation stress in *X. humilis* is different to the abiotic stress response of *A. thaliana*.

Of the 1468 unique contigs identified, 639 were differentially expressed between desiccated leaf, desiccated root and seed in *X. humilis*. These were annotated and the best *A. thaliana* match was identified. Of these genes, only those that were in both the *X. humilis* and *A. thaliana* data sets were compared. The expression values for these 373 genes were clustered and the expression profiles of the groups revealed a significant overlap between desiccated vegetative tissue and seed in *X. humilis*. Of particular interest was the identification of a set of seed-specific genes in *A. thaliana* and a number of these were found to be expressed in desiccated vegetative tissue and seed in *X. humilis*. These included a *Lea6* gene, previously found to be expressed in the vegetative tissue of *X. humilis* in response to desiccation (Illing et al., 2005) and a *Lea3* gene. A number of the *Lea3* family member genes were expressed in vegetative tissue and seed in both plants (At1g52690, At5g44310, At2g42560 and At4g15910) while only At3g53030 was seed-specific in *A. thaliana* and expressed in stressed vegetative tissue and seed of *X. humilis*. Genes encoding two-seed specific antioxidants were identified in addition to the seed-specific 1-cys periredoxin previously found to be expressed in desiccated vegetative tissue of *X. humilis*. In addition to these genes, two genes encoding seed-specific storage proteins were found to be expressed in the desiccated leaves and roots of *X. humilis*. The activation of these genes, whose expression is restricted to seeds in *A. thaliana*, points towards an overlap in the mechanisms employed during desiccation between mature orthodox seeds and desiccated vegetative tissue of *X. humilis*.

Assessment of functional enrichment of the clusters proved to be problematic as FatiGO failed to identify significant biological terms on any of the levels tested (data not shown). In an attempt to circumvent this we also clustered the data into smaller more specific groups of genes to test for significance, but this was not successful (data not shown). This apparent lack of significant representation of biological terms within clusters is most likely a consequence of the incomplete GO database, as it was noted that a number of LEA proteins did not have GO annotation data available for example. As a result an alternative approach was undertaken. This involved identifying genes encoding LEAs, antioxidants, HSPs, ribosomal-related proteins and sugar metabolism-related proteins and comparing their representation within the clusters for both

plants. Genes encoding LEAs and HSPs were significantly represented in *X. humilis* in response to desiccation and genes encoding ubiquitin-related proteins were over represented in *A. thaliana* in response to water loss. These results suggest that *X. humilis* and seeds employ similar mechanisms of protection during desiccation. *A. thaliana*, however, is not able to do this and as a result the only mechanism it shares in common with seeds is the removal of damaged proteins through the ubiquitin pathway. Overall these results provide support for the theory that desiccation tolerant angiosperms, like *X. humilis*, have adapted a seed-specific developmental program to survive vegetative desiccation.

Chapter 4

Real-time PCR Verification of Microarray Data

4.1. Introduction

Although care is taken to minimize noise at all stages of a microarray experiment, artifacts can still be introduced as a result of factors which are difficult to control. The presence of repeat sequences (e.g. polyA tails) and sequence motifs common to gene families can result in non-specific and cross hybridization giving rise to false signal detection. Additionally, genes with low expression levels can be difficult to detect as they are often close to that of background and expression of these genes must be confirmed. The background subtraction and normalization strategies employed can also adversely affect the expression values within a data set and therefore data from array-based experiments must be interpreted cautiously and should be verified by an independent means.

The expression levels of a set of genes from the *X. humilis* microarray dataset were confirmed by real-time PCR for all the conditions investigated. The expression data was obtained from boutique arrays which lacked control spots or "housekeeping" genes to assist with normalization. Additionally, these printed cDNA single channel arrays, are challenging to normalize as there is no reference channel available to adjust intensity levels within and across slides and also differences in spot morphology between slides can introduce variation. It was therefore necessary to perform real-time PCR to confirm the normalized expression levels of the genes identified as differentially expressed for *X. humilis* across the conditions tested. A number of genes from each of the clusters generated in chapter 3 were selected and real-time PCR was conducted to validate the expression profiles for the selected genes and to validate their assignment to their particular cluster. The average expression profile generated by PAMSAM to represent each cluster was confirmed at the same time.

Real-time quantitative PCR is a reliable and sensitive method for gene expression analysis and is the most rigorous and commonly used technology for the verification of microarray results. It

is a rapid, relatively inexpensive method requiring minimal starting template (Rajeevan et al., 2001; Walker, 2002) and it provides independent experimental verification of gene expression levels usually performed on the samples studied in the initial microarray experiment (Chauqui et al., 2002). Other applications of this technology include mRNA expression studies, genomic (Nigro et al., 2001; Ginzinger et al., 2000) or viral DNA copy number measurement (Desire et al., 2001), transgene copy number (Ingham et al., 2001) and allelic discrimination assays (Oliver et al., 2000; Walburger et al., 2001)

During real-time PCR the amount of PCR product generated in the exponential phase of the PCR process is measured and this is directly proportional to the amount of starting template (Schmittgen et al., 2000; Yin et al., 2001). The exponential phase of PCR is characterized by a doubling of product during each cycle if the reaction efficiency is close to 100% which is dependent on PCR conditions, primer design, template purity and amplicon length (Yuan et al., 2006). It is only possible to extrapolate back to determine the starting amount of template during the exponential phase as end point quantitation of PCR products is unreliable due to the presence of inhibitors in the template, reagent limitation or the accumulation of pyrophosphate molecules during the reaction. These accumulate to different amounts in the various PCR reactions and can result in either less or more product in the final stages of the reaction and will not accurately reflect the amount of starting template (Ginzinger et al., 2002). The amount of product accumulated is measured through the fluorescence of DNA binding dyes like SYBR green or DNA hybridization probes such as molecular beacons or Taqman probes (Bustin, 2000). SYBR green specifically binds double stranded DNA by intercalating between base pairs, and fluoresces only when bound to dsDNA. The fluorescent signal is detected during the PCR cycle at the end of the annealing or extension step when the greatest amount of dsDNA product is present and amplification plots show the cycle number versus fluorescence signal during the reaction. Real-time reactions measure when the amplification of a PCR product is first detected above background during the exponential phase and this is known as the threshold cycle (Ct) value (Fig. 4.1). The concentration of starting target is inversely proportional to the Ct value and the higher the concentration of target there is in the starting material, the lower the Ct. The

threshold must be in the linear region of the amplification plot representing the exponential accumulation of the PCR product (Nolan et al., 2006).

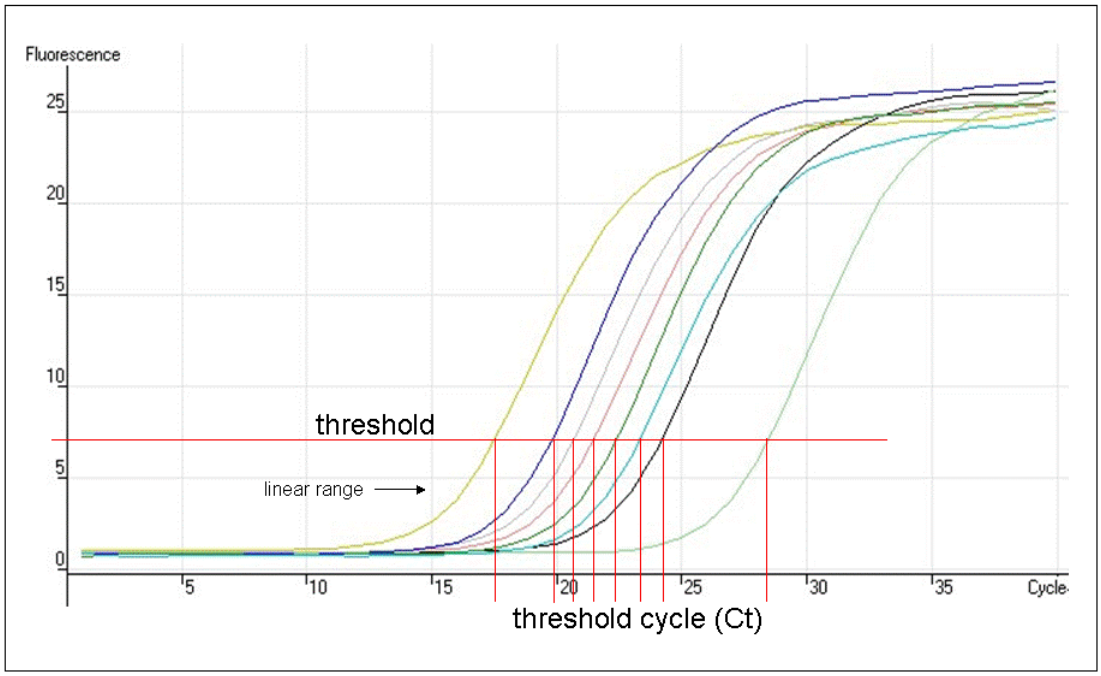


Figure 4.1. Amplification plot showing cycle number versus fluorescence measured for each reaction. The threshold cycle (Ct) for each reaction is recorded when its fluorescent reading increases significantly above the that of background

A standard curve can be generated using the Ct values of a sample of known concentration and usually a fivefold or tenfold serial dilution of the sample is used to generate the curve. The threshold cycle for the each of the standard curve reactions are plotted against the fold dilution of the template cDNA on a semi-logarithmic (base 10) plot and a linear trend line is determined and the slope and correlation factor (R^2) are calculated (Fig. 4.2).The experimental Ct values must lie in the dynamic range of the standard curve and this curve is then used to extrapolate relative level of expression of the genes of interest and the housekeeping gene for all experimental samples.

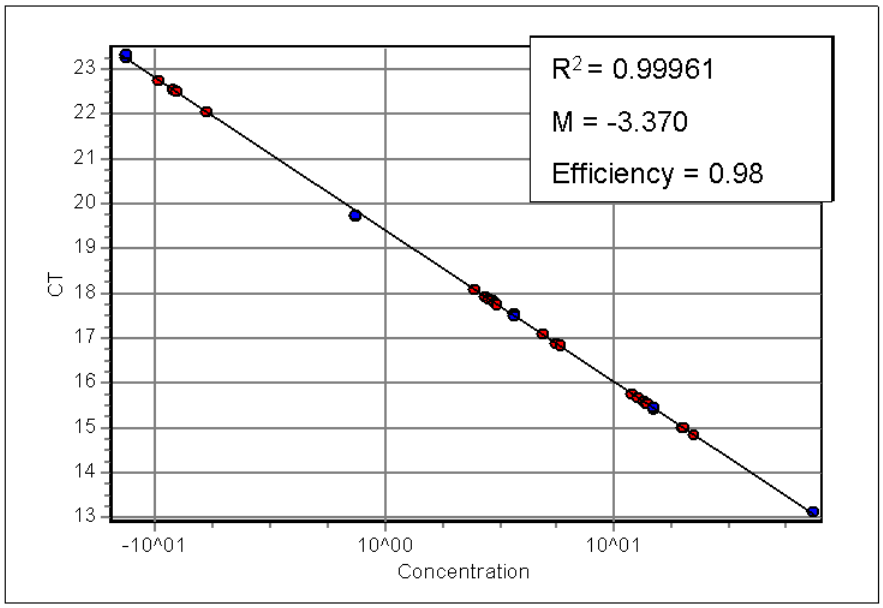


Figure 4.2. Plot of standard curve for calculation of PCR efficiency and sample quantitation. The threshold cycle for each standard curve reaction is plotted against its corresponding dilution factor. The concentration of the experimental samples is determined using their threshold cycles and the standard curve. Blue circles=standard curve samples, Red circles=samples of unknown concentration.

The standard curve can also be used for calculating the efficiency of the reaction using the resulting slope of the line fit to the data with the equation $Eff=10^{(-1/slope)} -1$. Ideally one would expect the amplicon concentration to double with each PCR cycle which would give an efficiency of 100% (or 1 in some software packages) and a slope of -3.323. However, a number of variables can affect the efficiency of the PCR including the length of the amplicon, presence of inhibitors, secondary structure within the template, $MgCl_2$ concentration, pipetting errors and primer design and concentration. An efficiency range of 80% to 120% (0.8 to 1.2) is generally regarded as acceptable in quantification of mRNA by real-time PCR and optimised assays generally have standard curves with a gradient between -3.1 and -3.323 (Nolan et al., 2006). For all the genes to be tested a constant annealing temperature is recommended so that all assays are run under the same PCR cycling conditions to allow comparison of efficiencies (Ginzinger et al., 2000). Reproducibility of the replicate reactions also reflects assay stability, with R^2 values of 0.98 or above indicating a stable reliable assay (Nolan et al., 2006).

In real-time PCR it is important to distinguish amplified DNA of the target gene from primer-dimers and non-specific amplification products. This can be done by plotting a melting curve of PCR products at the end of the run. Every dsDNA PCR product has a specific melting temperature (Tm) at which 50% of the DNA is single stranded. This temperature is dependent on both the length and the GC content of the PCR product can be used as a means to distinguish between desired product, amplification artifacts and primer dimers.

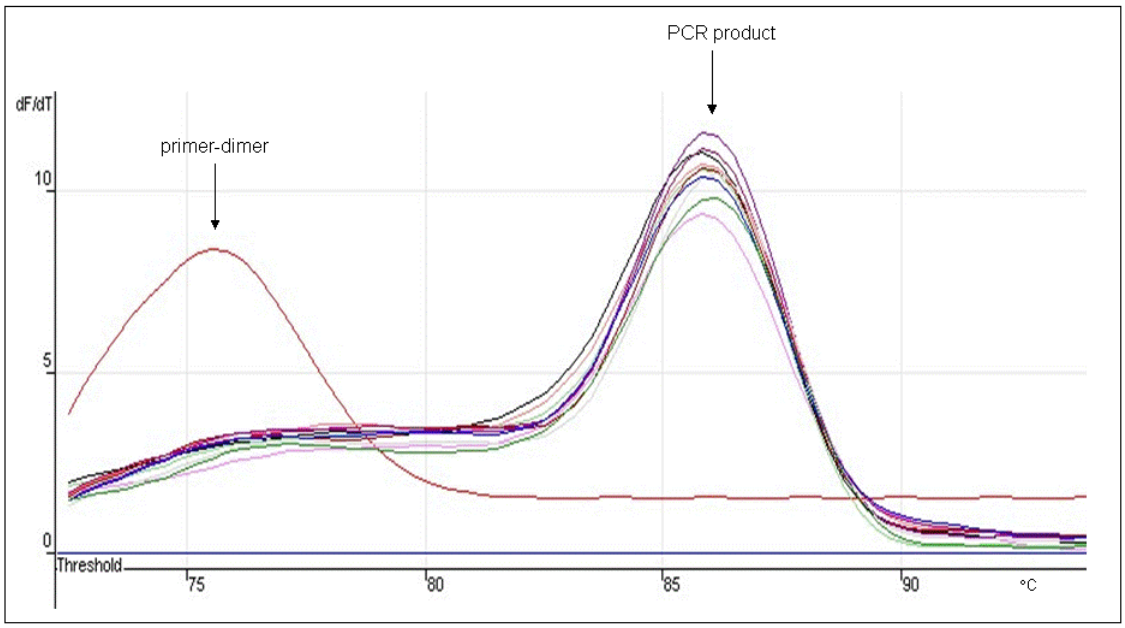


Figure 4.3. An example of first derivative melting curves. For each reaction shown on the plot only one peak is detected indicating that the reaction generates one specific product. The peak on the left (in red) indicates primer dimers detected in a reaction without template.

During a melt curve analysis the fluorescence of each sample is measured as it is heated from a temperature below the melting temperature of the products to a temperature above their melting point. The fluorescent dye (SYBR green) is released when the DNA product melts or dissociates, causing a sudden decrease in fluorescence when the Tm is reached thus providing accurate Tm data for every single amplified product. Melting peaks (Fig. 4.3) are calculated by taking the differential of the melt curve and they are the equivalent of the bands on an electrophoresis gel allowing a means to assess the products at the end of the run. Short primer dimers melt at lower temperatures than the longer target amplicon products. Single peaks indicate a single product and multiple peaks usually indicate multiple products (Nolan et al., 2006).

The most common analytic approaches to real-time PCR involve absolute or relative quantitation. Absolute quantification relies on an internal or external calibration curve to derive the input template copy number. This method is useful when the exact transcript copy number needs to be determined however in most cases relative quantification is sufficient (Yuan et al., 2006). Relative quantitation allows a comparison to be made between the expression of a gene of interest and a control gene within a sample (Pfaffl, 2001) by subtracting the cycle threshold (Ct) of the control gene from the Ct of the gene of interest. The exponent of the base2 (because of the doubling nature of the PCR reaction) of the difference in cycle number represents the fold difference of template for these two genes (Gingzinger et al., 2002).

In this study real-time PCR was used to independently verify the expression data of a selected number of genes from the *X. humilis* microarray experiments (chapter 2 and 3). A number of genes from each of the clusters discussed in chapter 3 were selected and the expression levels of mRNA transcripts were measured. The pattern of expression of mRNA transcript abundance between samples measured by real-time PCR correlated with the pattern of expression measured on the microarrays.

4.2. Materials and Methods

RNA samples

The same RNA samples from the microarray experiments were used for the real-time analysis, except for the hydrated root samples. These were freshly extracted from the same pool of tissue samples used for the microarray experiments. Two biological repeats from each condition were randomly chosen (Chapter 2) and included DL1, DL2, HL2, HL3, DR1, DR2, seed2, seed3. For hydrated root fresh RNA samples (HR1 and HR2) were extracted according to the RNA extraction protocol in chapter 2, representing 2 biological replicates unlike the microarray data.

Primer design

Gene-specific primers for selected genes were designed using Primer3 software (Rozen and Skaletsky, 2000) using the default settings. The primers were targeted to the 5' end of the cDNA sequence. This was to check that the reverse transcriptase had successfully transcribed full length cDNAs during the reverse transcription step of the reaction. They were designed to produce an amplicon between 100 and 150 bp in length and to have an annealing temperature of 60°C. The primers were all between 20 and 22 bp in length. Self complementarity was assessed in DNAman v4 (Lynnon Biosoft, 1997). Gene and primer sequences are shown in the appendix (A.4.1)

cDNA synthesis

Contaminating DNA was removed from RNA samples with DNase I (Ambion, USA) in a 50 µl reaction containing 10 µg RNA, 1x buffer, 1 U DNase I. The reaction was incubated at 37°C for 30 min after which 1/10 volume of 10 x inactivation buffer was added and incubated at room temperature for 2 mins with occasional mixing. The tube was centrifuged at 10 000 g for 1.5 mins and the supernatant containing the RNA was removed. The quality of the recovered RNA was assessed on a denaturing formamide gel. cDNA was synthesized from 5 µg of total DNA-free RNA. The RNA together with 500 ng oligodT primer in a volume of 10 µl was denatured at 65°C for 10 mins after which it was placed on ice and a final concentration of 1x First Strand Buffer, 50 mM DTT, 500 uM dNTP mix, 2 U RNAse Inhibitor, 10 U Superscript (Invitrogen, Germany) was added. The mix was incubated at 46°C for 1 hour after which an additional 10 U of Superscript was added and the reaction was further incubated at 46°C overnight. The reaction was stopped by incubation at 70°C for 15 mins and each sample was diluted 1/10 with RNase-free water. To test for genomic DNA contamination, equal aliquots of each RNA sample were mixed together to a final of 5 µg in an eppendorf and treated in the same manner above with the exception that Superscript was not added (i.e. –RT control).

Standard PCR

Standard PCR was performed to 1) optimize the concentration of primer pairs and 2) to confirm the size of the produced amplicon for each primer pair. Primer concentrations of 200 mM, 400 mM, 600 mM and 800 mM were checked for one of the genes, contig1463. All subsequent PCRs were performed using a concentration of 600 nM. The PCR reactions were performed in a volume of 25 µl containing 50 ng cDNA template, 1x PCR reaction buffer, 200 µM dNTP mix, 3 mM $MgCl_2$, 200 mM – 800 mM forward and reverse primers and 0.5 U Supertherm Taq polymerase (Southern Cross Biotechnology). The PCR was performed under the following conditions: 94°C for 3 mins, followed by 35 cycles of 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds and a final elongation at 72°C for 7 min. PCR products were electrophoresed on a 2% agarose gel and a FastRuler low range DNA ladder (Fermentas, Cape Town) was used as a molecular marker.

Standard curve

Equal aliquots of each cDNA were pooled, quantitated on a Nanodrop ND-1000 and serially diluted 1:5, 1:20, 1:100 and 1:1000 respectively to generate a standard curve. Each sample was aliquoted into single reaction volumes and stored at -80°C until required to avoid freeze-thaw induced degradation of cDNA.

Real-time PCR

Reactions were performed in a 25 µl volume containing 50 ng template, 1x SensiMix, 600 mM forward and reverse primer, 1 x SYBR Green (Quantace, London). The real-time PCR was performed on the Rotor-Gene 2000 Real-Time Cycler (Corbett Research, Sydney) using the following default parameters: Initial activation of enzyme at 95°C for 10 min, followed by 35 cycles of 95°C for 15 seconds, 60°C for 10 seconds and 72°C for 20 seconds. Melting curve analysis was performed at the end of each PCR run. Three technical repeats were amplified for each biological sample. A no-template and –RT were included as negative controls.

Calculations and Normalization

The real-time PCR software package (Roto-Gene 6, Corbett Research) was used to plot standard curves by plotting Ct values versus dilution. The concentrations of the experimental samples were calculated and log transformed. A housekeeping gene was used to correct for uneven amounts of cDNA template between samples. For this housekeeping gene the average and variance of the technical replicates of each sample for each sample was calculated (e.g. DL1, DL2). For the gene of interest the average value of the housekeeping gene for the corresponding sample was subtracted from each technical repeat value to give a normalized log ratio value for each technical repeat. The average and variance for all the normalized ratios (biological and technical repeats) for the gene of interest were then calculated. The exponential of this ratio is the normalized expression value for the gene for each of the conditions tested. The variance of the housekeeping gene and the variance of the replicates for the gene of interest were added together and the antilog calculated. The square root of this value is the standard deviation for the normalized expression value. All calculations were performed in excel.

4.3. Results and Discussion

Selection of genes

Real-time PCR quantification was performed to verify gene expression patterns from each of the clusters discussed in chapter 3. Genes (Table 4.1) were selected from the original 424 cDNA *X. humilis* geneset (Collett et al., 2004) as this was the only sequence information available at the commencement of the experiment. A total of 30 genes were originally selected to represent the clusters and two genes were selected as housekeeping genes to correct for differences in cDNA concentrations between the samples (Table 4.1). Contig1360, Contig1367, and Contig1463 were selected as potential housekeeping genes as they showed little change in expression across all conditions in the microarray experiments (Fig. 4.4).

Table 4.1: A list of genes selected from the microarray analysis for verification by real- time PCR

XH contig number	*annotated gene name*	*XH cluster*	*PCR pass*	*Real-time pass*	*R^2*	*slope (M)*	*Efficiency*	*Appendix reference*
Contig1046	LEA7	3	Y	Y	0.99692	-3.545	0.91	A.4.2.1
Contig1169	No match	3	Y	Y	0.99528	-3.301	1.01	A.4.2.2
Contig1360	Origin recognition complex subunit like	hkg	N	N				
Contig1367	Unknown Protein	hkg	Y	Y	0.99289	-3.285	1.02	A.4.2.3
Contig1463	No Match	hkg	Y	Y	0.99146	-3.231	1.04	A.4.2.4
Contig1468	No match	3	Y	Y	0.9816	-3.546	0.91	A.4.2.5
Contig1480	basic chitinase	1	Y	Y	0.99608	-3.5	0.93	A.4.2.6
Contig1561	OEC-33, photosystem II,	5	Y	Y	0.99888	-3.692	0.87	A.4.2.7
Contig1567	Unknown Protein	2	Y	Y	0.99801	-3.176	1.06	A.4.2.8
Contig23	senescence-associated protein	2	Y	Y	0.99899	-3.408	0.97	A.4.2.9
Contig348	psbY, chloroplast precursor	5	Y	Y	0.99607	-3.354	0.99	A.4.2.10
Contig408	tubulin beta-1	1	N	N	0.18416	-0.356	643.09	
Contig498	PSAF (photosystem I subunit F)	5	Y	Y	0.99764	-3.477	0.94	A.4.2.11
Contig531	histone H3.2	2	Y	Y	0.99449	-3.312	1	A.4.2.12
Contig541	PIP1 aquaporin	5	Y	Y	0.99701	-4.123	0.75	A.4.2.13
Contig580	Glyoxylase-I like	4	Y	Y	0.99874	-3.183	1.06	A.4.2.14
Contig617	No match	4	Y	Y	0.99928	-3.221	1.04	A.4.2.15
Contig656	Galactinol synthase, GolS-1	5	Y	Y	0.99864	-3.371	0.98	A.4.2.16
Contig693	LEA2	4	Y	Y	0.99769	-3.307	1.01	A.4.2.17
Contig701	LEA3	4	Y	Y	0.98973	-3.483	0.94	A.4.2.18
Contig721	LEA6	3	Y	N	0.98141	1.647	-0.75	A.4.2.19
Contig727	RNA-binding region RNP-1	5	Y	Y	0.99345	-3.661	0.88	A.4.2.20
Contig760	No Match	3	N	N				
Contig766	LEA10	4	Y	Y	0.99915	-3.821	0.83	A.4.2.21
Contig772	Metallothionein	3	Y	Y	0.99675	-3.305	1.01	A.4.2.22
Contig813	Nitrate transporter	3	Y	N	0.16996	-1.077	7.48	A.4.2.23
Contig859	No Match	4	Y	Y	0.9989	-3.368	0.98	A.4.2.24
Contig864	LEA3	3	Y	Y	0.98966	-3.501	0.93	A.4.2.25
Contig865	LEA8	3	Y	Y	0.99791	-3.569	0.91	A.4.2.26
Contig915	LEA10	4	Y	Y	0.99764	-3.3403	0.97	A.4.2.27
Contig959	No Match	4	Y	Y	0.9959	-3.645	0.88	A.4.2.28
Contig977	protein kinase	3	Y	Y	0.99778	-3.353	0.99	A.4.2.29
Contig989	OLEOSIN1	4	Y	Y	0.99731	-3.615	0.89	A.4.2.30

hkg – housekeeping gene.
Contig number – unique ID assigned to clones.
Annotated gene name - Name assigned to gene from BLASTX results.
XH cluster number – Microarray cluster to which the gene was assigned by pamsam (chapter 3).
PCR pass – Based on the production of a band of the correct size by standard PCR.
Real-time pass – Based on the melting curve analysis and efficiency of the real-time PCR reaction.

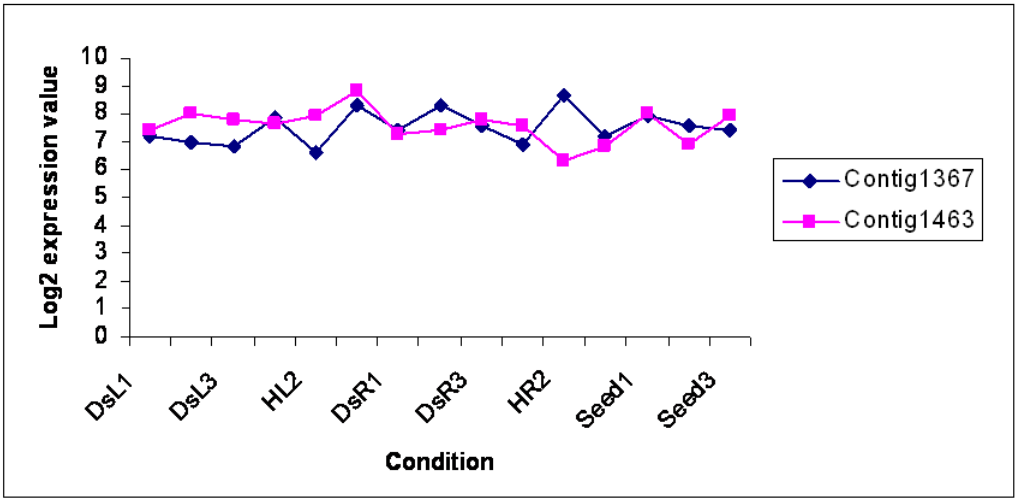


Figure 4.4: Microarray Log_2 expression values of selected potential housekeeping genes across the conditions tested in the microarray analysis.

Optimization of PCR conditions

All primer pairs were tested on cDNA made from an equal mix of RNA from all the conditions to be assayed. The primer concentration was optimized first and a range of concentrations from 200 nM to 800 nM was tested using primers for Contig1463 (one of the housekeeping genes) using standard PCR conditions. Figure 4.5 shows that 600 nM gave the best product. This primer concentration was used for the remaining primer sets and was not tested by standard PCR but the efficiencies were monitored for all the primer sets in the real-time analysis. Additionally, in fig 4.5 it can be seen that the negative RT control produced a band of the same size as the expected product (lane 2 for each primer concentration) but the water control was negative thus indicating a possible genomic DNA contamination of the RNA samples. All the RNA samples were subsequently treated with DNaseI and a representative few primers were used to confirm the removal of genomic DNA contamination by standard PCR (fig 4.6). In this figure there is no indication of a band in the sample that did not contain cDNA template (lanes 2 for each primer set) showing that the DNAse treatment had removed genomic DNA contamination. Standard PCR was performed on all the primer sets (data not shown) and those showing amplification of an amplicon of the expected size were selected to continue with real-time analysis (summarized in Table 4.1).

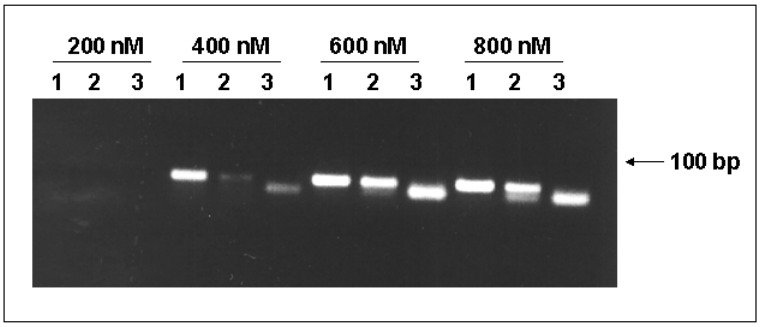


Figure 4.5: Agarose gel of standard PCR to optimize primer concentration for real time PCR using Contig1463 primers. Lane 1, cDNA template; lane 2, -RT control; lane 3, water control; lane Fragment sizes were determined using a FastRuler low range DNA ladder. Primer concentrations tested are indicated above the lane numbers. –RT, cDNA synthesis reaction excluding reverse transcriptase.

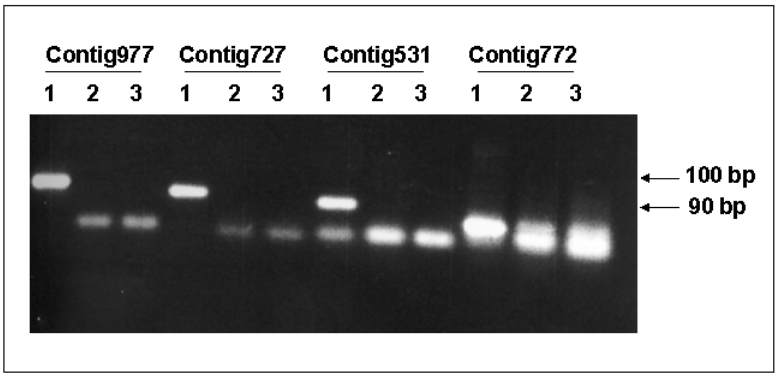


Figure 4.6: Agarose gel of standard PCR to test for the presence of genomic DNA using primers for 5 different contigs. Lane 1, cDNA template; lane 2, -RT control; lane 3, water control. Fragment sizes were determined using a FastRuler low range DNA ladder. The primer pairs used are indicated above the lane numbers. –RT, cDNA synthesis reaction excluding reverse transcriptase.

Contig1463 was also used to check that the cDNA synthesis reactions were successful across all samples (Fig. 4.7) in standard PCR reaction. An amplicon of the correct size was amplified in all the samples showing that cDNA was produced for all the samples to be tested. These cDNA samples were diluted 1/10 for real-time assays. A cDNA pool for the standard curve was generated by mixing equal aliquots of each of the diluted samples. The concentration of this pool was determined and then serially diluted 1:5, 1:20, 1:100 and 1:1000.

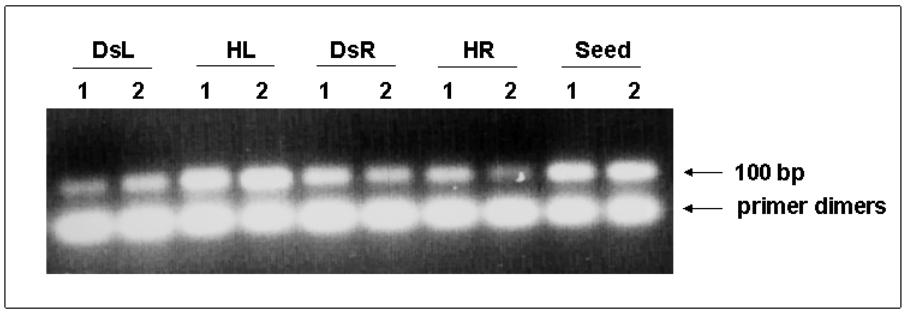


Figure 4.7: A 100 bp product was amplified from all cDNA samples using primers for Contig1463. 1) Biological replicate 1, 2) Biological replicate 2. Fragment sizes were determined using a FastRuler low range DNA ladder. DsL - Desiccated Leaf, HL – Hydrated Leaf, DsR – Desiccated Root, HR – Hydrated Root.

Quality check of the real-time PCR results

Quantitative PCR was performed to verify the *X. humilis* microarray data and to independently quantify transcript levels of representative genes from each of the five *X. humilis* gene clusters (Table 4.1). Amplification profiles, standard curves and melting curves were plotted for all selected genes (A.4.2). Values for the standard curve regression, slope and efficiency of each reaction are summarized in Table 4.1. With the exception of Contig721 and Contig813 the amplification plots for each gene show that the amplicons were amplified and detected between cycles 15 and 30. The primers for Contig721 and Contig813 failed to amplify a product and as a result these genes were excluded from further analysis. The melting curves for each of the remaining genes, except Contig408 and Contig865, show one clear peak across the samples. This shows that the primer pairs used in each case were specific and resulted in the amplification of a single product. The multiple peaks for Contig408 and Contig865 indicate the presence of multiple bands and possible contamination and these genes were excluded from further analysis. A peak on the left of the main peak is observed for Contig 617, Contig864 and Contig959 and this represents the presence of primer-dimers within the reaction. The efficiencies for these reactions were within the acceptable range, and these genes were included in further downstream analysis.

The R^2 values for the standard curves of the remaining genes are all 0.98 and above indicating a reliable and stable amplification and therefore an accurate curve to allow extrapolation of

starting template amounts. The concentrations of the samples tested all fell within the range of the standard curve except for those that were known not to be expressed within those tissues from the microarray data (A.4.2). Contig1169, Contig617, Contig772, Contig864, Contig915 and Contig959, are not expressed in HL. Contig498, Contig541 and Contig656, Contig1561are not expressed in seed, HR and DR. Contig348 is not expressed in DR and seed and Contig727 is not expressed in HR and seed. These correlated with the microarray data and amplification at higher concentrations did not change the calculated value. Additionally, the efficiency was consistently between 0.9 and 1 and the slopes were mainly between -3.1 and -3.6 indicating good amplification efficiency (Table 4.1). The efficiency was a little below the recommended level (Nolan et al., 2006) for Contig541, Contig727, Contig766, Contig959 and Contig1561, however, they still showed good correlation with the microarray data (discussed below). Those genes with single peaks in the melting curve analysis (A.4.2), R^2 values of at least 0.98 and efficiencies between 0.8 and 1 were used for verification of the microarray data (Table 4.1).

Real-time Verification of Microarray Data

Contig1463 was selected as a housekeeping gene to normalize the expression level of each gene. It was chosen as it showed no change in expression levels across conditions in the microarray experiments (Fig 4.4) and the real-time amplification efficiency was close to 100% (Table 4.1). Contig1367 was excluded as a possible housekeeping gene. Although the amplification efficiency was high some of the Ct values were not covered in the dynamic range of the standard curve and as a result an accurate concentration could not be calculated for these samples. Furthermore the melting curve was noisy. The expression level for each gene was normalized by dividing the expression level for that gene by the expression level Contig1463.

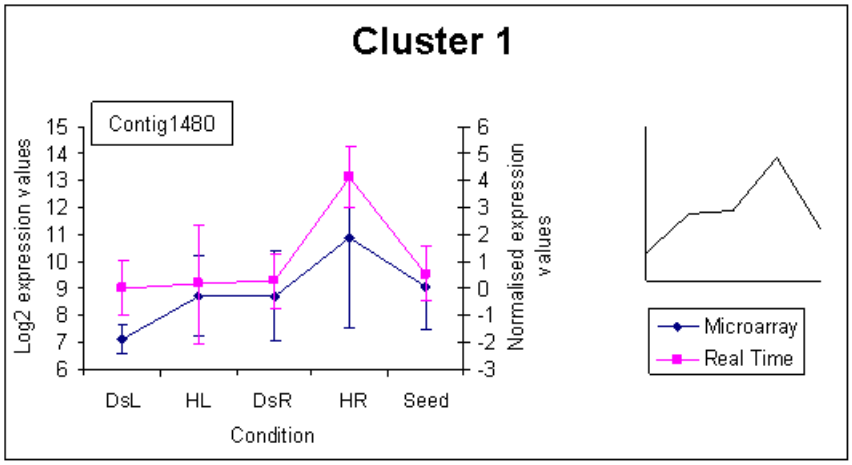


Figure 4.6: Comparison of microarray and quantitative RT-PCR gene expression values for a representative gene from *X. humilis* gene cluster 1. Contig1480 from cluster 1 ($n_{microarray}$=12). Transcript levels were assessed in leaves and roots from hydrated and desiccated *X. humilis* plants and mature seed. The microarray data (blue) is reported as $\log_2$ (fluorescent intensity) while the real-time PCR data (pink) is reported as ΔC_T normalized values. The real-time data represents the mean of triplicate real time PCR reactions from two different biological samples. The ΔC_T values are calculated as follows: ΔC_T (target gene) = ΔC_T (target gene) - ΔC_T (housekeeping gene; Contig1463). Log2 expression values of microarray data is shown on the y axis on the left hand side. Normalized expression values of real-time PCR data are shown on y-axis on the right hand side. Error bars represent ± standard deviation. The insert shows the profile of cluster 1 of the microarray data.

The genes in cluster 1 belong to a group whose expression is upregulated in HR. The real-time data confirmed the microarray expression profile for Contig997 (figure 4.6) reflecting the increase in expression in HR. A second gene, Contig408, failed the quality check of the real-time data. Cluster 2 contains genes that are upregulated in HL and HR in the microarray experiments (Fig. 4.7). The real-time expression data of Contig23, Contig531 and Contig1567 correlated well with the microarray data showing the increase in gene expression in HL and HR. This confirmed the microarray expression levels and validated the assignment of these genes to the group with this average expression profile.

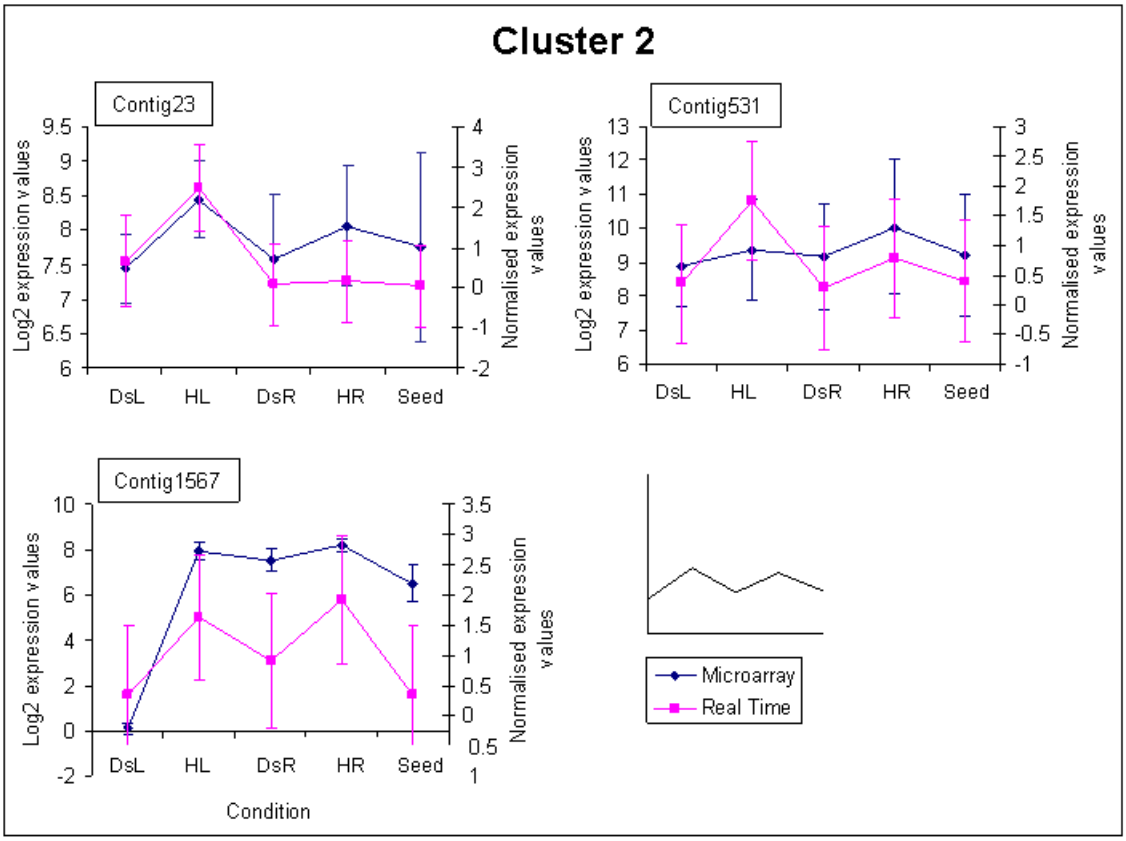


Figure 4.7: Comparison of microarray and quantitative RT-PCR gene expression values for representative genes from *X. humilis* gene cluster 2. Contig23 ($n_{microarray}$=18), Contig531 ($n_{microarray}$=18), Contig1567 ($n_{microarray}$=6). Transcript levels were assessed in leaves and roots from hydrated and desiccated *X. humilis* plants and mature seed. The microarray data (blue) is reported as $\log_2$ (fluorescent intensity) while the real-time PCR data (pink) is reported as ΔC_T normalized values. The real-time PCR data represents the mean of triplicate real time PCR reactions from two different biological samples. Error bars represent ± standard deviation. The ΔC_T values are calculated as follows: ΔC_T (target gene) = ΔC_T (target gene) - ΔC_T (housekeeping gene; Contig1463). Log2 expression values of microarray data is shown on the y axis on the left hand side. Normalized expression values of real-time PCR data are shown on y-axis on the right hand side. Error bars represent ± standard deviation. The insert shows the profile of cluster 2 of the microarray data.

The microarray data shows an increase in mRNA transcript levels in DsL, DsR and seed for genes in cluster 3 (Fig. 4.8) and this trend is also reflected in the real time data for all seven of the genes tested. The correlation between the microarray expression data and the real-time data confirms the observed expression pattern and the membership of these genes within this particular cluster.

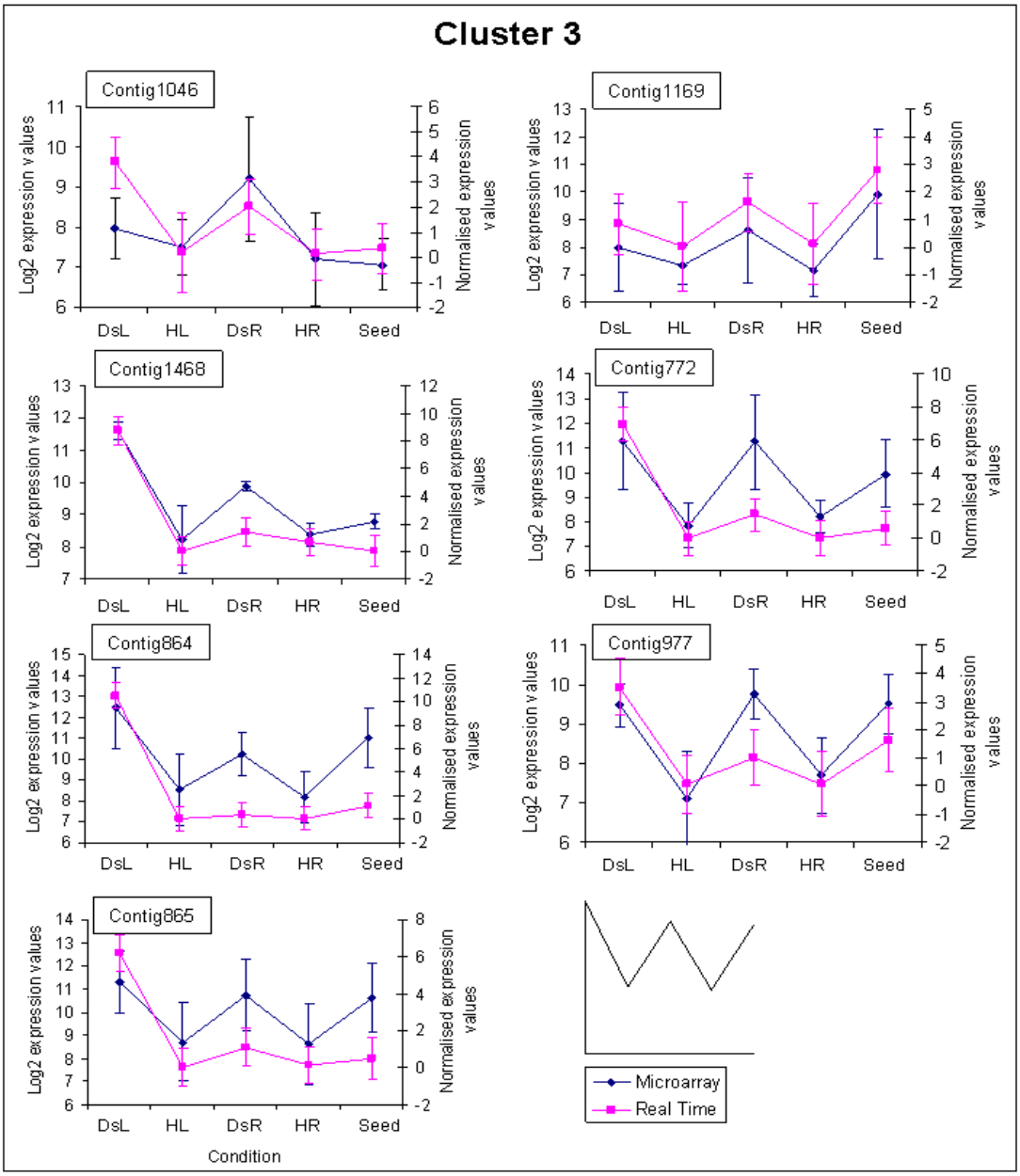


Figure 4.8: Comparison of microarray and quantitative RT-PCR gene expression values for representative genes from *X. humilis* gene cluster 3. Contig1046 ($n_{microarray}$=27), Contig1169 ($n_{microarray}$=18), Contig1468 ($n_{microarray}$=6), Contig772 ($n_{microarray}$=105), Contig864 ($n_{microarray}$=150), Contig977 ($n_{microarray}$=18), Contig865 ($n_{microarray}$=27). Transcript levels were assessed in leaves and roots from hydrated and desiccated *X. humilis* plants and mature seed. The microarray data (blue) is reported as $\log_2$ (fluorescent intensity) while the real-time PCR data (pink) is reported as ΔC_T normalized values. The real-time PCR data represents the mean of triplicate real time PCR reactions from two different biological samples. Error bars represent ± standard deviation. The ΔC_T values are calculated as follows: ΔC_T (target gene) = ΔC_T (target gene) - ΔC_T (housekeeping gene; Contig1463). Log2 expression values of microarray data is shown on the y axis on the left hand side. Normalized expression values of real-time PCR data are shown on y-axis on the right hand side. Error bars represent ± standard deviation. The insert shows the profile of cluster 3 of the microarray data.

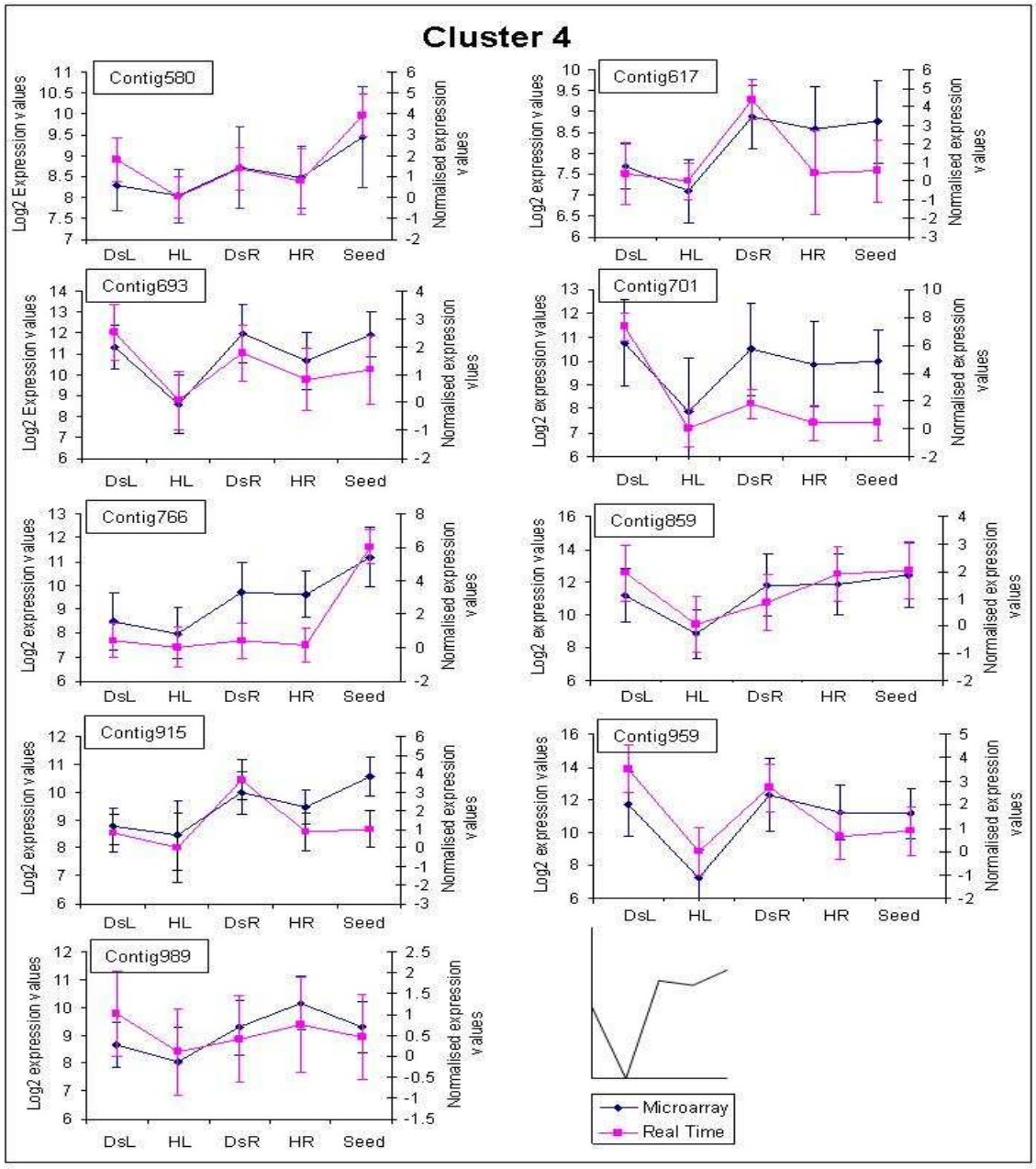


Figure 4.9: Comparison of microarray and quantitative RT-PCR gene expression values for representative genes from *X. humilis* gene cluster 4. Contig580 ($n_{microarray}$=45), Contig617 ($n_{microarray}$=36), Contig693 ($n_{microarray}$=102), Contig701 ($n_{microarray}$=42), Contig766 ($n_{microarray}$=93), Contig859 ($n_{microarray}$=342), Contig915 ($n_{microarray}$=30), Contig959 ($n_{microarray}$=60), Contig989 ($n_{microarray}$=69). Transcript levels were assessed in leaves and roots from hydrated and desiccated *X. humilis* plants and mature seed. The microarray data (blue) is reported as $\log_2$ (fluorescent intensity) while the real-time PCR data (pink) is reported as ΔC_T normalized values. The real-time PCR data represents the mean of triplicate real time PCR reactions from two different biological samples. Error bars represent ± standard deviation. The ΔC_T values are calculated as follows: ΔC_T (target gene) = ΔC_T (target gene) - ΔC_T (housekeeping gene; Contig1463). Log2 expression values of microarray data is shown on the y axis on the left hand side. Normalized expression values of real-time PCR data are shown on y-axis on the right hand side. Error bars represent ± standard deviation. The insert shows the profile of cluster 4 of the microarray data.

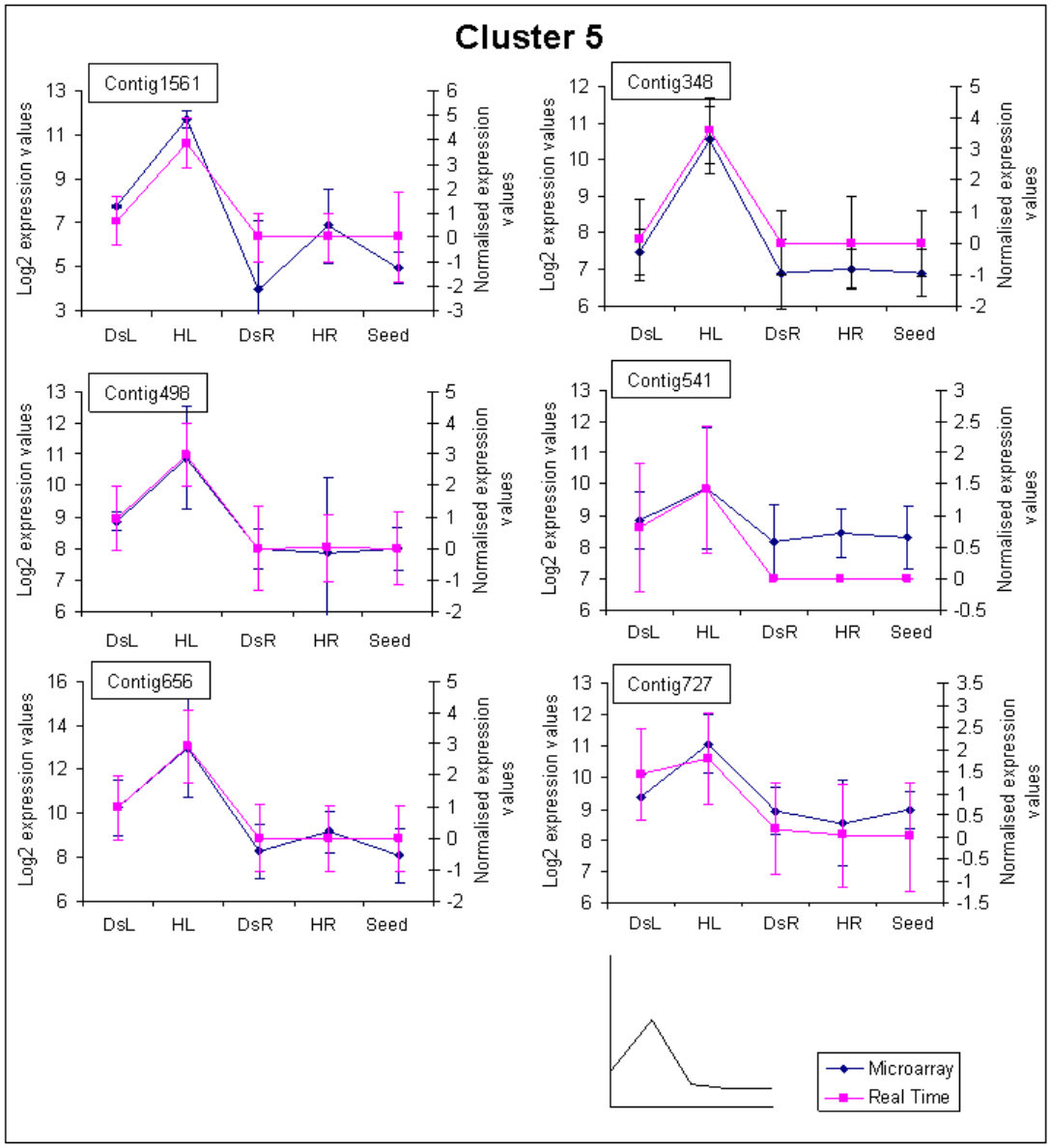


Figure 4.10: Comparison of microarray and quantitative RT-PCR gene expression values for representative genes from *X. humilis* gene cluster 5. Contig1561 ($n_{microarray}$=6), Contig348 ($n_{microarray}$=18), Contig498 ($n_{microarray}$=18), Contig541 ($n_{microarray}$=48), Contig656 ($n_{microarray}$=78), Contig727 ($n_{microarray}$=3). Transcript levels were assessed in leaves and roots from hydrated and desiccated *X. humilis* plants and mature seed. The microarray data (blue) is reported as $\log_2$ (fluorescent intensity) while the real-time PCR data (pink) is reported as ΔC_T normalized values. The real-time PCR data represents the mean of triplicate real time PCR reactions from two different biological samples. Error bars represent ± standard deviation. The ΔC_T values are calculated as follows: ΔC_T (target gene) = ΔC_T (target gene) - ΔC_T (housekeeping gene; Contig1463). Log2 expression values of microarray data is shown on the y axis on the left hand side. Normalized expression values of real-time PCR data are shown on y-axis on the right hand side. Error bars represent ± standard deviation. The insert shows the profile of cluster 5 of the microarray data.

The average expression profile of cluster 4 shows increased mRNA transcript levels in DsL, DsR, HR and seed (Fig. 4.9). The real-time and microarray expression profiles were comparable for all of the 9 genes tested. The correlation between the real-time and microarray data confirmed the membership of these genes in this cluster and the expression profile of this group. Finally, the transcript levels of the genes in cluster 5 increase in abundance in HL. This pattern was observed for both the microarray and real-time PCR experiments for all five of the genes tested (Fig. 4.10) confirming their expression pattern and their assignment to this group.

In all the cases presented above the error bars for the microarray data and the qRT-PCR data are large and this is most likely due to the biological variation between the samples. A total of three biological replicates, each with a number technical replicates were averaged to represent the microarray data. Variation may result as a consequence in differences in age of leaves and roots and small differences in RWC of the tissues. In addition, variation may also arise as a consequence of the difficulties associated with the normalization of 'non-standard' microarrays. For the qTR-PCR data, two biological replicates, each with three technical replicates were averaged to represent the data and biological variation can add to the inaccuracy of these results. Additionally differences in efficiency of cDNA synthesis can increase the error. This is because the primers in this study were designed to amplify the 5'end of the tested genes and suboptimal reverse transcription reactions can result in incomplete reverse transcription of longer transcripts. As a consequence, variation in the amount of template available to the primers in the different samples may increase the error in the data. Despite the large error bars the similarities between the qRT-PCR and microarray expression values are agreeable.

4.4. Conclusion

Although care was taken at all stages of the microarray experiment to minimise the introduction of noise, independent verification of the microarray expression values is still necessary. This is because the expression values in this data set, obtained from "boutique" arrays, did not follow a

normal distribution and the control genes (Lucidea spikes) originally included to assist with normalisation were not suitable due to poor hybridisation of these DNA targets. Also, the lack of "housekeeping" or reference genes on the slide as a result of the use of a focused desiccation library restricted the choice of normalisation strategies available. The expression of a small number of genes was verified by comparison with northern blots for the DsL and HL samples (chapter 2) but data was not available for DsR, HR and seed and gene expression levels in these samples still require validation. Additionally, false signal from non-specific and cross hybridisation could result in the identification of genes which are in fact not differentially expressed. The normalised microarray gene expression values must be verified to ensure the validity of the chosen approach and, therefore, real-time analysis was performed on a number of genes from each of the clusters generated in chapter 3. This was to confirm the expression patterns of genes in the microarray experiments and to confirm the assignment of genes to their respective clusters.

Although the microarray and real-time PCR platforms are very different, the resultant expression profiles for both methods were very similar for all the genes tested. The microarray protocol relies on detection of signal from thousands of probes hybridised to targets immobilised to a glass slide while real-time PCR relies on the detection of signal during the amplification of a single product of interest. As a result the downstream data processing differs between the two approaches. The microarray data is subjected to background subtraction, within- and between-slide normalisation using data from all the genes on the slide while the real-time PCR data is only normalised against one "housekeeping" or reference gene. The agreement in gene expression profiles between these two different platforms shows the reliability of the microarray expression data. The assignment of genes to their respective clusters and the expression pattern of these clusters were also confirmed by the real-time PCR results thus validating both the microarray results and the clustering algorithm implemented.

Therefore, the independent confirmation of the expression data and clustering patterns indicate that these clusters are most likely valid and reflect a significant overlap in seed and desiccated

vegetative tissue in *X. humilis*. This supports the original hypothesis that vegetative desiccation tolerance may be due to an activation of a seed desiccation programme.

Chapter 5
Conclusions

The results presented here support the hypothesis of the acquisition of desiccation tolerance in the vegetative tissues of the resurrection angiosperm, *X. humilis*, through the activation of genes believed to be induced during maturation drying of orthodox seeds. A normalized cDNA library prepared from RNA isolated from desiccating leaf and root from *X. humilis* (Collett et al., 2004) was used to assess the relative abundance of transcripts in DsL, DsR and mature seed. A similar analysis was performed on affymetrix AtGenExpress microarray data from the desiccation sensitive plant, *A. thaliana*, in which the relative abundance of transcripts in OsL, OsR and mature seed was determined.

For *X. humilis* a total of 3501 transcripts were printed on microarray slides and probed with RNA from DsL, HL, DsR, HR and mature seed. Genes that were differentially expressed between DsL and HL, DsR and HR, and between seed and both HL and HR were identified. These 1169 differentially expressed cDNAs represented 639 unique contigs. A second strategy was employed to identify those contigs that were differentially expressed between DsL and HL and between DsR and HR and the corresponding seed expression values were retrieved for these contigs. This approach allowed a more balanced comparison between the *X. humilis* dataset, which did not contain seed-specific genes, and the *A. thaliana* full genome dataset. Cluster analysis of these 421 differentially expressed contigs revealed a significant similarity in gene expression profiles between desiccated leaf and root and seed in *X. humilis*. The gene expression profiles in *A. thaliana*, however, were more similar between tissue types (e.g. osmotically stressed leaf and control leaf) than between stressed vegetative tissue and seed. A large group of genes expressed in desiccated vegetative tissue and seed was identified in *X. humilis*. These results were confirmed by real-time PCR and the cluster assignment and gene expression profiles were verified supporting the hypothesis of the activation of those genes that accumulate in mature, dry seed in the vegetative tissue of *X. humilis* in response to desiccation.

The corresponding gene expression profiles of 373 of the 639 differentially expressed contigs originally identified were investigated in *A. thaliana*. The expression of a considerable percentage of these homologues was restricted to mature seed in *A. thaliana* and a significant number of these were expressed in desiccated vegetative tissue of *X. humilis*. These included genes encoding antioxidants, seed storage proteins, HSPs, LEA family members, desiccation-related proteins and RNA-binding proteins in addition to a number of other genes whose products are known to be involved in signaling and sugar metabolism.

Three antioxidants, which have been shown to only be expressed in the mature seed of *A. thaliana*, were expressed in desiccated vegetative tissue of *X. humilis*. These included a 1-cys peroxiredoxin, previously found to be expressed in desiccated vegetative tissue of two *Xerophyta* species (Mowla et al., 2002; Collett et al., 2004), in addition to two oxidoreductases. It has been proposed that housekeeping antioxidants are upregulated during desiccation and that these enzymes are protected at lower RWC in desiccation tolerant species (Illing et al., 2005; Farrant, 2007). The housekeeping antioxidants could function together with homologues or orthologues of those antioxidants shown to be expressed in mature, dry seed. These "specialized" antioxidants may be able to function at low levels of hydration to afford increased tolerance to desiccation.

The transcripts of genes encoding oleosins accumulate during maturation drying in seeds and are also expressed in response to desiccation in the vegetative tissues of *X. humilis.* During desiccation orthodox seeds and the vegetative tissues of *X. humilis* replace their large cytoplasmic vacuole with a number of smaller ones, and the water within these smaller vacuoles is replaced with non-aqueous substances (Farrant, 2000; Mundree and Farrant, 2000). Oleosins stabilize and protect lipid bodies from premature degradation during desiccation (Kai and Huang, 2005) and these proteins could thus function to stabilize lipids within vacuoles of the resurrection plants during desiccation.

Genes encoding two LEA family members, including *Lea3* and *Lea6* were identified. While a number of genes encoding LEAs have been found to be upregulated in response to desiccation in vegetative tissues of resurrection plants the majority are seed-specific in *A. thaliana* and only a few are induced in the stressed vegetative tissue of this plant (Illing et al., 2005). The *Lea6* gene identified in this study has been shown to be expressed only in the seed of *A. thaliana* and has previously been shown to be expressed in dry vegetative tissue of *X. humilis* (Illing et al., 2005). However, this is the first report of the accumulation of *Lea3* transcripts in response to desiccation in the vegetative tissues of *X. humilis*. The transcripts for this gene are known to accumulate in dry, mature *A. thaliana* seeds and the expression of these genes in desiccated vegetative tissue and mature dry seeds indicates the importance of LEA proteins during extreme desiccation.Further support for the involvement of LEAs in seed and vegetative desiccation tolerance is derived from the study of the re-induction of desiccation tolerance in *M. truncatula* seedlings. It was found that 18 out of a total of 187 genes expressed both during seed maturation drying stage and re-establishment of desiccation tolerance in seedlings, encoded LEA proteins (Buitink et al., 2006).

The importance of LEA proteins in desiccation tolerance was once again highlighted when we considered the enrichment of genes overlapping between seed and desiccated vegetative tissue in *X. humilis*. A significant representation of genes encoding both LEA proteins and HSPs was observed in these tissues. This phenomenon appears to be unique to *X. humilis*, as *A. thaliana* only showed a significant representation of genes related to protein degradation in stressed vegetative tissue and seed. HSPs are believed to have a role in the acquisition of desiccation tolerance as they increase in level during seed maturation and persist in the dry state (Vierling, 1991; Coca et al., 1994; Wehmeyer et al., 1996; Kermode and Finch-Savage, 2002). They are constitutively expressed in *C. plantagineum* and increase in response to water stress and application of ABA induces the expression of sHSPs concomitant with the re-induction of desiccation tolerance in sensitive callus (Alamillo et al., 1995).

The persistence of these transcripts to such low water contents in seeds and the vegetative tissues of *X. humilis* would suggest stabilization of the mRNAs during desiccation and tight control of translation upon rehydration. The 3'UTR of transcripts represent an area extremely rich in diverse translational control mechanisms and stability and while they are highly diverse sequences they often contain regulatory motifs that are common to members of the same family (Mazumder et al., 2003). Many transcripts contain a polyadenylation signal (AU rich element) in their 3' UTR. This region consisting mainly of adenine or uridine nucleotides can either stabilize or destabilize the transcript depending on the protein bound to it (Mitchell and Tollervey, 2001). A number of examples of regulation of translation and stabilization by 3'UTR binding proteins have been described (Shaw and Kamen, 1986; Copeland et al., 2000; Piecyk et al., 2000) and analysis of the 3' UTRs of genes that are expressed in desiccated vegetative tissue and seed may reveal the enrichment of regulatory motifs within this cluster of genes and give insight into stabilization during desiccation and translation upon rehydration.

Although genes isolated specifically from mature *X. humilis* seed were not included in the library used in this study, the representation of genes expressed in seeds and vegetative tissue was significant and the overlap between desiccated vegetative tissue and seed was greater than that between hydrated tissue and seed. The expression patterns of a large number of genes were compared here, but analysis of transcript abundance of a larger number of genes may reveal a more significant overlap in gene expression between seed and desiccated vegetative tissue and potentially identify additional genes involved in extreme desiccation. The expression profiles of transcripts in extremely desiccated vegetative tissue and mature dry seed were investigated and these most likely represent genes necessary for survival of extreme desiccation and those required during early rehydration in plants and germination in seeds. A more comprehensive comparison of the expression of genes at various stages of seed maturation drying and vegetative tissue desiccation could reveal additional pathways and protective measures activated in response to water deficit in seeds and resurrection plants.

Additionally, the use of a data set with reported RWC values would provide a more reliable comparative model as this would ensure that the gene expression profiles analysed represent changes that occur in response to a definite physiological stress from the imposed water deficit stress. In this case the authors reported that 10% of fresh weight had been lost during the dehydration stress (Kilian et al., 2007) but it was not possible to ascertain the degree of cellular stress as a consequence of the dehydration stress imposed. While the same holds true for the hyperosmotic stress treatment used in this study, it has been shown that treatment with mannitol at concentrations greater than 200 mM result in significant changes at the transcriptional level and several stress-responsive genes are shown to be upregulated (Kreps et al., 2002; Kilian et al., 2007). It has to be taken into consideration, however, that a number of these stress-responsive genes have been shown to be stimulus-specific (Kreps et al., 2002) and as a result we may have identified some genes that respond specifically to the osmotic stress and not necessarily to water deficit as a result of dehydration. To overcome these shortcomings it would be better to compare the expression data from *X. humilis* with a data set from a model system that has been dehydrated to RWC levels that are physiologically comparable for that desiccation-sensitive system, i.e. it has to be dehydrated to levels that are stressful but non-lethal to the desiccation-sensitive plant. Unfortunately at the time of this study, a data set such as this was not available to us.

In the future, as genomic sequence becomes available for *X. humilis*, analysis of the cis elements in promoter regions of the genes that are induced in seed and desiccated vegetative tissue may identify common regulatory motifs. This could reveal the presence of ABREs, for example, and begin to unravel the regulatory pathways involved in this conserved desiccation response. In addition the 3' UTR of transcripts are often involved in translational control and stability (Mazumder et al., 2003) and analysis of these regions of genes that are expressed in desiccated vegetative tissue and seed may reveal the enrichment of regulatory motifs and give insight into stabilization of these transcripts during desiccation and translation upon rehydration.

In conclusion, this study has revealed a significant overlap in gene expression between vegetative tissue and seed in response to extreme desiccation in *X. humilis*. The results are in agreement with the observations of Buitink et al. (2006) who showed genes that were

upregulated during the re-establishment of desiccation tolerance in seedlings exposed to PEG, are comparable with those involved in late seed maturation. Additionally the overlap in expression and enrichment of gene involved in protection in *X. humilis* would suggest similarities in the protective measures activated in both seed and vegetative tissue during desiccation. The adaptation of the developmentally regulated seed desiccation program in vegetative tissue has most likely afforded this resurrection angiosperm a competitive advantage for survival in semi-arid niches such as those found in shallow soil rocky outcrops. The prevalence of desiccation tolerance in seeds and the overlap in gene expression between seed and desiccated vegetative tissue in this desiccation tolerant angiosperm gives one the reason to believe that the genetic potential to tolerate desiccation exists in all plants.

References

Aalen, R.B. 1999. Peroxiredoxin antioxidants in seed Physiology. *Seed Sci. Res.* 9: 285.

Alamillo, J., Almoguera, C., Bartels, D. and Jordano, J. 1995. Constitutive expression of small heat shock proteins in vegetative tissues of the resurrection plant *Craterostigma plantagineum. Plant Mol. Biol.* 29:1093

Alpert, P and Oliver, M.J. 2002. "Drying without Dying." In *Desiccation and Survival in Plants – Drying without Drying*, edited by Michael Black and Hugh William Pritchard, pp. 3-43. Wallingford and New York: CABI Publishing.

Al-Shahrour, F., Minguez, P., Tárraga, J., Montaner, D., Alloza, E., Vaquerizas, J.MM., Conde, L., Blaschke, C., Vera, J. and Dopazo, J. 2006. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiment, *Nucl. Acids Res.* 34: W472.

Al-Shahrour, F., Minguez, P., Vaquerizas, J.M., Conde, L. and Dopazo, J. 2005. Babelomics: a suite of web-tools for functional annotation and analysis of group of genes in high-throughput experiments, *Nucl. Acids Res.* 33: W460.

Al-Shahrour, F., Díaz-Uriarte, R. and Dopazo, J. 2004. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* 20: 578.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res* 25:3389.

Amuti, K.S. and Pollard, C.J. 1977. Soluble carbohydrates of dry and developing seeds. *Phytochem.* 16: 529

Apel, K. and Hirt, H. 2004. REACTIVE OXYGEN SPECIES: Metabolism, Oxidative Stress, and Signal Transduction. *Annu. Rev. Plant Biol.* 55:373

Arrigoni, O., de Gara, L., Tommasi, F. and Liso, R. 1992. Changes in the ascorbate system during seed development of *Vicia faba* L. *Plant Phys.* 99:235.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29: 365

Bailly C. 2004. Active oxygen species and antioxidants in seed biology. *Seed Sci Res*. 93:107

Baker, J., Steele, C. and Dure, L. III. 1988. Sequence and characterization of 6 LEA proteins and their genes from cotton. *Planta* 175: 485

Baker, E.H., Bradford, K.J. Bryant, J.A. and Rost, T.L. 1995. A comparison of desiccation-related proteins (dehydrin and QP47) in peas (*Pisum sativum*). *Seed Sci. Res.* 5: 185

Bartels, D., Singh, M. & Salamini, F. 1988. Onset of desiccation tolerance during development of the barley embryo. *Planta* 175:485

Bartels, D and Salamini, F. 2001. Desiccation tolerance in the resurrection plant *Craterostigma plantigineum.* A contribution to the study of drought tolerance at the molecular level. *Plant Phys.* 127: 1346

Bartels, D. 2005. Desiccation tolerance studied in the resurrection plant *Craterostigma plantigineum. Interg. Comp Biol* 45: 696.

Becker, R. A., Chambers, J. M. and Wilks, A. R. 1988. The New S Language. Wadsworth & Brooks/Cole.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Series* B,57: 289.

Benjamini, Y. and Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat*. 29:1165-1188.

Berjak, P., Farrant, J.M. and Pammenter, N.W. 2007. Seed desiccation-tolerance mechanisms. In: *Plant Desiccation Tolerance.* Eds:Jenks, M.A. and Wood, A.J. CAB International Press, Wallingford, UK.

Bewley, J. D. 1979. Physiological aspects of desiccation tolerance. Ann. Rev. *Plant Phys.* 30: 195.

Bewley, J.D. and Black, M. 1994. *Seeds. Phys.ogy of Development and Germination,* 2nd ed. New York: Plenum Press.

Bewley, J. D. & Krochko, J. E. 1982. Desiccation tolerance. In: Lange, O. L., Nobel, P. S., Osmond, C. B., & Ziegler, H. (eds), *Encyclopedia of plant Phys.ogy*. Vol 12B. p235 Phys.ogical Ecology II. Springer-Verlag, Berlin.

Bettey, M., and W.E. Finch-Savage. 1998. Stress protein content of mature Brassica seeds and their germination performance. *Seed Sci. Res*. 8:347.

Bianchi, G., Gamba, A., Murelli, C., Salamini, F. and Bartels, D. 1991. Novel carbohydrate metabolism in the resurrection plant *Craterostigma plantigineum. Plant J.* 1: 355.

Blackman, S.A., Obendorf, R.L. and Leopold, A.C. 1992. Desiccation tolerance in developing soybean seeds: The role of stress proteins. *Phys. Planta.* 93: 630.

Blomstedt,C.K., Gianello, R.D., Gaff, D.F., Hamill, J.D. and Neale, A.D. 1998. Differential gene expression in desiccation tolerant and desiccation-sensitive tissue of the resurrection grass *Sporobolus stapfianus. Aus. J. Plant Phys.* 25: 937.

Bolstad, B. M., Irizarry R. A., Astrand, M, and Speed, T. P. 2003. A Comparison of Normalization Methods for High Density OligoNucl.leotide Array Data Based on Bias and Variance. *Bioinformatics* 19: 185.

Berjak, P. 2006. Unifying perspectives of some mechanism basic to desiccation tolerance across life forms. *Seed Sci. Res* 16: 1.

Bohnert, H.J. and Jenson R.G. 1996. Strategies for engineering water-stress tolerance in plants. *Trends Biotech.* 14: 89.

Bohren, K.M, Bullock, B., Wermuth, B. and Gubbay, K.H. 1989. The aldo-keto reductase superfamily. *J. Biol. Chem.* 264: 9547.

Boudet, J., Buitink, J., Hoekstra, F.A., Rogniaux, H., Larré , C., Satour, P. and Leprince, O. 2006. Comparative Analysis of the Heat Stable Proteome of Radicles of Medicago truncatula Seeds during Germination Identifies Late Embryogenesis Abundant Proteins Associated with Desiccation Tolerance. *Plant Phys.* 140: 1418.

Bradford, K.J. and Chandler, P.M. 1992. Expression of 'dehydrin-like' proteins in embryos of *Zizania palustris* and *Oryza sativa* during dehydration. *Plant Phys.* 99:488.

Bray, E.A. 1993. Molecular Responses to Water Deficit. *Plant Phys.* 103: 1035.

Bray, E.A. 1997. Plant responses to water deficit. *Trends Plant Sci.* 2: 48.

Bray, E.A. 2002. Classification of genes differentially expressed during water deficit stress in *Arabidopsis thaliana*: an analysis using microarray and differential expression data. *Annuals Bot. (London)* 89: 803.

Browne, J. A., Dolan, K. M., Tyson, T., Goyal, K., Tunnacliffe, A., and Burnell, A.M. 2004. Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *Aphelenchus avenae*. *Euk. Cell* 3:966.

Buitink, J., Walters-Vertucci, C., Hoekstra, F.A. and Leprince, O. 1996. Calorimetric properties of dehydrating pollen: analysis of a desiccation-tolerant and intolerant species. *Plant Phys.* 11: 235.

Buitink, J., Hemmings, M.A. and Hoekstra, F.A. 2000a. Is there a role for oligosaccharides in seed longevity? An assessment of intracellular glass stability. *Plant Phys.* 122:1217.

Buitink, J., Hemmings, M.A. and Hoekstra, F.A. 2000b. The effects of moisture and temperature on the ageing kinetics of pollen: interpretation based on cytoplasmic mobility. *Plant Cell Environ.* 23: 967.

Buitink, J., Leprince, O., Hemmings, M.A. and Hoekstra, F.A. 2000c. Molecular mobility in the cytoplasm: an approach to describe and predict lifespan of dry germplasm. *PNAS* 97: 2385.

Buitink, J., Leprince, O. and Hoekstra, F.A. 2000d. Dehydration-induced redistribution of amphiphilic molecules between cytoplasm and lipids is associated with desiccation tolerance in seeds. *Plant Phys.* 124: 1413.

Buitink, J. Hoekstra, F.A. and Leprince, O. 2002. "Biochemistry and biophysics of tolerance systems." In *Desiccation and Survival in Plants – Drying without Drying*, (eds) Michael Black and Hugh William Pritchard, p. 293. Wallingford and New York: CABI Publishing.

Buitink, J., Leger, J.J., Guisle, I., Vu1, B.L., Wuilléme, S., Lamirault, G., Le Bars, A., Le Meur, N., Becker, A., Küster, H. and Leprince, O. 2006. Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation sensitive to desiccation-tolerant stages in Medicago truncatula seeds. *Plant J.* 47: 735

Burke, M. J. 1986. The glassy state and survival of anhydrous biological systems. In: Membranes, metabolism and dry organisms. (eds) Leopold, A. C. Cornell University Press, Ithaca, New York, p. 358.

Bustin, S. 2000. Absolute quantification of mRNA using REAL-TIME reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 25:169.

Cadman, C.S.C., Toorop, P.E., Hilhorst, H.W.M. and Finch-Savage, W.E. 2006. Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicates a common underlying dormancy control mechanism. *Plant J.* 46: 805.

Carrera, E., Holman, T., Medhurst, A., Dietrich, D., Footitt, S., Theodoulou, F.L. and Holdsworth, M.J. 2008. Seed after-ripening is a discrete developmental pathway associated with specific gene networks in Arabidopsis. *Plant J.* 53:214.

Carles, C., Bies-Etheve, N., Aspert, L., Le´ on-Kloosterziel, K.M., Koornneef, M., Echeverria, M. and Delseny, M. 2002. Regulation of Arabidopsis thaliana Em genes: role of ABI5. *Plant J.* 30: 373.

Carpenter, J.F., Crowe, J.H. and Arakawa, T. 1990. Comparison of solute-induced protein stabilization in aqueous solution and in frozen and dried states. *J. Dairy Sci.* 73: 3627.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. 1983. Graphical Methods for Data Analysis._ Wadsworth & Brooks/Cole.

Chang, B.S., Beauvais, R.M., Dong, A. and Carpenter, J.F. 1996. Physical factors affecting the storage stability of freeze-dried interleukin-1 receptor antagonist: glass transition and protein conformation. *Arch. Biochem. Biophys.* 331: 249.

Chazotte, B. 1994. Comparisons of the relative effects of polyhydroxyl compounds on local versus long-range motions in the mitochondrial inner membrane. Fluorescence recovery after photobleaching, fluorescence lifetime, and fluorescence anisotropy studies. *Biochim Biophys Acta* 1194:315.

Chuaqui, R.F., Bonner, R.F., Best, C.J.M., Gillespie, J.W., Flaig, M.J., Hewitt, S.M., Phillips, J.L., Krizman, D.B., Tangrea, M.A., Ahram, M., Linehan, M.W., Knezevic, V. and Emmert-Buck, M.R. 2002. Post-analysis follow-up and validation of microarray experiments. *Nat. Gen. suppl.* 32:509

Clegg, J.S. 1986. The physical properties and metabolic status of Artemia cysts at low water contents: the 'water replacement hypothesis'. In: Leopold, A.C. (ed.) *Membranes, Metabolism, and Dry Organisms*. Cornell University Press, Ithaca, New York, p. 169.

Close, T.J. 1996. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Phys. Planta.* 97: 795.

Close, T.J. 1997. Dehydrins: a commonality in the response of plants to dehydration and low temperature. *Phys. Planta.* 100: 291.

Coca, M.A., Almoguera, C. and Jordano, J. 1994. Expression of sunflower low-molecular weight heat-shock proteins during embryogenesis and persistence after germination: localization and possible functional implications. *Plant Mol. Biol.* 25, 479.

Collett, H., Butowt, R., Smith, J., Farrant, J. and Illing, N. 2003. Photosynthetic genes are differentially transcribed during the dehydration rehydration cycle in the resurrection plant, *Xerophyta humilis*. *J. Exp. Bot*. 54: 2593.

Collett, H., Shen, A., Gardner, M., Farrant, J.M., Denby, K.J. and Illing, N. 2004. Towards profiling of desiccation tolerance in *Xerophyta humilis* (Bak.) Dur and Schinz: Construction of a normalized 11 k *X. humilis* cDNA set and microarray expression analysis of 424 cDNAs in response to dehydration *Phys. Planta.* 122: 39.

Copeland, P.R., Fletcher, J.E., Carlson, B.A., Hatfield, D.L. and Driscoll, D.M. 2000. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J*. 19: 306.

Crowe, L.M., Womersley, C., Crowe, J.H. Reid, D., Appel, L., and Rudolph, A. 1986. Prevention of fusion and leakage in freeze-dried liposomes by carbohydrates. *Biochimica et Biophysica Acta* 861: 131.

Crowe, J.H., Crowe, L.M. Carpenter, J.F and Wistrom, C.A. 1987. Stabilisation of dry phospholipid bilayers and proteins by sugars. *Biochem. J* 242: 1.

Crowe, J. H., Hoekstra, F. A. & Crowe, L. M. 1992. Anhydrobiosis. *Ann. Rev. Phys.* 54: 579.

Crowe, J.H, Carptenter, J.F. and Crowe, L.M. 1998. The role of vitrification in anhydrobiosis. *Annual Rev. of Phys.* 60: 73.

Cui, X., Hwang, J.T., Qiu, J., Blades, N.J. and Churchill, G.A. 2005. Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* 6:59.

Cuming, A. C. 1999. LEA proteins. *In* P. R. Shewry and R. Casey (eds.), *Seed proteins*, pp. 753. Kluwer Academic Publishers, Netherlands.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. and Scheible, W-R. 2005. Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. *Plant Phys.* 139: 5.

Dalla Vecchia, F., El Asmar, T., Calamassi, R., Rascio, N., and Vazzana, C. 1998. Morphological and ultrastructural aspects of dehydration and rehydration in leaves of *Sporobolus stapfianus. Plant Growth Reg.* 24: 219.

Dace, H., Sherwin, H.W., Illing, N. and Farrant, J.M. 1998. Use of metabolic inhibitors to elucidate mechanisms of recovery from desiccation stress in the resurrection plant *Xerophyta humilis*. *Plant Growth Reg*. 24:171.

Dean, R.T., Gieseg, S. and Davies, M.J. 1993. Reactive species and their accumulation on radical-damaged proteins. *Trends of Biochem. Sci.* 18: 437.

Desire, N., Dehee, A. and Schneider, V. 2001. Quantification of human immunodeficiency virus type 1 proviral load by a TaqMan real-time PCR assay. *J. Clin. Microbiol.* 39:1303

Dhindsa, R.S. 1987. Protein synthesis during rehydration of rapidly dried *Tortula ruralis. Plant Phys.* 85: 1094.

Dickie, J. B and H. W. Pritchard. 2002. Systematic and evolutionary aspects of desiccation tolerance in seeds. *In* M. Black and H.W. Pritchard (eds.), *Desiccation and survival in plants: Drying without dying,* pp. 239. CABI publishing, Wallingford and New York.

Dickinson, C.D., Evans, R.P. and Nielsen, N.C. 1988. RY repeats are conserved in the 5¢-flanking regions of legume seed-protein genes. *Nucl. Acids Res.* 16: 371

Dong, J., Keller, W.A., Yan, W. and Georges, F. 2004. Gene expression at early stages of Brassica napus seed development as revealed by transcripts profiling of seed-abundant cDNAs. *Planta* 218: 483.

Drãghici, S. 2003. *Data Analysis tools for microarrays*. Chapman & Hall / CRC Press, London, UK.

Dubouzet. JG., Sakuma Y., Ito. Y., Kasuga, M., Dubouzet, E.G., Miura, S., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2003. OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33: 751

Duff, R.J., Oliver, M.J. and Wood, A.J. 1999. A *Tortula ruralis* cDNA encoding small-subunit ribosomal protein S3a: polysomal retention of transcript in response to desiccation and rehydration. *Bryologist.* 102: 418.

Dure, L. III. 1993. A repeating 11-mer amino acid motif and plant desiccation. *Plant J.* 3: 363.

Eisen, M.B. et al. (1998) Cluster analysis and display of genome-wide expression patterns. *PNAS.* 95: 14863

Elstner, E.F. and Oswald, W. 1994. Mechanisms of oxygen activation during plant stress. *Proc of the Royal Society Edinburgh* 102: 131.

Esmann, M., Hideg, K. and Marsh, D. 1994. Influence of poly(ethylene glycol) and aqueous viscosity on the rotational diffusion of membranous Na,K-ATPase. *Biochem.* 33:3693.

Ewing, B., Hillier, L., Wendl, M. and Green. P. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Gen Res* 8:175.

Ezcurra, I., Ellerström, M., Wycliffe, P., Stäberg, K. and Rask, L.1999. Interaction between composite elements in the napA promoter: both the B-Box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol. Biol.* 40:699.

Farrant, J.M., Pammenter, N.W. and Berjak, P. 1992. A contribution to an understanding of desiccation tolerance from a study of a desiccation-sensitive seed species. *Fourth International Workshop on seeds*, Angers, France.

Farrant, J.M., Berjak, P., Walters, C. and Pammenter, N.W. 1997. Subcellular organisation and metabolic activity in seeds which develop different degrees of tolerance to water loss. *Seed Sci. Res.* 7: 135.

Farrant, J.M. and Walters, C. 1998. Ultrastructural and biophysical changes in developing embryos of *Aesculus hippocastanum* in relation to the acquisition of tolerance to drying. *Phys. Planta.* 104: 513.

Farrant, J.M., Cooper, K., Kruger, L.A. and Sherwin, H.W. 1999. The effect of rate of drying on three different resurrection plants. *Ann. Bot.* 84: 371.

Farrant, J.M. 2000. Comparison of mechanisms of desiccation tolerance among three angiosperm resurrection plants. *Plant Ecol.* 151: 29.

Farrant, J.M. and Kruger, L.A. 2001. Effects of long-term drying on the resurrection plant *Myrothamnus flabellifolius*. *Plant Growth Reg.* 35: 109.

Farrant, J.M., Bartsch, S., Loffell, D., Vander Willigen, C. and Whittaker, A. 2003. An investigation into the effects of light on the desiccation of three resurrection plants species. *Plant Cell Environ.* 26: 1275.

Farrant, J.M. 2007. Mechanisms of desiccation tolerance in Angiosperm resurrection plants. In: *Plant Desiccation Tolerance.* Eds: Jenks, M.A. and Wood, A.J. CAB International Press, Wallingford, UK.

Fato, R., Cavazzoni, M., Castlluccio, C., Parenti Castelli, G., Palmer, G., Degli Esposti, M. and Lenaz, G. 1993 Steady-state kinetics of ubiquinol-cytochrome c reductase in bovine heart submitochondrial particles: diffusional effects. *Biochem. J.* 290:225

Ferguson, D., Guikema, J. and Paulsen, G. 1990. Ubiquitin pool modulation and protein degradation in wheat roots during high temperature stress. *Plant Phys.* 92: 740.

Finch-Savage, W.E., Pramanik, S.K. and Bewley, J.D. 1994. The expression of dehydrin proteins in desiccation-sensitive (recalcitrant) seeds of temperate trees. *Planta* 193: 478.

Finkelstein, R.R., Wang, M.L., Lynch, T.J., Rao, S. and Goodman, H. 1998. The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA2 domain protein. *Plant Cell* 10: 1043.

Finkelstein, R.R., Gampala, S.S.L. and Rock, C.D. 2002. Abscisic acid signaling in seeds and seedlings. *Plant Cell* S15.

Finch-Savage, W.E. 1996. The role of developmental studies in research on recalcitrant and intermediate seeds. pp. 83–97 in Ouédraogo, A.-S.; Poulsen, K.; Stubsgaard, F. (Eds) Intermediate/recalcitrant tropical forest tree seeds. Rome, IPGRI.

Finch-Savage, W.E., Cadman, C.S, Toorop, P.E., Lynn, J.R. and Hilhorst, H.W.M. 2007. Seed dormancy release in Arabidopsis Cvi by dry afterripening, low temperature, nitrate and light shows common quantitative patterns of gene expression directed by environmentally specific sensing. *Plant J.*51:60.

Fisher, R. A. 1950. *Statistical Methods for Research Workers*. Oliver and Boyd.

Fraley, C. and Raftery, A. E. 2002a. Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Stat. Assoc*. 97:611.

Fraley, C. and Raftery, A. E. 2002b. MCLUST: Software for model-based clustering, density estimation and discriminant analysis. Technical Report, Department of Statistics, University of Washington. http://www.stat.washington.edu/mclust.

Foyer, C.H., Lelandias, M and Kunert, K.J. 1994. Photooxidative stress in plants. *Phys. Planta.* 92: 696.

Foyer, C.H., Perry, M. and Noctor, G. 2003. Markers and signals associated with nitrogen metabolism in higher plants. *J. Exp. Bot.* 54:585

Franks, F., Hatley, R.H.M. and Mathias, S.F. 1991. Materials Sci. and production of shelf stable biologicals. *Bio-pharm* 4, 38.

Gaff, D.F. 1977. Desiccation tolerant vascular plants of Southern Africa. *Oecologia* 31:95.

Gaff, D. and Bole, P. 1986. Resurrection grasses in India. *Oecologia* 71: 159.

Gaff, D. F. 1987. Desiccation tolerant plants in South America. Oecologia 74:133.

Gaff, D. F. 1989. Responses of desiccation tolerant ‘‘resurrection’’ plants to water stress. *In* K. H. Kreeb., H. Richter, and T. M. Hinckley (eds.), *Structural and functional responses to environmental stress: Water shortage*, pp. 255. SPB Academic Publishers, The Hague.

Galau, G.A., Wang, H.Y.C. and Hughes, D.W. 1993. Cotton Lea5 and Lea14 encode atypical late embryogenesis-abundant proteins. *Plant Phys.* 101: 695.

Gallardo, K., Job, C., Groot, S. P.C., Puype, M., Demol, H., Vandekerckhove, J., Job, D. 2002. Proteomics of Arabidopsis seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Phys.* 129: 823

Garay-Arroyo, A., Colmenero-Flores, J.M., Garciarrubio, A., and Covarrubias, A.A. 2000. Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem*. 275:5668

Ghasempour, H.R., Gaff, D.F., Williams, R.D. and Gianello, R.D. 1998. Contents of sugars in leaves of drying desiccation tolerant flowering plants, particularly grasses. *Plant Growth Reg.* 24: 185.

Ginzinger, D.G., Godfrey, T.E. and Nigro, J. 2000. Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. *Cancer Res.* 60:5405

Ginzinger, D.G. 2002. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Exp. Hemat.* 30:503.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. and Goodman, H.M. 1992. Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4: 1251.

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

Golovina, E., Hoekstra, F.A. and Van Aelst, A.C. 2001. The competence to acquire cellular desiccation-tolerance is not dependent on seed morphological development. *J. Exp. Bot*. 52: 1015.

Goyal, K. et al. 2003. Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *J. Biol.Chem.* 278:12977

Goyal, K., Walton, L.J. and Tunnacliffe, A. 2005. LEA proteins prevent protein aggregation due to water stress. *Biochem J.* 388: 151.

Halliwell, B. 1987. Oxidative damage, lipid peroxidation and antioxidant protection in chloroplasts. *Chem. Physics. Lipids* 44: 3227.

Halliwell, B. and Gutteridge, J.M.C. 1999. *Free radicals in biology and medicine, 3rd ed.* Oxford: Oxford University Press.

Han, B., Berjak, P., Pammenter, N., Farrant, J. and Kermode, A.R. 1997. The recalcitrant plant species, *Castanospermum australe* and *Trichilia dregeana*, differ in their ability to produce dehydrin-related polypeptides during seed maturation and in response to ABA or water-deficit-related stresses. *J. Exp. Bot.* 48:1717

Hand, S.C. and Hardewig, I. 1996. Down-regulation of cellular metabolism during environmental stress: mechanisms and implications. *Annu Rev Physiol* 58:539.

Hardie, D.G., Carling, D. and Carlson, M. 1998. The AMP-activated/ SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cells? Annu Rev Biochem 67:821.

Haslekas, C., Grini, P.E., Nordgard, S.H., Thorstensen, T., Viken, M.K., Nygaard, V., Aalen, R.B. 2003, ABI3 mediates expression of the peroxiredoxin antioxidant AtPER1 gene and induction by oxidative stress. *Plant Mol. Biol.* 53**:** 313

Hasselgren, P.O. and Fischer, 1997. The ubiquitin-proteasome pathway: review of a novel intracellular mechanism of muscle protein breakdown during sepsis and other catabolic conditions. *Ann. Surg*. 225: 307.

Haslekas, C., R. A. Stacy., V. Nygaard., F. A. Culianez-Marcia, and R. B. Aalen. 1998. The expression of a peroxiredoxin antioxidant gene, AtPerl, in *Arabidopsis thaliana* is seed-specific and related to dormancy. *Plant Mol. Biol*. 36:833.

Hattori, T., Vasil, V., Rosenkrans, L., Hannah, L.C., McCarty, D.R. and Vasil, I.K. (1992) The Viviparous-1 gene and abscisic acid activate the C1 regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Dev*. 6: 609.
Huang, B. 2006. Plant-environment interactions. CRC Press.

Hendry, G. A. F., W. E. Finch-Savage., P. C. Thorpe., N. M. Atherton., S. M. Buckland., K. A. Nilsson, and W. E. Seel. 1992. Freeradical processes and loss of seed viability during desiccationin the recalcitrant species *Quercus robur*. *New Phytol*. 122:273

Hendry, G.A.F. 1993. Oxygen and Free radical processes in seed longevity. *Seed Sci. Res.* 3: 141.

Hennig, L., Gruissem, W., Grossniklaus, U. and Köhler, C. 2004. Transcriptional programs of early reproductive stages in Arabidopsis. *Plant Phys.* 135: 1765.
Hoekstra, F.A., Golovian, E.A. and Buitink, J. 2001. Mechanisms of plant desiccation tolerance. *Trends Plant Sci.* 6: 431.

Hilhorst, H.W.M. (1995) A critical update on seed dormancy. 1. Primary dormancy. *Seed Sci. Res*. 5: 61.
Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75: 800-803.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Sca. J. Stat*. 6:65-70.

Hommel, G. 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 75:383-386.

Illing, N., Denby, K., Collett, H., Shen, A. and Farrant, J.M. 2005. The signature of seeds in resurrection plants: a molecular and Phys.ogical comparison of desiccation tolerance in seeds and vegetative tissues. *Interg. Comp. Biol.* 45: 771.

Imai, R., L. Chang., A. Ohta., E. A. Bray, and M. Takagi. 1996. A lea-class gene of tomato confers salt and freezing tolerance when expressed in *Saccharomyces cerevisae*. *Gene* 170:243.

Ingham, D., Beer, S., Money, S. and Hansen, G. 2001. Quantitative real-time PCR assay for determining transgene copy number in transformed plants. *Biotechniques* 31:132

Ingram, J. and Bartels, D. 1996. The molecular basis of dehydration tolerance in plants. *Ann. Rev. Plant Phys. Plant Mol. Biol.* 47: 377.

Ingram, J., Chandler, J.W., Gallagher, L., Salamini, F. and Bartels, D. 1997. Analysis of cDNA clones encoding sucrose-phosphate synthase in relation to sugar interconversions associated with dehydration in the resurrection plant Craterostigma plantagineum Hochst. *Plant Phys.* 115: 113.

Kahkonen, M.P., Hopia, A.I., Vuorela, H.J., Ruaha, J.P., Pihlaja, K.K. and Heinonen, T.S. 1999. Antioxidant activity of plant extracts containing phenolic compounds. *J. Agric. Food Chem.* 47: 3562.

Kappen, L. and F. Valladares. 1999. Opportunistic growth and desiccation tolerance: The ecological success of poikilohydrous autotrophs. *In* F. I. Pugnaire and F. Valladares (eds.), *Handbook of funct. plant ecology,* pp. 10. Marcel Dekker, New York.

Kai, H. and Anthony H.C. Huang, A.H.C. 2005. Lipid-rich tapetosomes in *Brassica* tapetum are composed of oleosin-coated oil droplets and vesicles, both assembled in and then detached from the endoplasmic reticulum. *Plant J.* 43: 889.

Kaufman, L. and Rousseeuw, P.J. 1990. "Finding Groups in Data" Chichester: Wiley.

Kaushik, J.K. and Bhat, R. 2003. Why is trehalose an exceptional protein stabilizer? An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose. *J. Biol. Chem.* 278: 26458.

Kazutoshi Yamagishi, K., Nagata, N., Yee, K.M., Braybrook, S.A., Pelletier, J., Fujioka, S., Yoshida, S., Fischer, R.L., Goldberg, R.B. and John J. Harada, J.J. 2005. *TANMEI/EMB2757* Encodes a WD Repeat Protein Required for Embryo Development in Arabidopsis. *Plant Phys.* 139:163.

Kermode, A.R. 1990. Regulatory mechanisms in the transition from seed development to germination. *Critical Rev. Plant Sci* 9: 155.

Kermode, A.R. 1995. Regulatory mechanisms in the transition from seed development to germination: interactions between the embryo and the seed environment. In: Galili, G. and Kigel, J. (eds.) *Seed Development and Germination*. Marcel Dekker, New York, pp. 273.

Kermode, A.R. and Finch-Savage, W.E. 2002. "Desiccation sensitivity in orthodox and recalcitrant seeds in relation to development." In *Desiccation and Survival in Plants: Drying without Dying*. edited by M. Black and H.W. Pritchard, pp. 149. Wallingford UK: CABI publishing.

Kermode, A.R. 1997. Approaches to elucidate the basis of desiccation-tolerance in seeds. *Seed Sci. Res.* 7: 75.

Kerr, M.K. and Churchill, G.A. 2001. Experimental design for gene expression microarrays. *Biostatistics* 2:183.

Khojasteh, M., Lam, W.L., Ward, R.K. and MacAulay, C. 2005. A stepwise framework for the normalization of array CGH data. *BMC Bioinf.* 6:274

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J. and Harter, K. 2007. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* 50: 347.

Koster, K.L. and Leopold, A.C. 1988. Sugars and desiccation tolerance in seeds. *Plant Phys.* 88: 829

Koster, K.L. and Bryant, G. 2005. "Dehydration in model membranes and protoplasts: contrasting effects at low, intermediate and high hydrations." In *Cold Hardiness in Plants:*

Molecular Genetics, Cell Biology and Phys., edited by T.H.H. Chen, M. Uemura and S. Fujikawa, pp. 219. Wallingford, UK: CAB International.

Kovach, D.A. and Bradford, K.J. 1992. Imbibitional damage and desiccation tolerance of wild rice *(Zizania palustris)* seeds. *J Exp Bot* 43:747.

Kranner, I**.,** Beckett, R.P., Wornik, S, Zorn, M and Pfeifhofer, H.W**.** (2002). Revival of a resurrection plant correlates with its antioxidant status. *Plant J.* 31: 13

Kranner, I. and Grill, D. 1996. Significance of thiol disulphide exchange in resting stages of plant development. *Botanica Acta* 109: 8.

Kranner, I. and D. Grill. 1997. Desiccation and the subsequent recovery of cryptogamics that are resistant to drought. *Phyton* 37:139.

Kranner, I**.,** Beckett, R.P., Wornik, S, Zorn, M and Pfeifhofer, H.W**.** 2002. Revival of a resurrection plant correlates with its antioxidant status. *The Plant J.* 31: 13

Kranner, I. and Birtić, S. 2005. A modulating role for antioxidants in desiccation tolerance. *Interg. Comp. Biol.* 45: 734.

Kranner, I., Birtić, S., Anderson, K.M. and Pritchard, H.W. 2006. Glutathione half-cell reduction potential: a universal stress marker and modulator of programmed cell death? *Free Radical Biol. Medicine* 40: 2155.

Kreps JA, Wu Y, Chang H-S, Zhu T, Harper JF (2002) Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Phys.* 130: 2129

Lancien, M., Gadal, P. and Hodges, M. 2000. Enzyme redundancy and the importance of 2 oxoglutarate in higher plant ammonium assimilation. *Plant Phys.* 123: 817.

Larson, D.W. 1988. The impact of ten years at -20°C on gas exchange in five lichen species. *Oecologia* 78, 97.

Lehner, A., Chopera, D.R., Peters, S.W., Keller, F., Mundree, S.G., Thomson, J.A. and Farrant, J.M. 2008. Protection mechanisms in the resurrection plant *Xerophyta viscosa*: cloning, expression, characterisation and role of *XvINO1*, a gene coding for a *myo*-inositol 1-phosphate synthase. *Fun. Plant Biol.* 35: 26

Leopold, A.C, Sun, W.Q and Bernal-Lugo, I. 1994. The glassy state in seeds: analysis and function. *Seed Sci. Res* 4: 267.

Leprince, O., Hendry, G. A. F. and McKersie, B. D. 1993. The mechanisms of desiccation tolerance in developing seeds. *Seed Sci. Res.* 3: 231.

Leprince, O., Atherton, N.M., Deltour, R. and Hendry, G.A.F .1994. The involvement of respiration in free radical processes during loss of desiccation tolerance in germinating *Zea mays* L.: an electron paramagnetic resonance study. *Plant Physiol.* 104:1333.

Leprince, O. and Walters-Vertucci, C. 1995. A calorimetric study of the glass transition behaviours in axes of bean with relevance to storage stability. *Plant Phys.* 109: 1471.

Leprince, O. and Hoekstra, F.A. 1998. The responses of cytochrome redox state and energy metabolism to dehydration support a role for cytoplasmic viscosity in desiccation tolerance. *Plant Phys.* 118: 1253.

Leprince, O., Buitink, J. and Hoekstra, F.A. 1999. Axes and cotyledons of recalcitrant seeds of Castanea sativa Mill. exhibit contrasting responses of respiration to drying in relation to desiccation sensitivity. *J. Exp. Bot.* 50: 1515.

Leprince, O., Harren, F.J.M., Buitink, J., Alberda, M. and Hoekstra, F.A. 2000. Metabolic dysfunction and unabated respiration precede the loss of membrane integrity during dehydration of germinating radicles. *Plant Phys.* 122: 597.

Lerner, H.R. 1999. Plant responses to environmental stresses. From phytohormones to genome reorganization. Edited by H.R. Lerner. Marcel Dekker Inc. New York.

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

Lunn, J.E. and Furbank, R.T. 1999. Transley Review No. 105. Sucrose biosynthesis in C4 plants. *New Phytol*.143: 221.

Marcotte, Jr, W.R., Russell, S.H. and Quatrano, R.S. 1989. Abscisic acid-responsive sequences from the Em gene of wheat. *Plant Cell* 1: 969.

Mateos, A., Herrero, J., Tamames, J. and Dopazo, J. 2002. Supervised neural networks for clustering conditions in DNA array data after reducing noise by clustering gene expression profiles. In Lin, S. and Johnson, K. (eds), *Methods of Microarray Data Analysis II*. Kluwer Academic Publishers, Boston. pp. 91.

Mazumder, B., Seshadri, V. and Fox, P.L. 2003. Translational control by the 3' UTR: the ends specify the means. *Trends Bioch. Sci.* 28:91.

McCubbin, W.D., Kay, C.M. and Lane, B.G. 1985. Hydrodynamic and optical properties of the wheat germ Em protein. *Canadian J. of Biochem. Cell Biol.* 63: 803.

McKersie, B.D., Senaratna, T., Walker, M.A., Kendall, E.J. and Hetherington, R. 1988. Deterioration of membranes during aging in plants: evidence for free radical mediation. In:

Nooden, L.D. and Leopold, A.C. (eds) *Senescence and Aging in Plants*. Academic Press, New York, pp. 412.

McKersie, B.D., Crowe, J.H. and Crowe, L.M. 1989. Free fatty acid effects on leakage, phase properties and fusion of fully hydrated model membranes. *Biochimica et Biophysica Acta* 982: 156.

McKersie, B.D. 1991. The role of oxygen free radicals in mediating freezing and desiccation stress in plants. In: Pell, E. and Staffen, K. (eds.) *Active Oxygen and Oxidative Stress in Plant Metabolism. Current Topics in Plant Phys.* American Society of Plant Phys. Series, Vol. 8, pp.107.

Mc-Queen Mason, S.J. and Cosgove, D.J. 1995. Expansin mode of action on cell walls: analysis of wall hydrolysis, stress relaxation, and binding. *Plant Phys.* 107: 87.

Møller, I. M., Jensen, P.E. and Hansson, A. 2007. Oxidative Modifications to Cellular Components in Plants. *Annu. Rev. Plant Biol*. 58:459

Mönke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.-P., Bäumlein, H. and Conrad, U. 2004. Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta* 219: 158.

Moore, J., Farrant, J.M., Brandt, W and Lindsey, GG 2005a. The South African and Namibian populations of the resurrection plant *Myrothamnus flabellifolius* are genetically distinct and display variation in their galloylquinic acid composition. *J. Chem. Ecol.* 31: 2823.

Moore, J., Westall, K.L., Ravenscroft, N., Farrant, J.M., Lindsey, G.G and Brandt, W.F. 2005b. The predominant polyphenol in the leaves of the resurrection plant *Myrothanmnus flabellifolius,* 3,4,5 tri-O-galloylquinic acid, protects membranes against desiccation and free radical induced oxidation. *Biochem. J.* 385: 301.

Moore, JP, Cannesan, MA, Chevalier, LM, Lindsey GG, Brandt W, Lerouge P, Farrant, JM & Driouich, A. 2006. The response of the leaf cell wall to desiccation in the resurrection plant *Myrothamnus flabellifolius. Plant Phys.* 141: 651.

Moore, JP, Hearshaw, M, Ravenscroft, N., Lindsey, GG, Farrant, JM and Brandt, WF. 2007a. Desiccation-induced ultrastructural and biochemical changes in the leaves of the resurrection plants *Myrothanmus flabellifolia. Australian J. Bot.* 55: 482

Moore, J., Lindsey, G.G., Farrant, J.M. and Brandt, W.F. 2007b. An overview of the biology of the desiccation-tolerant plant *Myrothamnus flabellifolia. Ann. Bot.* 99: 211

Mowla, S.B., Thomson, J.A., Farrant, J.M. and Mundree, S.G 2002. A novel stress-inducible antioxidant enzyme identified from the resurrection plant *Xerophyta viscosa. Planta* 215: 716.

Mtwisha, L., Brandt, W., McCready, S. and Lindsey, G.G. 1998. HSP 12 is a LEA-like protein in *Saccharomyces cerevisiae. Plant Mol. Biol.* 37: 513.

Mtwisha, L., Farrant, J., Brandt, W. and Lindsey, G. 2006. "Protection Mechanisms against Water Deficit Stress: Desiccation Tolerance in Seeds as a Study Case. In *Drought Adaptation in Cereals,* edited by Jean-Marcel Ribaut, pp531. New York, London, Oxford; The Haworth Press Inc.

Mundree, S.G. and Farrant, J.M. 2000. Some Phys.ogical and molecular insights into the mechanisms of desiccation tolerance in the resurrection plant *Xerophyta viscosa* Baker. IN *Plant tolerance to abiotic stresses in agriculture: role of genetic engineering,* edited by J. Cherry, pp 201. The Netherlands: Kluwer Academic Publishers.

Mundree, S.G., Baker, B., Mowla, S., Peters, S., Marais, S., Vander Willigen, C., Govender, K., Maredza, A., Farrant, J.M. and Thomson, J.A. 2002. Phys.ogical and molecular insights into drought tolerance. *African J. Biotech.*1:28.

Munne-Bosch, S and Alegre, L. 2002. The function of tocopherols and tocotrienols in plants. *Crit. Rev. Plant Sci.* 21:31.

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

Nakashima, K., Kiyosue, T., Yamaguchi-Shinozaki, K. and Shinozaki, K. 1997. A Nucl.lear gene, erd1, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally up-regulated during senescence in Arabidopsis thaliana. *Plant J.* 12: 851.

Nambara, E. and Marion-Poll, A. 2005. Abscisic acid biosynthesis and catabolism. *Ann. Rev. Plant Biol*. 56: 165.

Ndima, T.B., Farrant, J.M., Thomson, J.A. & Mundree, S.G. 2001. Molecular characterization of XVT8, a stress-responsive gene from the resurrection plant Xerophyta viscosa Baker. *Plant Growth Reg*. 35: 137.

Nigro, J.M., Takahashi, M.A., Ginzinger, D.G. 2001. Detection of 1p and 19q loss in oligodendroglioma by quantitative microsatellite analysis, a real-time quantitative PCR assay. *Am. J. Pathol.* 4: 1253

Niklas, K.J. 1997. The Evolutionary Biology of Plants. University of Chicago Press, Chicago, USA

Niu, X., Helentjaris, T. and Bate, N.J. 2002. Maize ABI4 binds coupling element1 in abscisic acid and sugar response genes. *Plant Cell* 14: 2565.

Noctor, G. and Foyer, C.H. 1998. Ascorbate and glutathione: keeping active oxygen under control. *Ann. Rev. Plant Phys. Plant Mol. Biol.* 49: 249.

Nolan, T., Hands, R.E. and Bustin, S.A. 2006. Quantification of mRNA using real-time RT-PCR. *Nature Protocols* 1:1559.

Oliver, D.H., Thompson, R.E., Griffin, C.A. and Eshleman, J.R. 2000. Use of single Nucl.leotide polymorphisms (SNP) and real-time polymerase chain reaction for bone marrow engraftment analysis. *J. Mol. Diagn.* 2:202

Oliver, M. J. 1996. Desiccation tolerance in vegetative plant cells. *Phys. Plant* 97:779.

Oliver, M.J. and Bewley, J.D. 1997. Desiccation tolerance of plant tissues: A mechanistic overview. *Hort. Rev.* 18: 171.

Oliver, M.J., Wood, A.J. and O'Mahony, P. 1998. "To dryness and beyond" – preparation for the dried state and rehydration in vegetative desiccation-tolerant plants. *Plant Growth Reg.* 24: 193.

Oliver, M. J., Z. Tuba, and B. D. Mishler. 2000. The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol.* 151: 85.

Oliver, M.J., Velten, J. And Mishler, B.D. 2005. Desiccation Tolerance in Bryophytes: A Reflection of the Primitive Strategy for Plant Survival in Dehydrating Habitats? *Integr. Comp. Biol.*,45:788

Öpik, H. 1985. The fine structure of some dry seed tissues observed after completely anhydrous chemical fixation. *Ann. Bot.* 56, 453.

Oztur Z.N., Talame V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R. and Bohnert, H.J 2002. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol. Biol.* .48: 551

Pammenter, N.W. and Berjak, P. 1999. A review of recalcitrant seed Phys. in relation to desiccation-tolerance mechanisms. *Seed Sci. Res.* 9: 13.

Parkinson,J., Guiliano,D. and Blaxter, M. 2002. Making sense of EST sequences by CLOBBing them. *BMC Bioinf.* 3:31.

Peters, S., Mundree, S.G., Thomson, J.A., Farrant, J.M. and Keller, F. 2007. Protection Mechanisms in the Resurrection Plant *Xerophyta viscosa* (Baker): Both Sucrose and Raffinose Family Oligosacharides (RFOs) Accumulate in Leaves in Response to Water Deficit. *J. Exp. Bot.* (In Press).

Phillips J.R., Oliver, M.J. and Bartels, D. 2002. "Molecular Genetics of Desiccation and Tolerant Systems." In *Desiccation and Survival in Plants – Drying without Drying,* edited by Michael Black and Hugh William Pritchard, 319. Wallingford, New York: CABI Publishing.

Piatkowski, D., Schneider, K., Salamini, F and Bartels, D. 1990. Characterisation of five abscisic acid-responsive cDNA clones isolated from the desiccation-tolerant plant *Craterostigma plantigineum* and their relationship to other water-stress genes. *Plant Phys.* 94: 1682.

Piecyk, M., Wax, S., Beck, A. R. P., Kedersha, N., Gupta, M., Maritim, B, Chen, S., Gueydan, C., Kruys, V., Streuli, M. and Anderson, P. 2000. TIA-1 is a translational silencer that selectively regulates the expression of TNF-a. *EMBO J.* 19: 4154

Porembski, S and Barthlott, W. 2000. Granitic and gneissic outcrops (inselbergs) as centers of diversity for desiccationtolerant vascular plants. *Plant Ecol.* 151: 19.

Poulsen, K.N. and Eriksen, E.N. 1992. Physiological aspects of recalcitrance in embryonic axes of Quercus robur L. Seed Sci. Res. 2:215.

Pretrelski, S.J., Tedeschi, N., Arkawa, T. and Carpenter, J.F. 1993. Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers. *Biophysical J.* 65: 661.

Proctor, M. C. F. 1990. The Phys.ogical basis of bryophyte production. *Bot. J. Linn. Soc*. 104:61.

Proctor, M. C. F. and V. C. Pence. 2002. Vegetative Tissues: Bryophytes, vascular resurrection plants, and vegetative propogules. *In* M. Black and H.W. Pritchard (eds.), *Desiccation and survival in plants: Drying without dying*, pp. 207–237. CABI Publishing, Wallingford, Oxon.

Raychaudhuri, S. 2000. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac. Symp. Biocomp*. 455: 100

Rabbani M.A., Maruyama K., Abe H., Khan M.A., Katsura K., Ito Y., Yoshiwara K., Seki M., Shinozaki K. and Yamaguchi-Shinozaki K. 2003. Monitoring expression profiles of rice genes under cold,.drought, and high-salinity stresses and abscisic acid application.using cDNA microarray and RNA gel-blot analyses. *Plant Phys.* 133: 1755.

Rajeevan, M.S., Vernon, S.D., Taysavang, N. and Unger, E.R. 2001. Validation of array based gene expression profiles by real-time (kinetic) RT-PCR. *J. Mol. Diagn* 3: 26.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C. and Job, D. 2004. The effect of α-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Phys.* 134: 1598

Ramanjulu, S., Bartels, D. 2002. Drought and dessication-induced modulation of gene expression in plants. *Plant Cell. Environ.* 25: 141.

R Development Core Team 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reyes J.M., Rodrigo, M., Colmenero-Flores, J. Gil, A., Garay-Arroyo, F., Campos, F., Salamini, F., Bartels, D. and Covarrubias., A. 2005. Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects *in vitro. Plant Cell Environ.* 28**:** 709.

Rinne, P. L. H., Kaikuranta. P. L. M., van der Plas. L. H. W., and C. van der Schoot. 1999. Dehydrins in cold-acclimated apices of birch (*Betula pubescens* Ehrh.): Production, localization and potential role in rescuing enzyme function during dehydration. *Planta* 209:377.

Roberts, E.H. 1973. Predicting the storage life of seeds. *Seed Sci Technol* 1: 499

Rogerson NE, Matthews S (1977) Respiratory and carbohydrate changes in developing pea (*Pisum sativum*) seeds in relation to their ability to withstand desiccation. *J Exp Bot* 28:304.

Roos, Y.H. 1995. *Phase transitions in Food*. Academic Press, London.

Rozen, S and Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Eds: Krawetz, S., Misener, S. and Totowa, N.J. *Humana Press*:365.

Russouw, P.S., Farrant, J., Brandt, W., Maeder, D. and Lindsey, G.G. 1995. Isolation and characterization of a heatsoluble protein from pea (*Pisum sativum*) embryos. *Seed Sci. Res* 5: 137.

Russouw, P.S., Brandt, W.F., Farrant, J.M. and Lindsey, G.G. 1997. Chaperone-like activity of pea seed P11 protein. *Seed Sci. Res.* 7: 117.

Ruuska, S.A., Girke, T., Benning, C. and ohlrogge, J.B. 2002. Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* 14: 1191.

Saal, L.H., Troein, C., Vallon-Christersson, J., Gruvberger, S., Borg, A. and Peterson, C. 2002. BioArray Software Environment: A Platform for Comprehensive Management and Analysis of Microarray Data. *Genome Biol.* 3(8): software 0003.1-0003.6

Sales, K., Brandt, W., Rumbak, E. and Lindsey, G. 2000. The LEA like protein HSP 12 in *Saccharomyces cerevisiae* has a plasma membrane location and protects membranes against desiccation and ethanol-induced stress. *Biochim. Biophys. Acta—Biomembranes* 1463:267.

Salmen Espindola, L, Noin, M., Corineau, F. and Côme, D. 1994. Cellular and metabolic damage induced by desiccation in recalcitrant *Araucaria angustifolia* embryos. *Seed Sci. Res.* 4: 193.

Schena, M., Shalon, D., Davis, R. and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a cDNA microarray. *Sci.* 270: 467.

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. & Lohmann, J. U. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nat. Gen.* 37: 501.

Scott, P. 2000. Resurrection plants and the secrets of eternal leaf. *Ann. Bot.* 85:159.

Seel, W., Hendry, G.A.F and Lee, J.A. 1992a. Effects of desiccation on some activated oxygen processing enzymes and antio-oxidants in mosses. *J. Exp. Bot.* 43:1031.

Seel, W.E., Baker, N.R. and Lee, J.A. 1992b. Analysis of the decrease in photosynthesis on desiccation of mosses from xeric and hydric environments. *Phys. Planta.* 86: 451.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. 2002. Monitoring the expression of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. 2004. RIKEN Arabidopsis full length (RAFL) cDNA and its application for expression profiling under abiotic stress conditions. *J. Exp. Bot.* 55: 213.

Senaratna, T. and McKersie, B.D. 1983. Dehydration injury in germinating soybean (*Glycine max* L. Merr) seeds. *Plant Phys.* 72: 620.

Senaratna, T. and McKersie, B.D. 1986. Loss of desiccation tolerance during seed germination: a free radical mechanism of injury. In: Leopold, A.C. (ed.) *Membranes, Metabolism and Dry Organisms*. Cornell University Press, Ithica, New York, pp. 85.

Shaw, G. and Kamen, R. 1986. A conserved AU sequence from the 30 untranslated region of GM–CSF mRNA mediates selective mRNA degradation. *Cell* 46: 659

Sherwin, H.W and Farrant, J.M. 1998. Protection mechanism against excess light in the resurrection plants *Craterostigma wilmsii* and *Xerophyta viscosa. Plant Growth Reg.* 24: 203.

Shen, Q., Uknes, S.J. and Ho, T.-H.D. 1993. Hormonal response complex in a novel abscisic acid and cycloheximide-inducible barley gene. *J. Biol. Chem*. 268: 23652.

Shen, Q. and Ho, T.-H.D. 1995. Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA responsive complexes each containing a G-box and a novel cisacting element. *Plant Cell* 7: 295.

Shinozaki, K., Yamaguchi-Shinozaki, K. and Seki, M. 2003. Regulatory network of gene expression in the drought and cold stress responses. *Curr. Opin. Plant Biol.* 6: 410.

Simpson, S.D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K. and Yamaguhci-Shinozaki K. 2003. Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. *Plant J.* 33: 259.

Sivamani, E., Bahieldrin, A., Wraith, J. M., Al-Niemi, T., Dyer, W. E., Ho, T-H. D and Qu. R. 2000. Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. *Plant Sci.* 155:1.

Slade, L. and Levine, H. 1991. A food polymer Sci. approach to structure-property relationships in aqueous food systems: nonequilibrium behaviour of carbohydrate-water systems. In: Levine, H. and Slade, L. (eds.) *Water Relationships in Food.* Plenum Press, New York, pp. 29-101.

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. and Lohmann, J. U. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nat. Gen.* 37: 501.

Smyth, G. K. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Gen. Mol. Biol* 3:Article 3.

Smart, R.E. and Bingham, G.E. 1974. Rapid Estimates of Relative Water Content. *Plant Physiol.* 53:258

Smirnoff, N. 1993. The role of active oxygen in the response of plants to water deficit and desiccation. *New Phytol.* 125: 214.

Smyth, G. K. and Speed, T. P. 2003. Normalization of cDNA microarray data. *Methods* 31: 265.

Schmittgen, T.D., Zakrajsek, B.A., Mills, A.G., Gorn, V., Singer, M.J. and Reed, M.W. 2000. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Ann. Biochem.* 285:194

Soeda, Y., Konings, M.C.J.M., Vorst, O., van Houwelingen, A.M.M.L., Stoopen, G.M., Maliepaard, C.A., Kodde, J., Bino, R.J., Groot, S.P.C. and van der Geest, A.H.M. 2005. Gene expression programs during Brassica oleracea seed maturation, osmopriming, and germination are indicators of the progression of the germination process and the stress tolerance level. *Plant Phys.*137: 354.

Stacy, R.A.P., Nordeng, T.W., Culiáñez-Maciá F.A. and Aalen, R.B. 1999. The dormancy-related peroxiredoxin antioxidant PER1, is located to the Nucl.leus in barley embryo and aleurone. *Ann. Bot.* 19: 1.

Stekel, D. 2003. *Microarray Bioinformatics*. Cambridge University Press, Cambridge.

Steponkus, P.L. 1979. Freeze-thaw induced lesions in the plasma membrane. In: Lyons, J.M., Graham, D.G. and Raison, J.K. (eds.) *Low Temperature Stress in Crop Plants*. Academic Press, *New York*, pp. 231.

Steponkus, P.L. and Lynch, D.V. 1989. Freeze/thaw-induced destabilization of the plasma membrane and the effects of cold acclimation. *J. Bioenerg. Biomemb.* 21: 21.

Steponkus, P.L., Uemura, M. and Webb, M.S. 1995. Freeze-induced destabilization of cellular membranes and lipid bilayers. In: Disalvo, E.A. and Simon, S.A. (eds.) *Permeability and Stability of Lipid Bilayers.* CRC Press, Boca Raton, Florida, pp. 77.

Stefanov, K., Markovska, Y., Kimenov, G. and Popov, S. 1992. Lipid and sterol changes in leaves of *Haberlea rhodopensis* and *Ramonda serbica* at transition from biosis into anabiosis and vice versa caused by water stress. *Phytochem*. 31:2309.

Still, D.W., Kovach, D.A. and Bradford, K.J. 1994. Development of desiccation tolerance during embryogenesis in rice (*Oryza sativa*) and wild rice (*Zizania palustris*). *Plant Physiol* .104:431

Stitt, M., Müller, C., Matt, P., Gibon, Y., Carillo, P., Morcuende, R., Scheible, W-R. and Krapp, A. 2002. Steps towards an integrated view of nitrogen metabolism. *J. Exp. Bot*. 53: 959.

Storey, J.D. and Tibshirani, R. 2003. Statistical significance for genomewide studies. *PNAS.* 100:9440

Sun, W.Q. and Leopold, A.C. 1997. Cytoplasmic vitrification and survival of anhydrobiotic organisms. *Comp. Biochem. Phys.* 117A: 327.

Swire-Clark, G. A. and Marcotte, W.R. Jr. 1999. The wheat LEA protein Em functions as an osmoprotective molecule in *Saccharomyces cerevisiae*. *Plant Mol. Biol*. 39:117.

Tamayo, P. et al. 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *PNAS*. 96: 2907.

Tavazoie, S. et al. 1999. Systematic determination of genetic network architecture. *Nat. Genet*. 22: 281

Tetteroo, F.A.A., Peters, A.H.J.L., Hoekstra, F.A., Van der Plas, L.H.W. and Hagendoorn, M.J.M. 1995. ABA reduces respiration and sugar metabolism in developing carrot (Daucus carota L.) embryoids. *J. Plant Physiol.* 145:477.

Tran, L.S.P., Nakashima, K., Sakuma, Y., Simpson, S.D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2004. Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* 16: 2481.

Tuba Z., Lichtenthaler, H.K., Csintalan, Zs. and Pocs T. 1993a. Regreening of the desiccated leaves of the poikilochlorophyllous *Xerophyta scabrida* upon rehydration. *Plant Phys.* 142: 103.

Tuba Z., Lichtenthaler, H.K., Maroti, I. and Csintalan Zs. 1993b. Resynthesis of thylakoids and functional chloroplasts in the desiccated leaves of the poikilochlorophyllous *Xerophyta scabrida* upon rehydration. *Plant Phys.* 142: 742.

Tuba Z., Proctor M., and Csintalan Zs. 1998. EcoPhys.ogical responses of homoichlorophyllous and poikilochlorophyllous desiccation tolerant plants: a comparison and an ecological perspective. *Plant Growth Reg.* 24: 211.

Van Heerden, J., Walford, S.A., Shen, A and Illing, N. 2007. A framework for the informed normalization of printed microarrays. *SA. J. Sci*. In press.

Van der Willigen C., Mundree, S.G., Pammenter, N.W. and Farrant, J.M. 2004. Mechanical stabilisation in desiccated vegetative tissues of the resurrection grass *Eragrostis nindensis*: does an alpha TIP and/or subcellular compartmentalization play a role? *J . Exp. Bot.* 55: 651.

Vasil, V., Marcotte, Jr, W.R., Rosenkrans, L., Cocciolone, S.M., Vasil, I.K., Quatrano, R.S. and McCarty, D.R. 1995. Overlap of Viviparous 1 (VP1) and abscisic acid response elements in the Em promoter:G-box elements are sufficient but not necessary for VP1 transactivation. *Plant Cell* 7: 1511.

Vertucci, C.W. and Farrant, J.M. 1995. "Acquisition and loss of desiccation tolerance." In *Seed Development and Germination*, Edited by J. Kigel and G. Galili. pp 237. New York: Marcel Dekker Press Inc.

Vertucci, C.W. and Leopold, A.C. 1984. Bound water in soybean seed and its relation to respiration and imbibitional damage. *Plant Phys.* 75: 114.

Vertucci, C.W. and Leopold, A.C. 1986. Phys.ogical activities associated with hydration levels in seeds. In: Leopold, A.C. (ed.) *Membranes, Metabolism and Dry Organisms*. Cornell University Press, Ithica, New York, pp. 35.

Vertucci, C.W. and Roos, E.E. 1990. Theoretical basis of protocols for seed storage. *Plant Phys.* 94: 1019.

Vicre, M., Sherwin, H.W., Driouich, A., Jaffer, M., Jauneau, A., and Farrant, J.M. 1999. Cell wall properties of hydrated and dry leaves of the resurrection plant *Craterostigma wilmsii. Plant Phys.* 155: 719.

Vicre M., Farrant, J.M., Gibouin, D and Driouich, A. 2003. Resurrection plants: how to cope with desiccation? *Rec. Res Dev. Plant Biol.* 3: 69.

Vicre, M., Farrant, J.M. and Driouich, A. 2004a Insights into the mechanisms of desiccation tolerance among resurrection plants. *Plant Cell. Environ.* 27: 1329.

Vicre, M., Lerouxel, O., Farrant, J.M., Lerouge, P., and Driouich, A. 2004b. Composition and desiccation induced alterations of the cell wall in the resurrection plant *Craterostigma wilmsii Physiol. Planta.* 120: 229

Vierling, E. 1991. The roles of heat shock proteins in plants. *Annual Rev.Plant Phys. Plant Mol. Biol.* 42: 579.

Walburger, D.K, Afonina, I.A. and Wydro, R. 2001. An improved real time PCR method for simultaneous detection of C282Y and H63D mutations in the HFE gene associated with hereditary hemochromatosis. *Mutat Res*. 432:69

Walford S-A, Thomson JA, Farrant JM, Mundree SG (2004) Isolation and characterisation of a novel dehydration-induced Grp94 homologue from the resurrection plant *Xerophyta viscosa*. *S .A J.Bot* .70:1

Walker, N. 2002. A technique whose time has come. *Sci.* 296:557.

Walters, C., Pammenter, N.W., Berjak, P. And Crane, J. 2001. Desiccation damage, accelerated aging and metabolism in desiccation tolerant and sensitive seeds. *Seed Sci. Res.* 11, 135.

Walters, C., Farrant, J.M., Pammenter, N.W. and Berjak, P. 2002. “Desiccation and Damage.” In *Desiccation and Survival in Plants: Drying without Dying,* Edited by M. Black and H. Pritchard, pp 263. London: CABI Publishing.

Wang, H., Cao, G.H. and Prior, R.L. 1996. Total antioxidant capacity of fruits. *J. Agri. Food Chem.* 44: 701.

Wasmuth, J.D. and Blaxter, M.L. 2004. prot4EST: Translating Expressed Sequence Tags from neglected genomes. *BMC Bioinf.* 5:187

Waters, E.R., Lee, G.J. and Vierling, E. 1996. Evolution, structure and function of the small heat shock proteins in plants. *J. Exp. Bot.* 47:325.

Wehmeyer, N., Hernandez, L., Finkelstein, R. and Vierling, E. 1996. Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. *Plant Phys.* 112: 747.

Wehmeyer, N. and Vierling, E. 2000. The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. *Plant Phys.* 122: 1099.

Whittaker, A., Bochicchio, A., Vazzana, C., Lindsey, G. and Farrant, J.M. 2001. Changes in leaf hexokinase activity and metabolite levels in response to drying in the desiccation tolerant species *Sporobolus stapfianus* and *Xerophyta viscosa. J.Exp Bot*. 352: 961.

Whittaker, A., Martinelli, T., Farrant, J.M, Bochicchio, A. and Vazzana, C. 2007. Sucrose phosphate synthase activity and the co-ordination of carbon partitioning during sucrose and amino acid accumulation in desiccation-tolerant leaf material of the C4 resurrection plant *Sporobolus stapfianus* during Dehydration. *J. Exp. Bot.* 58: 3775

Williams, R.J. and Leopold, A.C. 1989. Measuring and modeling environmental influences on photosynthetic gas exchange in *Sphagnum* and *Pleurozium. Plant Cell Environ.* 21: 555.

Wise, M.J. 2003. LEAping to conclusions: a computational reanalysis of late emryogenesis abundant proteins and their possible roles. *BMC bioinf.* 4: 52

Wise, M.J. and Tunnacliff, A. 2004. POPP the question: what do LEA proteins do. *Trends Plant Sci.* 9: 1360.

Wit, E. and McClure, J. 2004. Statistics for Microarrays: Design, Analysis and Inference. Eds. Wit, E. and McClure, J., Chichester, Wiley.

Wolff, S.P., Garner, A. and Dean, R.T. 1986. Free radicals, lipids and protein degradation. *Trends. Biochem. Sci.* 11: 27.

Wolkers, W.F., Alberda, M., Koornneef, M., Léon-Kloosterziel, K.M. and Hoekstra, F.A. 1998a. Properties of proteins and the glassy matrix in maturation-defective mutant seeds of *Arabidopsis thaliana*. *Plant J.* 16: 133.

Wolkers, W.F., Oldenhof, H., Alberda, M.,and Hoekstra, F.A. 1998b. A fourier transform infrared microspectroscopy study of sugar glasses: application to anhydrobiotic higher plant cells. *Biochem. Biophys. Acta* 1379: 83.

Wolkers, W. F., McCready,S., Brandt,W., Lindsey,G. G. and F. A. Hoekstra. 2001. Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro. *Biochem. Biophys. Acta* 1544:196.

Wood, A.J. and Oliver, M.J. 1999. Translational control in plant stress: the formation of messenger riboNucl.leoprotein particles (mRNPs) in response to desiccation of Tortula ruralis gametophytes. *Plant J.* 18:359.

Wood, A.J., Duff, R.J. and Oliver, M.J. 2000. The translational apparatus of Tortula ruralis: polysomal retention of transcripts encoding the ribosomal proteins RPS14, RPS16 and RPL23 in desiccated and rehydrated gametophytes. *J. Exp. Bot*. 51:1655.

Xu, D., Duan, X., Wang, B., Hong,B., Ho, T-H. D. and Wu, R. 1996. Expression of a late embryogenesis abundant protein gene, *HVA1,* from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Phys.* 110:249.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucl. Acids Res.* 30:e15.

Yin, J.L., Shackel, N.A. and Zekry, A. 2001. Real-time reverse-transcriptase polymerase chain reaction (RT-PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green. *J. Immunol. Cell Biol*.79:213.

Yuan, J.S., Reed, A., Chen, F and Stewart Jr, C.N. 2006. Statistical analysis of real-time PCR data *BMC Bioinf.* 7:85.

Zeng, Y. and Kermode, A.R. 2004. A gymnosperm ABI3 gene functions in 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 Mol. Biol* .56:731.

Zhang, L., Ohta, A., Takagi,M. and Imai, R. 2000. Expression of plant group 2 and group 3 *lea* genes in *Saccharomyces cerevisiae* revealed functional divergence among LEA proteins. *J. Biochem.* 127: 611.

Appendix

A.2.1.

```
setwd("C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr files")

library(Biobase)
library(limma)
library(smida)

#######Read GPR data files
DL1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DL1.txt",sep="\t",header=TRUE)
DL2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DL2.txt",sep="\t",header=TRUE)
DL3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DL3.txt",sep="\t",header=TRUE)

HL1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HL1.txt",sep="\t",header=TRUE)
HL2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HL2.txt",sep="\t",header=TRUE)
HL3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HL3.txt",sep="\t",header=TRUE)

DR1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DR1.txt",sep="\t",header=TRUE)
DR2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DR2.txt",sep="\t",header=TRUE)
DR3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DR3.txt",sep="\t",header=TRUE)

HR1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HR1.txt",sep="\t",header=TRUE)
HR2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HR2.txt",sep="\t",header=TRUE)
HR3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HR3.txt",sep="\t",header=TRUE)

seed1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\seed1.txt",sep="\t",header=TRUE)
seed2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\seed2.txt",sep="\t",header=TRUE)
seed3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\seed3.txt",sep="\t",header=TRUE)

########## Extract intensity values and put in 7200x15 matrix
genedata<-matrix(0,nrow(HL3),ncol=15)
genedata[,1]<-DL1[,"F532.Mean"]
genedata[,2]<-DL2[,"F532.Mean"]
genedata[,3]<-DL3[,"F532.Mean"]
genedata[,4]<-HL1[,"F532.Mean"]
genedata[,5]<-HL2[,"F532.Mean"]
genedata[,6]<-HL3[,"F532.Mean"]
genedata[,7]<-DR1[,"F532.Mean"]
genedata[,8]<-DR2[,"F532.Mean"]
genedata[,9]<-DR3[,"F532.Mean"]
genedata[,10]<-HR1[,"F532.Mean"]
```

```
genedata[,11]<-HR2[,"F532.Mean"]
genedata[,12]<-HR3[,"F532.Mean"]
genedata[,13]<-seed1[,"F532.Mean"]
genedata[,14]<-seed2[,"F532.Mean"]
genedata[,15]<-seed3[,"F532.Mean"]
genedata<-log2(genedata)

#Spatial plots
library(sma)

XHsetup<-init.grid()
#8
#4
#15
#15

#plot F532.Mean values or log2 data
#use crit1=0.4 or 0.5 because this shows the variation in the data
plot.spatial(genedata[,1],XHsetup,crit1=0.5,main="Spatial Plot of DL1", sub="Mean Log Intensity Raw Data")
#after spatial normalisation
plot.spatial(dat.spat[1,],XHsetup,crit1=0.5,main="Spatial Plot of DL1",sub="Mean Log Intensity Spatially Normalised")

#Boxplot
boxplot(genedata[,1],genedata[,2],genedata[,3],genedata[,4],genedata[,5],genedata[,6],
genedata[,7],genedata[,8],genedata[,9],genedata[,10],genedata[,11],genedata[,12],
genedata[,13],genedata[,14],genedata[,15],names=c("DL1","DL2","DL3","HL1","HL2","HL3","DR1","DR2","DR3","HR1","H
R2","HR3","Seed1","Seed2","Seed3"),main="Boxplots of Raw Data",ylab="log2 expression value",
xlab="Condition")

#histogram
library(affy)
plotDensity(genedata)
```

A.2.2.

```
setwd("C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis")

library(Biobase)
library(limma)
library(smida)

#######Read csv data files
controls<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\controls.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
drought<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\drought.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
seed<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\seed silique.csv",header =
TRUE, sep = ",", quote="\"", dec=".")

########## Extract intensity values and put in 22810 x 215 matrix
genedata<-matrix(0,nrow(controls),ncol=87)
genedata[,1]<-controls[,"Signal_Shoots_0h_Rep1"]
genedata[,2]<-controls[,"Signal_Shoots_0h_Rep2"]
genedata[,3]<-controls[,"Signal_Roots_0h_Rep1"]
genedata[,4]<-controls[,"Signal_Roots_0h_Rep2"]
genedata[,5]<-controls[,"Signal_Shoots_0.25h_Rep1"]
genedata[,6]<-controls[,"Signal_Shoots_0.25h_Rep2"]
genedata[,7]<-controls[,"Signal_Roots_0.25h_Rep1"]
genedata[,8]<-controls[,"Signal_Roots_0.25h_Rep2"]
genedata[,9]<-controls[,"Signal_Shoots_0.5h_Rep1"]
```

```
genedata[,10]<-controls[,"Signal_Shoots_0.5h_Rep2"]
genedata[,11]<-controls[,"Signal_Roots_0.5h_Rep1"]
genedata[,12]<-controls[,"Signal_Roots_0.5h_Rep2"]
genedata[,13]<-controls[,"Signal_Shoots_1.0h_Rep1"]
genedata[,14]<-controls[,"Signal_Shoots_1.0h_Rep2"]
genedata[,15]<-controls[,"Signal_Roots_1.0h_Rep1"]
genedata[,16]<-controls[,"Signal_Roots_1.0h_Rep2"]
genedata[,17]<-controls[,"Signal_Shoots_3.0h_Rep1"]
genedata[,18]<-controls[,"Signal_Shoots_3.0h_Rep2"]
genedata[,19]<-controls[,"Signal_Roots_3.0h_Rep1"]
genedata[,20]<-controls[,"Signal_Roots_3.0h_Rep2"]
genedata[,21]<-controls[,"Signal_Shoots_4.0h_Rep1"]
genedata[,22]<-controls[,"Signal_Shoots_4.0h_Rep2"]
genedata[,23]<-controls[,"Signal_Roots_4.0h_Rep1"]
genedata[,24]<-controls[,"Signal_Roots_4.0h_Rep2"]
genedata[,25]<-controls[,"Signal_Shoots_6.0h_Rep1"]
genedata[,26]<-controls[,"Signal_Shoots_6.0h_Rep2"]
genedata[,27]<-controls[,"Signal_Roots_6.0h_Rep1"]
genedata[,28]<-controls[,"Signal_Roots_6.0h_Rep2"]
genedata[,29]<-controls[,"Signal_Shoots_12.0h_Rep1"]
genedata[,30]<-controls[,"Signal_Shoots_12.0h_Rep2"]
genedata[,31]<-controls[,"Signal_Roots_12.0h_Rep1"]
genedata[,32]<-controls[,"Signal_Roots_12.0h_Rep2"]
genedata[,33]<-controls[,"Signal_Shoots_24.0h_Rep1"]
genedata[,34]<-controls[,"Signal_Shoots_24.0h_Rep2"]
genedata[,35]<-controls[,"Signal_Roots_24.0h_Rep1"]
genedata[,36]<-controls[,"Signal_Roots_24.0h_Rep2"]

genedata[,37]<-drought[,"Signal_Shoots_0.25h_Rep1"]
genedata[,38]<-drought[,"Signal_Shoots_0.25h_Rep2"]
genedata[,39]<-drought[,"Signal_Roots_0.25h_Rep1"]
genedata[,40]<-drought[,"Signal_Roots_0.25h_Rep2"]
genedata[,41]<-drought[,"Signal_Shoots_0.5h_Rep1"]
genedata[,42]<-drought[,"Signal_Shoots_0.5h_Rep2"]
genedata[,43]<-drought[,"Signal_Roots_0.5h_Rep1"]
genedata[,44]<-drought[,"Signal_Roots_0.5h_Rep2"]
genedata[,45]<-drought[,"Signal_Shoots_1.0h_Rep1"]
genedata[,46]<-drought[,"Signal_Shoots_1.0h_Rep2"]
genedata[,47]<-drought[,"Signal_Roots_1.0h_Rep1"]
genedata[,48]<-drought[,"Signal_Roots_1.0h_Rep2"]
genedata[,49]<-drought[,"Signal_Shoots_3.0h_Rep1"]
genedata[,50]<-drought[,"Signal_Shoots_3.0h_Rep2"]
genedata[,51]<-drought[,"Signal_Roots_3.0h_Rep1"]
genedata[,52]<-drought[,"Signal_Roots_3.0h_Rep2"]
genedata[,53]<-drought[,"Signal_Shoots_6.0h_Rep1"]
genedata[,54]<-drought[,"Signal_Shoots_6.0h_Rep2"]
genedata[,55]<-drought[,"Signal_Roots_6.0h_Rep1"]
genedata[,56]<-drought[,"Signal_Roots_6.0h_Rep2"]
genedata[,57]<-drought[,"Signal_Shoots_12.0h_Rep1"]
genedata[,58]<-drought[,"Signal_Shoots_12.0h_Rep2"]
genedata[,59]<-drought[,"Signal_Roots_12.0h_Rep1"]
genedata[,60]<-drought[,"Signal_Roots_12.0h_Rep2"]
genedata[,61]<-drought[,"Signal_Shoots_24.0h_Rep1"]
genedata[,62]<-drought[,"Signal_Shoots_24.0h_Rep2"]
genedata[,63]<-drought[,"Signal_Roots_24.0h_Rep1"]
genedata[,64]<-drought[,"Signal_Roots_24.0h_Rep2"]

genedata[,65]<-seed[,"Signal_76_B"]
genedata[,66]<-seed[,"Signal_76_C"]
```

```
genedata[,67]<-seed[,"Signal_77_D"]
genedata[,68]<-seed[,"Signal_77_E"]
genedata[,69]<-seed[,"Signal_77_F"]
genedata[,70]<-seed[,"Signal_78_D"]
genedata[,71]<-seed[,"Signal_78_E"]
genedata[,72]<-seed[,"Signal_78_F"]
genedata[,73]<-seed[,"Signal_79_A"]
genedata[,74]<-seed[,"Signal_79_B"]
genedata[,75]<-seed[,"Signal_79_C"]
genedata[,76]<-seed[,"Signal_81_A"]
genedata[,77]<-seed[,"Signal_81_B"]
genedata[,78]<-seed[,"Signal_81_C"]
genedata[,79]<-seed[,"Signal_82_A"]
genedata[,80]<-seed[,"Signal_82_B"]
genedata[,81]<-seed[,"Signal_82_C"]
genedata[,82]<-seed[,"Signal_83_A"]
genedata[,83]<-seed[,"Signal_83_B"]
genedata[,84]<-seed[,"Signal_83_C"]
genedata[,85]<-seed[,"Signal_84_A"]
genedata[,86]<-seed[,"Signal_84_B"]
genedata[,87]<-seed[,"Signal_84_D"]

genedata<-log2(genedata)

#Boxplots
boxplot(genedata[,45],genedata[,46],genedata[,13],genedata[,14],genedata[,47],genedata[,48],
genedata[,15],genedata[,16],genedata[,85],genedata[,86],genedata[,87],
names=c("DehL1","DehL2","CL1","CL2","DehR1","DehR2","CR1","CR2","Seed1","Seed2","Seed3"),main="Boxplots of A.
thaliana Normalised Data",ylab="log2 expression value",
xlab="Condition")

#Histogram
library(affy)
plotDensity(genedata)
```

A.2.3.

```
setwd("C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis")

library(Biobase)
library(limma)
library(smida)

#######Read csv data files
controls<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\controls.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
osmotic<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\osmotic.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
seed<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\seed silique.csv",header =
TRUE, sep = ",", quote="\"", dec=".")

########## Extract intensity values and put in 22810 x 215 matrix
genedata<-matrix(0,nrow(controls),ncol=83)
genedata[,1]<-controls[,"Signal_Shoots_0h_Rep1"]
genedata[,2]<-controls[,"Signal_Shoots_0h_Rep2"]
genedata[,3]<-controls[,"Signal_Roots_0h_Rep1"]
genedata[,4]<-controls[,"Signal_Roots_0h_Rep2"]
genedata[,5]<-controls[,"Signal_Shoots_0.25h_Rep1"]
genedata[,6]<-controls[,"Signal_Shoots_0.25h_Rep2"]
genedata[,7]<-controls[,"Signal_Roots_0.25h_Rep1"]
```

```
genedata[,8]<-controls[,"Signal_Roots_0.25h_Rep2"]
genedata[,9]<-controls[,"Signal_Shoots_0.5h_Rep1"]
genedata[,10]<-controls[,"Signal_Shoots_0.5h_Rep2"]
genedata[,11]<-controls[,"Signal_Roots_0.5h_Rep1"]
genedata[,12]<-controls[,"Signal_Roots_0.5h_Rep2"]
genedata[,13]<-controls[,"Signal_Shoots_1.0h_Rep1"]
genedata[,14]<-controls[,"Signal_Shoots_1.0h_Rep2"]
genedata[,15]<-controls[,"Signal_Roots_1.0h_Rep1"]
genedata[,16]<-controls[,"Signal_Roots_1.0h_Rep2"]
genedata[,17]<-controls[,"Signal_Shoots_3.0h_Rep1"]
genedata[,18]<-controls[,"Signal_Shoots_3.0h_Rep2"]
genedata[,19]<-controls[,"Signal_Roots_3.0h_Rep1"]
genedata[,20]<-controls[,"Signal_Roots_3.0h_Rep2"]
genedata[,21]<-controls[,"Signal_Shoots_4.0h_Rep1"]
genedata[,22]<-controls[,"Signal_Shoots_4.0h_Rep2"]
genedata[,23]<-controls[,"Signal_Roots_4.0h_Rep1"]
genedata[,24]<-controls[,"Signal_Roots_4.0h_Rep2"]
genedata[,25]<-controls[,"Signal_Shoots_6.0h_Rep1"]
genedata[,26]<-controls[,"Signal_Shoots_6.0h_Rep2"]
genedata[,27]<-controls[,"Signal_Roots_6.0h_Rep1"]
genedata[,28]<-controls[,"Signal_Roots_6.0h_Rep2"]
genedata[,29]<-controls[,"Signal_Shoots_12.0h_Rep1"]
genedata[,30]<-controls[,"Signal_Shoots_12.0h_Rep2"]
genedata[,31]<-controls[,"Signal_Roots_12.0h_Rep1"]
genedata[,32]<-controls[,"Signal_Roots_12.0h_Rep2"]
genedata[,33]<-controls[,"Signal_Shoots_24.0h_Rep1"]
genedata[,34]<-controls[,"Signal_Shoots_24.0h_Rep2"]
genedata[,35]<-controls[,"Signal_Roots_24.0h_Rep1"]
genedata[,36]<-controls[,"Signal_Roots_24.0h_Rep2"]

genedata[,37]<-osmotic[,"Signal_Shoots_0.5h_Rep1"]
genedata[,38]<-osmotic[,"Signal_Shoots_0.5h_Rep2"]
genedata[,39]<-osmotic[,"Signal_Roots_0.5h_Rep1"]
genedata[,40]<-osmotic[,"Signal_Roots_0.5h_Rep2"]
genedata[,41]<-osmotic[,"Signal_Shoots_1.0h_Rep1"]
genedata[,42]<-osmotic[,"Signal_Shoots_1.0h_Rep2"]
genedata[,43]<-osmotic[,"Signal_Roots_1.0h_Rep1"]
genedata[,44]<-osmotic[,"Signal_Roots_1.0h_Rep2"]
genedata[,45]<-osmotic[,"Signal_Shoots_3.0h_Rep1"]
genedata[,46]<-osmotic[,"Signal_Shoots_3.0h_Rep2"]
genedata[,47]<-osmotic[,"Signal_Roots_3.0h_Rep1"]
genedata[,48]<-osmotic[,"Signal_Roots_3.0h_Rep2"]
genedata[,49]<-osmotic[,"Signal_Shoots_6.0h_Rep1"]
genedata[,50]<-osmotic[,"Signal_Shoots_6.0h_Rep2"]
genedata[,51]<-osmotic[,"Signal_Roots_6.0h_Rep1"]
genedata[,52]<-osmotic[,"Signal_Roots_6.0h_Rep2"]
genedata[,53]<-osmotic[,"Signal_Shoots_12.0h_Rep1"]
genedata[,54]<-osmotic[,"Signal_Shoots_12.0h_Rep2"]
genedata[,55]<-osmotic[,"Signal_Roots_12.0h_Rep1"]
genedata[,56]<-osmotic[,"Signal_Roots_12.0h_Rep2"]
genedata[,57]<-osmotic[,"Signal_Shoots_24.0h_Rep1"]
genedata[,58]<-osmotic[,"Signal_Shoots_24.0h_Rep2"]
genedata[,59]<-osmotic[,"Signal_Roots_24.0h_Rep1"]
genedata[,60]<-osmotic[,"Signal_Roots_24.0h_Rep2"]

genedata[,61]<-seed[,"Signal_76_B"]
genedata[,62]<-seed[,"Signal_76_C"]
genedata[,63]<-seed[,"Signal_77_D"]
genedata[,64]<-seed[,"Signal_77_E"]
```

```
genedata[,65]<-seed[,"Signal_77_F"]
genedata[,66]<-seed[,"Signal_78_D"]
genedata[,67]<-seed[,"Signal_78_E"]
genedata[,68]<-seed[,"Signal_78_F"]
genedata[,69]<-seed[,"Signal_79_A"]
genedata[,70]<-seed[,"Signal_79_B"]
genedata[,71]<-seed[,"Signal_79_C"]
genedata[,72]<-seed[,"Signal_81_A"]
genedata[,73]<-seed[,"Signal_81_B"]
genedata[,74]<-seed[,"Signal_81_C"]
genedata[,75]<-seed[,"Signal_82_A"]
genedata[,76]<-seed[,"Signal_82_B"]
genedata[,77]<-seed[,"Signal_82_C"]
genedata[,78]<-seed[,"Signal_83_A"]
genedata[,79]<-seed[,"Signal_83_B"]
genedata[,80]<-seed[,"Signal_83_C"]
genedata[,81]<-seed[,"Signal_84_A"]
genedata[,82]<-seed[,"Signal_84_B"]
genedata[,83]<-seed[,"Signal_84_D"]


genedata<-log2(genedata)

#Boxplot
boxplot(genedata[,57],genedata[,58],genedata[,33],genedata[,34],genedata[,59],genedata[,60],
genedata[,35],genedata[,36],genedata[,81],genedata[,82],genedata[,83],
names=c("OsL1","OsL2","CL1","CL2","OsR1","OsR2","CR1","CR2","Seed1","Seed2","Seed3"),main="Boxplots of A. thaliana
Normalised Data",ylab="log2 expression value",
xlab="Condition")

#Histogram
library(affy)
plotDensity(affy)
```

A.2.4.

```
setwd("C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr files")

library(Biobase)
library(limma)
library(smida)

#######Read GPR data files
DL1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DL1.txt",sep="\t",header=TRUE)
DL2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DL2.txt",sep="\t",header=TRUE)
DL3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DL3.txt",sep="\t",header=TRUE)

HL1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HL1.txt",sep="\t",header=TRUE)
HL2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HL2.txt",sep="\t",header=TRUE)
HL3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HL3.txt",sep="\t",header=TRUE)

DR1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DR1.txt",sep="\t",header=TRUE)
```

```
DR2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DR2.txt",sep="\t",header=TRUE)
DR3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\DR3.txt",sep="\t",header=TRUE)

HR1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HR1.txt",sep="\t",header=TRUE)
HR2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HR2.txt",sep="\t",header=TRUE)
HR3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\HR3.txt",sep="\t",header=TRUE)

seed1<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\seed1.txt",sep="\t",header=TRUE)
seed2<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\seed2.txt",sep="\t",header=TRUE)
seed3<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\seed3.txt",sep="\t",header=TRUE)

########## Extract intensity values and put in 7200x15 matrix
genedata<-matrix(0,nrow(HL3),ncol=15)
genedata[,1]<-DL1[,"F532.Mean"]
genedata[,2]<-DL2[,"F532.Mean"]
genedata[,3]<-DL3[,"F532.Mean"]
genedata[,4]<-HL1[,"F532.Mean"]
genedata[,5]<-HL2[,"F532.Mean"]
genedata[,6]<-HL3[,"F532.Mean"]
genedata[,7]<-DR1[,"F532.Mean"]
genedata[,8]<-DR2[,"F532.Mean"]
genedata[,9]<-DR3[,"F532.Mean"]
genedata[,10]<-HR1[,"F532.Mean"]
genedata[,11]<-HR2[,"F532.Mean"]
genedata[,12]<-HR3[,"F532.Mean"]
genedata[,13]<-seed1[,"F532.Mean"]
genedata[,14]<-seed2[,"F532.Mean"]
genedata[,15]<-seed3[,"F532.Mean"]
genedata<-log2(genedata)


############### Covert data into exprSet object
covdesc<- list("Condition")
names(covdesc) <-"Condition"
geneCov<-as.data.frame(c("DL1","DL2","DL3","HL1","HL2","HL3","DR1",
"DR2","DR3","HR1","HR2","HR3","seed1","seed2","seed3"))
pdata <- new("phenoData", pData=geneCov, varLabels=covdesc)
eset <- new("exprSet", exprs=genedata, phenoData=pdata)

## Write expression matrix to file (tab-delimited text).
## I add the column names before writing it.

ExpData<-exprs(eset)
colnames(ExpData)<-pData(eset)[,1]
#write.table(ExpData,file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr
files\\ExpDatfilt.txt",sep="\t",row.names=FALSE)

### Get IDs and layout info from grp file
genes <- data.frame(ID=DL1$ID,Name=DL1$Name,Row=DL1$Row,Column=DL1$Column)

# set control status
```

```
spottypes<-readSpotTypes()
genes$Status<-controlStatus(spottypes,genes=genes)

dat<-t(exprs(eset))
dim(dat)

# get indeces of control genes from gene ID list read in from GAL file above
# I had a quick browse through your IDs and it seems as though most controls
# contains the sequence "control" in their ID, so I pick all of these out
# (368 of them)

# returns indeces of control genes
control.ind<-grep("control",as.character(genes$ID))

dat.spat<-dat
dat.spat[1,]<-spat.norm(cbind(genes$Row,genes$Column,dat[1,]),ignore=control.ind)$spat.norm.xpr
dat.spat[2,]<-spat.norm(cbind(genes$Row,genes$Column,dat[2,]),ignore=control.ind)$spat.norm.xpr
dat.spat[3,]<-spat.norm(cbind(genes$Row,genes$Column,dat[3,]),ignore=control.ind)$spat.norm.xpr
dat.spat[4,]<-spat.norm(cbind(genes$Row,genes$Column,dat[4,]),ignore=control.ind)$spat.norm.xpr
dat.spat[5,]<-spat.norm(cbind(genes$Row,genes$Column,dat[5,]),ignore=control.ind)$spat.norm.xpr
dat.spat[6,]<-spat.norm(cbind(genes$Row,genes$Column,dat[6,]),ignore=control.ind)$spat.norm.xpr
dat.spat[7,]<-spat.norm(cbind(genes$Row,genes$Column,dat[7,]),ignore=control.ind)$spat.norm.xpr
dat.spat[8,]<-spat.norm(cbind(genes$Row,genes$Column,dat[8,]),ignore=control.ind)$spat.norm.xpr
dat.spat[9,]<-spat.norm(cbind(genes$Row,genes$Column,dat[9,]),ignore=control.ind)$spat.norm.xpr
dat.spat[10,]<-spat.norm(cbind(genes$Row,genes$Column,dat[10,]),ignore=control.ind)$spat.norm.xpr
dat.spat[11,]<-spat.norm(cbind(genes$Row,genes$Column,dat[11,]),ignore=control.ind)$spat.norm.xpr
dat.spat[12,]<-spat.norm(cbind(genes$Row,genes$Column,dat[12,]),ignore=control.ind)$spat.norm.xpr
dat.spat[13,]<-spat.norm(cbind(genes$Row,genes$Column,dat[13,]),ignore=control.ind)$spat.norm.xpr
dat.spat[14,]<-spat.norm(cbind(genes$Row,genes$Column,dat[14,]),ignore=control.ind)$spat.norm.xpr
dat.spat[15,]<-spat.norm(cbind(genes$Row,genes$Column,dat[15,]),ignore=control.ind)$spat.norm.xpr

save(dat.spat,file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr files\\datspat.R")
#load(file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr files\\datspat.R")

dat.bkg<-dat.spat
dat.bkg[1,]<-bkg.norm(xpr=dat.bkg[1,],empty=200,method="deterministic")
dat.bkg[2,]<-bkg.norm(xpr=dat.bkg[2,],empty=200,method="deterministic")
dat.bkg[3,]<-bkg.norm(xpr=dat.bkg[3,],empty=200,method="deterministic")
dat.bkg[4,]<-bkg.norm(xpr=dat.bkg[4,],empty=200,method="deterministic")
dat.bkg[5,]<-bkg.norm(xpr=dat.bkg[5,],empty=200,method="deterministic")
dat.bkg[6,]<-bkg.norm(xpr=dat.bkg[6,],empty=200,method="deterministic")
dat.bkg[7,]<-bkg.norm(xpr=dat.bkg[7,],empty=200,method="deterministic")
dat.bkg[8,]<-bkg.norm(xpr=dat.bkg[8,],empty=200,method="deterministic")
dat.bkg[9,]<-bkg.norm(xpr=dat.bkg[9,],empty=200,method="deterministic")
dat.bkg[10,]<-bkg.norm(xpr=dat.bkg[10,],empty=200,method="deterministic")
dat.bkg[11,]<-bkg.norm(xpr=dat.bkg[11,],empty=200,method="deterministic")
dat.bkg[12,]<-bkg.norm(xpr=dat.bkg[12,],empty=200,method="deterministic")
dat.bkg[13,]<-bkg.norm(xpr=dat.bkg[13,],empty=200,method="deterministic")
dat.bkg[14,]<-bkg.norm(xpr=dat.bkg[14,],empty=200,method="deterministic")
dat.bkg[15,]<-bkg.norm(xpr=dat.bkg[15,],empty=200,method="deterministic")

library(affy)
dat.cond<-normalize.quantiles(t(dat.bkg),copy=TRUE)
norm.eset<-new("exprSet", exprs=dat.cond , phenoData=pdata)
dat.cond<-exprs(norm.eset)

##Create a table of normalised data before averaging.
```

```
dat.cond.matrix<-data.frame(names=genes$ID,
DL1=dat.cond[,1],DL2=dat.cond[,2],DL3=dat.cond[,3],HL1=dat.cond[,4],HL2=dat.cond[,5],HL3=dat.cond[,6],DR1=dat.cond[,7
],DR2=dat.cond[,8],DR3=dat.cond[,9],
HR1=dat.cond[,10],HR2=dat.cond[,11],HR3=dat.cond[,12],Seed1=dat.cond[,13],Seed2=dat.cond[,14],Seed3=dat.cond[,15])
write.table(dat.cond.matrix,"dat.cond.matrix",sep="\t")

# Filter out all control genes
dat.cond.filt<-dat.cond[genes$Status!="lucidea",]
filt.ID<-(as.character(genes$ID)[genes$Status!="lucidea"])

# reorder data so that replicate spots are grouped together
# to facilitate averaging over spots
i <- order(filt.ID)
filt.ID<-filt.ID[i]
dat.cond.filt<-dat.cond.filt[i,]

# average over replicate spots for all 15 arrays
ave.dat.cond<-matrix(0,ncol=15,nrow=length(unlist(lapply(split(dat.cond.filt[,1],filt.ID),mean))))

for (j in 1:15)
{
ave.dat.cond[,j]<-unlist(lapply(split(dat.cond.filt[,j],filt.ID),mean))
}
ave.gene.IDs<-names(lapply(split(dat.cond.filt[,1],filt.ID),mean))

save(ave.dat.cond,file="C:\\Documents and Settings\\sally\\Desktop\\modified xerophyta gpr files\\datnormall.R")
```

A.2.5.

```
setwd("C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis")

library(Biobase)
library(limma)
library(smida)

#######Read csv data files
controls<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\controls.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
drought<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\drought.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
seed<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\seed silique.csv",header =
TRUE, sep = ",", quote="\"", dec=".")

########## Extract intensity values and put in 22810 x 215 matrix
genedata<-matrix(0,nrow(controls),ncol=87)
genedata[,1]<-controls[,"Signal_Shoots_0h_Rep1"]
genedata[,2]<-controls[,"Signal_Shoots_0h_Rep2"]
genedata[,3]<-controls[,"Signal_Roots_0h_Rep1"]
genedata[,4]<-controls[,"Signal_Roots_0h_Rep2"]
genedata[,5]<-controls[,"Signal_Shoots_0.25h_Rep1"]
genedata[,6]<-controls[,"Signal_Shoots_0.25h_Rep2"]
genedata[,7]<-controls[,"Signal_Roots_0.25h_Rep1"]
genedata[,8]<-controls[,"Signal_Roots_0.25h_Rep2"]
genedata[,9]<-controls[,"Signal_Shoots_0.5h_Rep1"]
genedata[,10]<-controls[,"Signal_Shoots_0.5h_Rep2"]
genedata[,11]<-controls[,"Signal_Roots_0.5h_Rep1"]
genedata[,12]<-controls[,"Signal_Roots_0.5h_Rep2"]
genedata[,13]<-controls[,"Signal_Shoots_1.0h_Rep1"]
genedata[,14]<-controls[,"Signal_Shoots_1.0h_Rep2"]
genedata[,15]<-controls[,"Signal_Roots_1.0h_Rep1"]
```

```
genedata[,16]<-controls[,"Signal_Roots_1.0h_Rep2"]
genedata[,17]<-controls[,"Signal_Shoots_3.0h_Rep1"]
genedata[,18]<-controls[,"Signal_Shoots_3.0h_Rep2"]
genedata[,19]<-controls[,"Signal_Roots_3.0h_Rep1"]
genedata[,20]<-controls[,"Signal_Roots_3.0h_Rep2"]
genedata[,21]<-controls[,"Signal_Shoots_4.0h_Rep1"]
genedata[,22]<-controls[,"Signal_Shoots_4.0h_Rep2"]
genedata[,23]<-controls[,"Signal_Roots_4.0h_Rep1"]
genedata[,24]<-controls[,"Signal_Roots_4.0h_Rep2"]
genedata[,25]<-controls[,"Signal_Shoots_6.0h_Rep1"]
genedata[,26]<-controls[,"Signal_Shoots_6.0h_Rep2"]
genedata[,27]<-controls[,"Signal_Roots_6.0h_Rep1"]
genedata[,28]<-controls[,"Signal_Roots_6.0h_Rep2"]
genedata[,29]<-controls[,"Signal_Shoots_12.0h_Rep1"]
genedata[,30]<-controls[,"Signal_Shoots_12.0h_Rep2"]
genedata[,31]<-controls[,"Signal_Roots_12.0h_Rep1"]
genedata[,32]<-controls[,"Signal_Roots_12.0h_Rep2"]
genedata[,33]<-controls[,"Signal_Shoots_24.0h_Rep1"]
genedata[,34]<-controls[,"Signal_Shoots_24.0h_Rep2"]
genedata[,35]<-controls[,"Signal_Roots_24.0h_Rep1"]
genedata[,36]<-controls[,"Signal_Roots_24.0h_Rep2"]

genedata[,37]<-drought[,"Signal_Shoots_0.25h_Rep1"]
genedata[,38]<-drought[,"Signal_Shoots_0.25h_Rep2"]
genedata[,39]<-drought[,"Signal_Roots_0.25h_Rep1"]
genedata[,40]<-drought[,"Signal_Roots_0.25h_Rep2"]
genedata[,41]<-drought[,"Signal_Shoots_0.5h_Rep1"]
genedata[,42]<-drought[,"Signal_Shoots_0.5h_Rep2"]
genedata[,43]<-drought[,"Signal_Roots_0.5h_Rep1"]
genedata[,44]<-drought[,"Signal_Roots_0.5h_Rep2"]
genedata[,45]<-drought[,"Signal_Shoots_1.0h_Rep1"]
genedata[,46]<-drought[,"Signal_Shoots_1.0h_Rep2"]
genedata[,47]<-drought[,"Signal_Roots_1.0h_Rep1"]
genedata[,48]<-drought[,"Signal_Roots_1.0h_Rep2"]
genedata[,49]<-drought[,"Signal_Shoots_3.0h_Rep1"]
genedata[,50]<-drought[,"Signal_Shoots_3.0h_Rep2"]
genedata[,51]<-drought[,"Signal_Roots_3.0h_Rep1"]
genedata[,52]<-drought[,"Signal_Roots_3.0h_Rep2"]
genedata[,53]<-drought[,"Signal_Shoots_6.0h_Rep1"]
genedata[,54]<-drought[,"Signal_Shoots_6.0h_Rep2"]
genedata[,55]<-drought[,"Signal_Roots_6.0h_Rep1"]
genedata[,56]<-drought[,"Signal_Roots_6.0h_Rep2"]
genedata[,57]<-drought[,"Signal_Shoots_12.0h_Rep1"]
genedata[,58]<-drought[,"Signal_Shoots_12.0h_Rep2"]
genedata[,59]<-drought[,"Signal_Roots_12.0h_Rep1"]
genedata[,60]<-drought[,"Signal_Roots_12.0h_Rep2"]
genedata[,61]<-drought[,"Signal_Shoots_24.0h_Rep1"]
genedata[,62]<-drought[,"Signal_Shoots_24.0h_Rep2"]
genedata[,63]<-drought[,"Signal_Roots_24.0h_Rep1"]
genedata[,64]<-drought[,"Signal_Roots_24.0h_Rep2"]

genedata[,65]<-seed[,"Signal_76_B"]
genedata[,66]<-seed[,"Signal_76_C"]
genedata[,67]<-seed[,"Signal_77_D"]
genedata[,68]<-seed[,"Signal_77_E"]
genedata[,69]<-seed[,"Signal_77_F"]
genedata[,70]<-seed[,"Signal_78_D"]
genedata[,71]<-seed[,"Signal_78_E"]
genedata[,72]<-seed[,"Signal_78_F"]
```

```
genedata[,73]<-seed[,"Signal_79_A"]
genedata[,74]<-seed[,"Signal_79_B"]
genedata[,75]<-seed[,"Signal_79_C"]
genedata[,76]<-seed[,"Signal_81_A"]
genedata[,77]<-seed[,"Signal_81_B"]
genedata[,78]<-seed[,"Signal_81_C"]
genedata[,79]<-seed[,"Signal_82_A"]
genedata[,80]<-seed[,"Signal_82_B"]
genedata[,81]<-seed[,"Signal_82_C"]
genedata[,82]<-seed[,"Signal_83_A"]
genedata[,83]<-seed[,"Signal_83_B"]
genedata[,84]<-seed[,"Signal_83_C"]
genedata[,85]<-seed[,"Signal_84_A"]
genedata[,86]<-seed[,"Signal_84_B"]
genedata[,87]<-seed[,"Signal_84_D"]

genedata<-log2(genedata)

############### Convert data into exprSet object
covdesc<- list("Condition")
names(covdesc) <-"Condition"
geneCov<-as.data.frame(c("controls_Shoots_0h_Rep1","controls_Shoots_0h_Rep2"
,"controls_Roots_0h_Rep1","controls_Roots_0h_Rep2","controls_Shoots_0.25h_Rep1",
"controls_Shoots_0.25h_Rep2","controls_Roots_0.25h_Rep1","controls_Roots_0.25h_Rep2",
"controls_Shoots_0.5h_Rep1","controls_Shoots_0.5h_Rep2","controls_Roots_0.5h_Rep1"
,"controls_Roots_0.5h_Rep2","controls_Shoots_1.0h_Rep1","controls_Shoots_1.0h_Rep2"
,"controls_Roots_1.0h_Rep1","controls_Roots_1.0h_Rep2","controls_Shoots_3.0h_Rep1"
,"controls_Shoots_3.0h_Rep2","controls_Roots_3.0h_Rep1","controls_Roots_3.0h_Rep2"
,"controls_Shoots_4.0h_Rep1","controls_Shoots_4.0h_Rep2","controls_Roots_4.0h_Rep1"
,"controls_Roots_4.0h_Rep2","controls_Shoots_6.0h_Rep1","controls_Shoots_6.0h_Rep2"
,"controls_Roots_6.0h_Rep1","controls_Roots_6.0h_Rep2","controls_Shoots_12.0h_Rep1"
,"controls_Shoots_12.0h_Rep2","controls_Roots_12.0h_Rep1","controls_Roots_12.0h_Rep2"
,"controls_Shoots_24.0h_Rep1","controls_Shoots_24.0h_Rep2","controls_Roots_24.0h_Rep1"
,"controls_Roots_24.0h_Rep2","drought_Shoots_0.25h_Rep1","drought_Shoots_0.25h_Rep2",
"drought_Roots_0.25h_Rep1","drought_Roots_0.25h_Rep2","drought_Shoots_0.5h_Rep1",
"drought_Shoots_0.5h_Rep2","drought_Roots_0.5h_Rep1","drought_Roots_0.5h_Rep2",
"drought_Shoots_1.0h_Rep1","drought_Shoots_1.0h_Rep2","drought_Roots_1.0h_Rep1",
"drought_Roots_1.0h_Rep2","drought_Shoots_3.0h_Rep1","drought_Shoots_3.0h_Rep2",
"drought_Roots_3.0h_Rep1","drought_Roots_3.0h_Rep2","drought_Shoots_6.0h_Rep1",
"drought_Shoots_6.0h_Rep2","drought_Roots_6.0h_Rep1","drought_Roots_6.0h_Rep2",
"drought_Shoots_12.0h_Rep1","drought_Shoots_12.0h_Rep2","drought_Roots_12.0h_Rep1",
"drought_Roots_12.0h_Rep2","drought_Shoots_24.0h_Rep1","drought_Shoots_24.0h_Rep2",
"drought_Roots_24.0h_Rep1","drought_Roots_24.0h_Rep2","seed_76_B","seed_76_C"
,"seed_77_D","seed_77_E","seed_77_F","seed_78_D","seed_78_E","seed_78_F","seed_79_A","seed_79_B"
,"seed_79_C","seed_81_A","seed_81_B","seed_81_C","seed_82_A","seed_82_B","seed_82_C","seed_83_A"
,"seed_83_B","seed_83_C","seed_84_A","seed_84_B","seed_84_D"))

pdata <- new("phenoData", pData=geneCov, varLabels=covdesc)
eset <- new("exprSet", exprs=genedata, phenoData=pdata)

## Write expression matrix to file (tab-delimited text).
## I add the column names before writing it.

ExpData<-exprs(eset)
colnames(ExpData)<-pData(eset)[,1]
write.table(ExpData,file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\ExpDat.txt",sep="\t",row.names=FALSE)

### Get IDs and layout info from grp file
genes <- data.frame(ID=controls$GENENAME,Name=controls$DESCRIPTION)
```

```
# set control status
spottypes<-readSpotTypes()
genes$Status<-controlStatus(spottypes,genes=genes)

##Normalise data
dat<-exprs(eset)
dim(dat)

library(affy)
dat.cond<-normalize.quantiles(dat,copy=TRUE)
norm.eset<-new("exprSet", exprs=dat.cond , phenoData=pdata)
dat.cond<-exprs(norm.eset)

## Filter out all control genes
dat.cond.filt<-dat.cond[genes$Status=="gene",]
filt.ID<-(as.character(genes$ID)[genes$Status=="gene"])

#save(dat.cond.filt,file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\datnorm.R")
#dump("filt.ID",file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\IDnorm.txt")


# reorder data so that replicate spots are grouped together
# to facilitate averaging over spots
i <- order(filt.ID)
filt.ID<-filt.ID[i]
dat.cond.filt<-dat.cond.filt[i,]

# average over replicate spots for all 215 conditions
ave.dat.cond<-matrix(0,ncol=87,nrow=length(unlist(lapply(split(dat.cond.filt[,1],filt.ID),mean))))

for (j in 1:87)
{
ave.dat.cond[,j]<-unlist(lapply(split(dat.cond.filt[,j],filt.ID),mean))
}
ave.gene.IDs<-names(lapply(split(dat.cond.filt[,1],filt.ID),mean))

save(ave.dat.cond,file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\datnorm.R")
```

A.2.6.

```
setwd("C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis")

library(Biobase)
library(limma)
library(smida)

#######Read csv data files
controls<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\controls.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
osmotic<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\osmotic.csv",header =
TRUE, sep = ",", quote="\"", dec=".")
seed<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\modified files\\seed silique.csv",header =
TRUE, sep = ",", quote="\"", dec=".")

########## Extract intensity values and put in 22810 x 215 matrix
genedata<-matrix(0,nrow(controls),ncol=83)
genedata[,1]<-controls[,"Signal_Shoots_0h_Rep1"]
genedata[,2]<-controls[,"Signal_Shoots_0h_Rep2"]
genedata[,3]<-controls[,"Signal_Roots_0h_Rep1"]
genedata[,4]<-controls[,"Signal_Roots_0h_Rep2"]
```

```
genedata[,5]<-controls[,"Signal_Shoots_0.25h_Rep1"]
genedata[,6]<-controls[,"Signal_Shoots_0.25h_Rep2"]
genedata[,7]<-controls[,"Signal_Roots_0.25h_Rep1"]
genedata[,8]<-controls[,"Signal_Roots_0.25h_Rep2"]
genedata[,9]<-controls[,"Signal_Shoots_0.5h_Rep1"]
genedata[,10]<-controls[,"Signal_Shoots_0.5h_Rep2"]
genedata[,11]<-controls[,"Signal_Roots_0.5h_Rep1"]
genedata[,12]<-controls[,"Signal_Roots_0.5h_Rep2"]
genedata[,13]<-controls[,"Signal_Shoots_1.0h_Rep1"]
genedata[,14]<-controls[,"Signal_Shoots_1.0h_Rep2"]
genedata[,15]<-controls[,"Signal_Roots_1.0h_Rep1"]
genedata[,16]<-controls[,"Signal_Roots_1.0h_Rep2"]
genedata[,17]<-controls[,"Signal_Shoots_3.0h_Rep1"]
genedata[,18]<-controls[,"Signal_Shoots_3.0h_Rep2"]
genedata[,19]<-controls[,"Signal_Roots_3.0h_Rep1"]
genedata[,20]<-controls[,"Signal_Roots_3.0h_Rep2"]
genedata[,21]<-controls[,"Signal_Shoots_4.0h_Rep1"]
genedata[,22]<-controls[,"Signal_Shoots_4.0h_Rep2"]
genedata[,23]<-controls[,"Signal_Roots_4.0h_Rep1"]
genedata[,24]<-controls[,"Signal_Roots_4.0h_Rep2"]
genedata[,25]<-controls[,"Signal_Shoots_6.0h_Rep1"]
genedata[,26]<-controls[,"Signal_Shoots_6.0h_Rep2"]
genedata[,27]<-controls[,"Signal_Roots_6.0h_Rep1"]
genedata[,28]<-controls[,"Signal_Roots_6.0h_Rep2"]
genedata[,29]<-controls[,"Signal_Shoots_12.0h_Rep1"]
genedata[,30]<-controls[,"Signal_Shoots_12.0h_Rep2"]
genedata[,31]<-controls[,"Signal_Roots_12.0h_Rep1"]
genedata[,32]<-controls[,"Signal_Roots_12.0h_Rep2"]
genedata[,33]<-controls[,"Signal_Shoots_24.0h_Rep1"]
genedata[,34]<-controls[,"Signal_Shoots_24.0h_Rep2"]
genedata[,35]<-controls[,"Signal_Roots_24.0h_Rep1"]
genedata[,36]<-controls[,"Signal_Roots_24.0h_Rep2"]

genedata[,37]<-osmotic[,"Signal_Shoots_0.5h_Rep1"]
genedata[,38]<-osmotic[,"Signal_Shoots_0.5h_Rep2"]
genedata[,39]<-osmotic[,"Signal_Roots_0.5h_Rep1"]
genedata[,40]<-osmotic[,"Signal_Roots_0.5h_Rep2"]
genedata[,41]<-osmotic[,"Signal_Shoots_1.0h_Rep1"]
genedata[,42]<-osmotic[,"Signal_Shoots_1.0h_Rep2"]
genedata[,43]<-osmotic[,"Signal_Roots_1.0h_Rep1"]
genedata[,44]<-osmotic[,"Signal_Roots_1.0h_Rep2"]
genedata[,45]<-osmotic[,"Signal_Shoots_3.0h_Rep1"]
genedata[,46]<-osmotic[,"Signal_Shoots_3.0h_Rep2"]
genedata[,47]<-osmotic[,"Signal_Roots_3.0h_Rep1"]
genedata[,48]<-osmotic[,"Signal_Roots_3.0h_Rep2"]
genedata[,49]<-osmotic[,"Signal_Shoots_6.0h_Rep1"]
genedata[,50]<-osmotic[,"Signal_Shoots_6.0h_Rep2"]
genedata[,51]<-osmotic[,"Signal_Roots_6.0h_Rep1"]
genedata[,52]<-osmotic[,"Signal_Roots_6.0h_Rep2"]
genedata[,53]<-osmotic[,"Signal_Shoots_12.0h_Rep1"]
genedata[,54]<-osmotic[,"Signal_Shoots_12.0h_Rep2"]
genedata[,55]<-osmotic[,"Signal_Roots_12.0h_Rep1"]
genedata[,56]<-osmotic[,"Signal_Roots_12.0h_Rep2"]
genedata[,57]<-osmotic[,"Signal_Shoots_24.0h_Rep1"]
genedata[,58]<-osmotic[,"Signal_Shoots_24.0h_Rep2"]
genedata[,59]<-osmotic[,"Signal_Roots_24.0h_Rep1"]
genedata[,60]<-osmotic[,"Signal_Roots_24.0h_Rep2"]

genedata[,61]<-seed[,"Signal_76_B"]
```

```
genedata[,62]<-seed[,"Signal_76_C"]
genedata[,63]<-seed[,"Signal_77_D"]
genedata[,64]<-seed[,"Signal_77_E"]
genedata[,65]<-seed[,"Signal_77_F"]
genedata[,66]<-seed[,"Signal_78_D"]
genedata[,67]<-seed[,"Signal_78_E"]
genedata[,68]<-seed[,"Signal_78_F"]
genedata[,69]<-seed[,"Signal_79_A"]
genedata[,70]<-seed[,"Signal_79_B"]
genedata[,71]<-seed[,"Signal_79_C"]
genedata[,72]<-seed[,"Signal_81_A"]
genedata[,73]<-seed[,"Signal_81_B"]
genedata[,74]<-seed[,"Signal_81_C"]
genedata[,75]<-seed[,"Signal_82_A"]
genedata[,76]<-seed[,"Signal_82_B"]
genedata[,77]<-seed[,"Signal_82_C"]
genedata[,78]<-seed[,"Signal_83_A"]
genedata[,79]<-seed[,"Signal_83_B"]
genedata[,80]<-seed[,"Signal_83_C"]
genedata[,81]<-seed[,"Signal_84_A"]
genedata[,82]<-seed[,"Signal_84_B"]
genedata[,83]<-seed[,"Signal_84_D"]

genedata<-log2(genedata)

############### Convert data into exprSet object
covdesc<- list("Condition")
names(covdesc) <-"Condition"
geneCov<-as.data.frame(c("controls_Shoots_0h_Rep1","controls_Shoots_0h_Rep2"
,"controls_Roots_0h_Rep1","controls_Roots_0h_Rep2","controls_Shoots_0.25h_Rep1",
"controls_Shoots_0.25h_Rep2","controls_Roots_0.25h_Rep1","controls_Roots_0.25h_Rep2",
"controls_Shoots_0.5h_Rep1","controls_Shoots_0.5h_Rep2","controls_Roots_0.5h_Rep1"
,"controls_Roots_0.5h_Rep2","controls_Shoots_1.0h_Rep1","controls_Shoots_1.0h_Rep2"
,"controls_Roots_1.0h_Rep1","controls_Roots_1.0h_Rep2","controls_Shoots_3.0h_Rep1"
,"controls_Shoots_3.0h_Rep2","controls_Roots_3.0h_Rep1","controls_Roots_3.0h_Rep2"
,"controls_Shoots_4.0h_Rep1","controls_Shoots_4.0h_Rep2","controls_Roots_4.0h_Rep1"
,"controls_Roots_4.0h_Rep2","controls_Shoots_6.0h_Rep1","controls_Shoots_6.0h_Rep2"
,"controls_Roots_6.0h_Rep1","controls_Roots_6.0h_Rep2","controls_Shoots_12.0h_Rep1"
,"controls_Shoots_12.0h_Rep2","controls_Roots_12.0h_Rep1","controls_Roots_12.0h_Rep2"
,"controls_Shoots_24.0h_Rep1","controls_Shoots_24.0h_Rep2","controls_Roots_24.0h_Rep1"
,"controls_Roots_24.0h_Rep2","osmotic_Shoots_0.5h_Rep1",
"osmotic_Shoots_0.5h_Rep2","osmotic_Roots_0.5h_Rep1","osmotic_Roots_0.5h_Rep2",
"osmotic_Shoots_1.0h_Rep1","osmotic_Shoots_1.0h_Rep2","osmotic_Roots_1.0h_Rep1",
"osmotic_Roots_1.0h_Rep2","osmotic_Shoots_3.0h_Rep1","osmotic_Shoots_3.0h_Rep2",
"osmotic_Roots_3.0h_Rep1","osmotic_Roots_3.0h_Rep2","osmotic_Shoots_6.0h_Rep1",
"osmotic_Shoots_6.0h_Rep2","osmotic_Roots_6.0h_Rep1","osmotic_Roots_6.0h_Rep2",
"osmotic_Shoots_12.0h_Rep1","osmotic_Shoots_12.0h_Rep2","osmotic_Roots_12.0h_Rep1",
"osmotic_Roots_12.0h_Rep2","osmotic_Shoots_24.0h_Rep1","osmotic_Shoots_24.0h_Rep2",
"osmotic_Roots_24.0h_Rep1","osmotic_Roots_24.0h_Rep2","seed_76_B","seed_76_C"
,"seed_77_D","seed_77_E","seed_77_F","seed_78_D","seed_78_E","seed_78_F","seed_79_A","seed_79_B"
,"seed_79_C","seed_81_A","seed_81_B","seed_81_C","seed_82_A","seed_82_B","seed_82_C","seed_83_A"
,"seed_83_B","seed_83_C","seed_84_A","seed_84_B","seed_84_D"))

pdata <- new("phenoData", pData=geneCov, varLabels=covdesc)
eset <- new("exprSet", exprs=genedata, phenoData=pdata)

## Write expression matrix to file (tab-delimited text).
## I add the column names before writing it.
```

```
ExpData<-exprs(eset)
colnames(ExpData)<-pData(eset)[,1]
write.table(ExpData,file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\ExpDat.txt",sep="\t",row.names=FALSE)

### Get IDs and layout info from grp file
genes <- data.frame(ID=controls$GENENAME,Name=controls$DESCRIPTION)
# set control status
spottypes<-readSpotTypes()
genes$Status<-controlStatus(spottypes,genes=genes)

##Normalise data
dat<-exprs(eset)
dim(dat)

library(affy)
dat.cond<-normalize.quantiles(dat,copy=TRUE)
norm.eset<-new("exprSet", exprs=dat.cond , phenoData=pdata)
dat.cond<-exprs(norm.eset)

## Filter out all control genes
dat.cond.filt<-dat.cond[genes$Status=="gene",]
filt.ID<-(as.character(genes$ID)[genes$Status=="gene"])

#save(dat.cond.filt,file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\datnorm.R")
#dump("filt.ID",file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\IDnorm.txt")


# reorder data so that replicate spots are grouped together
# to facilitate averaging over spots
i <- order(filt.ID)
filt.ID<-filt.ID[i]
dat.cond.filt<-dat.cond.filt[i,]

# average over replicate spots for all 83 conditions
ave.dat.cond<-matrix(0,ncol=83,nrow=length(unlist(lapply(split(dat.cond.filt[,1],filt.ID),mean))))

for (j in 1:83)
{
ave.dat.cond[,j]<-unlist(lapply(split(dat.cond.filt[,j],filt.ID),mean))
}
ave.gene.IDs<-names(lapply(split(dat.cond.filt[,1],filt.ID),mean))

save(ave.dat.cond,file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\datnorm.R")
```

A.3.1.

```
###Create a exprSet of subset data with new subset column names
covdesc<- list("Condition")
names(covdesc) <-"Condition"
geneCov<-as.data.frame(c("OsL1","OsL2","CL1","CL2","OsR1","OsR2","CR1","CR2","seed1","seed2","seed3"))
OR
as.data.frame(c("DsL1","DsL2","DsL3","HL1","HL2","HL3","DsR1","DsR2","DsR3","HR1","HR2","HR3","seed1","seed2","se
ed3"))

pdata <- new("phenoData", pData=geneCov, varLabels=covdesc)
subset.eset <- new("exprSet", exprs=subset.data, phenoData=pdata)

####add the subset expression data to the exprSet
###Read this table out to correct the -Inf values
ExpData.subset<-exprs(subset.eset)
```

```
colnames(ExpData.subset)<-pData(subset.eset)[,1]


###CLUSTERING GENES
## Linear model fits
design<-model.matrix(~-1+factor(c(1,1,2,2,3,3,4,4,5,5,5)))
colnames(design)<-c("OsL","CL","OsR","CR","seed")
OR colnames(design)<-c("DsL","HL","DsR","HR","seed")

fit<-lmFit(ExpData.subset,design)
contrast.matrix<-makeContrasts("OsL-CL","OsR-CR","seed-HL","seed-CR",levels=design)
OR contrast.matrix<-makeContrasts("OsL-CL","OsR-CR",levels=design)
OR contrast.matrix<-makeContrasts("DsL-HL","DsR-HR","seed-HL","seed-HR",levels=design)
OR contrast.matrix<-makeContrasts("DsL-HL","DsR-HR",levels=design)
fit2<-contrasts.fit(fit,contrast.matrix)
fit2<-eBayes(fit2)
results<-decideTests(fit2)
diffex<-apply(abs(results),1,sum)
diffex<-diffex>0

cluster.data<-ExpData.subset[diffex,]
# extracts the gene names for above data
cluster.names<-ave.gene.IDs[diffex]
rownames(cluster.data)<-cluster.names
unique(rownames(cluster.data))
dim(cluster.data)

show(cluster.data)
cluster.norm.eset<-new("exprSet", exprs=cluster.data, phenoData=pdata)
cluster.data<-exprs(cluster.norm.eset)

clustering<-exprs(cluster.norm.eset)
colnames(clustering)<-pData(cluster.norm.eset)[,1]

###AVERAGE THE BIOLOGICAL REPLICATES

write.table(cluster.data,"cluster.data.txt",sep="\t")
##average the replicates in excel and read back into R and create new exprs object (Remove gene names and arrive 3 replicates in
new spreadsheet)
ave.cluster.data<-read.table(file="C:\\Documents and Settings\\sally\\Desktop\\Arabidopsis\\ave.cluster.data.txt",sep="\t")

ave.cluster.data<-as.matrix(ave.cluster.data)

covdesc<- list("Condition")
names(covdesc) <-"Condition"
geneCov<-as.data.frame(c("OsL","CL","OsR","CR","Seed"))
OR geneCov<-as.data.frame(c("DsL","CL","DsR","CR","Seed"))

pdata <- new("phenoData", pData=geneCov, varLabels=covdesc)
eset <- new("exprSet", exprs=ave.cluster.data, phenoData=pdata)

## Write expression matrix to file (tab-delimited text).
## I add the column names before writing it.

AveClusterData<-exprs(eset)
colnames(AveClusterData)<-pData(eset)[,1]
#write.table(AveClusterData,file="C:\\Documents and
Settings\\sally\\Desktop\\Arabidopsis\\AveClusterData.txt",sep="\t",row.names=FALSE)
```

```
rownames(AveClusterData)<-cluster.names
unique(rownames(AveClusterData))

###change between cluster.data and ave.cluster.data for gene clustering
###and between clustering and AveClusterData for conditional clustering

###CLUSTERING GENES
library(mclust)
mclust<-Mclust(AveClusterData)
plot.Mclust(mclust,AveClusterData)

x11()
pamsam<-pamsam(AveClusterData,k=8,metric="correlation")

#CLUSTERS OF INTEREST
#extract cluster information
#cluster1
clust1names<-cluster.names[pamsam$clustering==1]
clust1DL.exp<-as.vector(AveClusterData[,1][pamsam$clustering==1])
clust1HL.exp<-as.vector(AveClusterData[,2][pamsam$clustering==1])
clust1DR.exp<-as.vector(AveClusterData[,3][pamsam$clustering==1])
clust1HR.exp<-as.vector(AveClusterData[,4][pamsam$clustering==1])
clust1seed.exp<-as.vector(AveClusterData[,5][pamsam$clustering==1])
cluster1<-
data.frame(names=clust1names,DL.exp=clust1DL.exp,HL.exp=clust1HL.exp,DR.exp=clust1DR.exp,HR.exp=clust1HR.exp,seed
.exp=clust1seed.exp)
write.table(cluster1,"cluster1",sep="\t")

#cluster2
clust2names<-cluster.names[pamsam$clustering==2]
clust2DL.exp<-as.vector(AveClusterData[,1][pamsam$clustering==2])
clust2HL.exp<-as.vector(AveClusterData[,2][pamsam$clustering==2])
clust2DR.exp<-as.vector(AveClusterData[,3][pamsam$clustering==2])
clust2HR.exp<-as.vector(AveClusterData[,4][pamsam$clustering==2])
clust2seed.exp<-as.vector(AveClusterData[,5][pamsam$clustering==2])
cluster2<-
data.frame(names=clust2names,DL.exp=clust2DL.exp,HL.exp=clust2HL.exp,DR.exp=clust2DR.exp,HR.exp=clust2HR.exp,seed
.exp=clust2seed.exp)
write.table(cluster2,"cluster2",sep="\t")

#cluster3
clust3names<-cluster.names[pamsam$clustering==3]
clust3DL.exp<-as.vector(AveClusterData[,1][pamsam$clustering==3])
clust3HL.exp<-as.vector(AveClusterData[,2][pamsam$clustering==3])
clust3DR.exp<-as.vector(AveClusterData[,3][pamsam$clustering==3])
clust3HR.exp<-as.vector(AveClusterData[,4][pamsam$clustering==3])
clust3seed.exp<-as.vector(AveClusterData[,5][pamsam$clustering==3])
cluster3<-
data.frame(names=clust3names,DL.exp=clust3DL.exp,HL.exp=clust3HL.exp,DR.exp=clust3DR.exp,HR.exp=clust3HR.exp,seed
.exp=clust3seed.exp)
write.table(cluster3,"cluster3",sep="\t")

#cluster4
clust4names<-cluster.names[pamsam$clustering==4]
clust4DL.exp<-as.vector(AveClusterData[,1][pamsam$clustering==4])
clust4HL.exp<-as.vector(AveClusterData[,2][pamsam$clustering==4])
clust4DR.exp<-as.vector(AveClusterData[,3][pamsam$clustering==4])
clust4HR.exp<-as.vector(AveClusterData[,4][pamsam$clustering==4])
clust4seed.exp<-as.vector(AveClusterData[,5][pamsam$clustering==4])
```

```
cluster4<-
data.frame(names=clust4names,DL.exp=clust4DL.exp,HL.exp=clust4HL.exp,DR.exp=clust4DR.exp,HR.exp=clust4HR.exp,seed
.exp=clust4seed.exp)
write.table(cluster4,"cluster4",sep="\t")

#cluster5
clust5names<-cluster.names[pamsam$clustering==5]
clust5DL.exp<-as.vector(AveClusterData[,1][pamsam$clustering==5])
clust5HL.exp<-as.vector(AveClusterData[,2][pamsam$clustering==5])
clust5DR.exp<-as.vector(AveClusterData[,3][pamsam$clustering==5])
clust5HR.exp<-as.vector(AveClusterData[,4][pamsam$clustering==5])
clust5seed.exp<-as.vector(AveClusterData[,5][pamsam$clustering==5])
cluster5<-
data.frame(names=clust5names,DL.exp=clust5DL.exp,HL.exp=clust5HL.exp,DR.exp=clust5DR.exp,HR.exp=clust5HR.exp,seed
.exp=clust5seed.exp)
write.table(cluster5,"cluster5",sep="\t")

#cluster6
clust6names<-cluster.names[pamsam$clustering==6]
clust6DL.exp<-as.vector(AveClusterData[,1][pamsam$clustering==6])
clust6HL.exp<-as.vector(AveClusterData[,2][pamsam$clustering==6])
clust6DR.exp<-as.vector(AveClusterData[,3][pamsam$clustering==6])
clust6HR.exp<-as.vector(AveClusterData[,4][pamsam$clustering==6])
clust6seed.exp<-as.vector(AveClusterData[,5][pamsam$clustering==6])
cluster6<-
data.frame(names=clust6names,DL.exp=clust6DL.exp,HL.exp=clust6HL.exp,DR.exp=clust6DR.exp,HR.exp=clust6HR.exp,seed
.exp=clust6seed.exp)
write.table(cluster6,"cluster6",sep="\t")

#cluster7
clust7names<-cluster.names[pamsam$clustering==7]
clust7DL.exp<-as.vector(AveClusterData[,1][pamsam$clustering==7])
clust7HL.exp<-as.vector(AveClusterData[,2][pamsam$clustering==7])
clust7DR.exp<-as.vector(AveClusterData[,3][pamsam$clustering==7])
clust7HR.exp<-as.vector(AveClusterData[,4][pamsam$clustering==7])
clust7seed.exp<-as.vector(AveClusterData[,5][pamsam$clustering==7])
cluster7<-
data.frame(names=clust7names,DL.exp=clust7DL.exp,HL.exp=clust7HL.exp,DR.exp=clust7DR.exp,HR.exp=clust7HR.exp,seed
.exp=clust7seed.exp)
write.table(cluster7,"cluster7",sep="\t")

CLUSTERING CONDITIONS
sample.cluster<-t(AveClusterData)

#pca of samples
x11()
pca<-cluster.samples(sample.cluster,method="pca")

#diana clustering
library(cluster)
diana<-diana(sample.cluster,metric="correlation")
plot(diana)
```

Table A.3.1: Annotation of genes differentially expressed between desiccated vegetative tissue and seed in *X. humilis*. The *X. humilis* Contig number and At locus tag, annotated gene name and Best match e-value of the corresponding *A. thaliana* homologue are presented.

Contig ID	***At locus tag***	***Annotated gene name***	***Best match E- value***
Contig190	At1g01440	extra-large G-protein-related	5.00E-09
Contig865	At1g01470	LEA8	2.00E-44
Contig587	At1g02130	Ras-related protein ARA-5	1.00E-86
Contig657	At1g02700	unknown protein	2.00E-16
Contig583	At1g02816	unknown protein	2.00E-39
Contig807	At1g02890	AAA-type ATPase family protein	9.00E-133
Contig484	At1g02940	Glutathione-S-transferase	2.00E-05
Contig766	At1g04560	LEA10	3.00E-52
Contig915	At1g04560	LEA10	8.00E-53
Contig44	At1g04760	XhVAMP726 (VESICLE-ASSOCIATED MEMBRANE PROTEIN)	4.00E-85
Contig361	At1g06040	STO (SALT TOLERANCE); transcription factor/ zinc ion binding	3.00E-68
Contig580	At1g07645	glyoxylase-I like	9.00E-48
Contig871	At1g07890	APX1 (ASCORBATE PEROXIDASE 1, MATERNAL EFFECT EMBRYO ARREST 6); L-ascorbate peroxidase	2.00E-86
Contig473	At1g08380	PSAO (photosystem I subunit O)	6.00E-45
Contig409	At1g08830	CSD1 (copper/zinc superoxide dismutase 1)	3.00E-55
Contig615	At1g09590	60S ribosomal protein L21 (RPL21A)	1.00E-75
Contig834	At1g09960	Sucrose transporter	2.00E-78
Contig189	At1g10500	XhCPISCA (CHLOROPLAST-LOCALIZED ISCA-LIKE PROTEIN, Iron-sulfur assembly protein IscA	2.00E-45
Contig1460	At1g11740	ankyrin repeat family protein	5.00E-03
Contig196	At1g11750	CLPP6 (Clp protease proteolytic subunit 6); endopeptidase Clp	1.00E-99
Contig1108	At1g11840	XhGLX1, glyoxalase I	2.00E-11
Contig769	At1g11910	aspartic proteinase	2.00E-129
Contig1088	At1g13060	PBE1 (20S proteasome beta subunit E1); peptidase	3.00E-50
Contig788	At1g14270	CAAX amino terminal protease family protein	1.00E-74
Contig122	At1g14290	acid phosphatase, putative	9.00E-99
Contig711	At1g14340	RNA-binding region RNP-1 (RNA recognition motif) domain containing protein	6.00E-45
Contig145	At1g15690	vacuolar-type H+-pumping pyrophosphatase, ATPase	4.00E-140
Contig86	At1g15710	Prephenate dehydrogenase domain containing protein	2.00E-37
Contig255	At1g16180	TMS membrane family protein / tumour differentially expressed (TDE) family protein	4.00E-119
Contig307	At1g17200	DUF588 integral membrane family protein	4.00E-42
Contig1461	At1g18080	RACK1A Receptor for Activated C Kinase 1	4.00E-62
Contig223	At1g18335	GCN5-related N-acetyltransferase (GNAT) family protein	6.00E-46
Contig813	At1g18880	Nitrate transporter NRT1-5 (Fragment)	2.00E-102
Contig1134	At1g19530	Unknown Protein	3.00E-07
Contig1523	At1g20020	NADPH dehydrogenase/ oxidoreductase (XhLFNR)	5.00E-89
Contig946	At1g20030	pathogenesis-related thaumatin family protein	3.00E-78
Contig423	At1g20696	HMGB3 (HIGH MOBILITY GROUP B 3), transcription factor	4.00E-19

Contig680	At1g22400	UDP-glucuronosyl/UDP-glucosyltransferase family protein	2.00E-54
Contig831	At1g22600	unknown protein	2.00E-17
Contig178	At1g23260	MMZ1 (MMS ZWEI HOMOLOGE 1); ubiquitin-protein ligase	5.00E-61
Contig224	At1g23740	oxidoreductase, zinc-binding dehydrogenase family protein	8.00E-51
Contig415	At1g25350	glutamine-tRNA ligase	8.00E-164
Contig781	At1g26800	Zn-finger, RING domain containing protein	5.00E-25
Contig1301	At1g26910	60S ribosomal protein L10 (RPL10B)	2.00E-25
Contig1067	At1g27990	unknown protein	7.00E-65
Contig1376	At1g28100	unknown protein	3.00E-32
Contig372	At1g28330	DRM1 (DORMANCY-ASSOCIATED PROTEIN 1), auxin-repressed protein	1.00E-31
Contig490	At1g29930	Chlorophyll a-b binding protein 2, chloroplast precursor (LHCII type I CAB-2) (CAB-140) (LHCP)	7.00E-112
Contig201	At1g29930	Chlorophyll a-b binding protein 2, chloroplast precursor (LHCII type I CAB-2) (CAB-140) (LHCP)	9.00E-126
Contig673	At1g29970	60S ribosomal protein L18a-1	7.00E-85
Contig367	At1g29990	prefoldin	5.00E-35
Contig498	At1g31330	PSAF (photosystem I subunit F)	2.00E-85
Contig1050	At1g32230	radical induced cell death	3.00E-24
Contig194	At1g32790	CID11; RNA binding / protein binding	7.00E-70
Contig1536	At1g36050	unknown protein	2.00E-54
Contig786	At1g36070	WD domain protein-like	2.00E-106
Contig923	At1g36070	WD domain protein-like	3.00E-109
Contig1373	At1g43190	polypyrimidine tract-binding protein, putative / heterogeneous nuclear ribonucleoprotein, putative, RNA-binding	5.00E-45
Contig586	At1g44910	protein binding	2.00E-24
Contig595	At1g47980	Desiccation related protein	3.00E-98
Contig628	At1g47980	Desiccation related protein	5.00E-100
Contig1306	At1g48130	1-cysteine peroxiredoxin; antioxidant	2.00E-05
Contig1064	At1g48130	1-cysteine peroxiredoxin; antioxidant	4.00E-78
Contig527	At1g48130	1-cysteine peroxiredoxin; antioxidant	5.00E-86
Contig1082	At1g48350	ribosomal protein L18 family protein ; 50S ribosomal protein L18, chloroplast precursor (CL18)	7.00E-24
Contig362	At1g49950	XhTRB1/TRB1 (TELOMERE REPEAT BINDING FACTOR 1); MYB transcription factor	1.00E-56
Contig616	At1g51200	zinc finger (AN1-like) family protein	3.00E-42
Contig321	At1g52230	XhPSA-H, photosystem I reaction center subunit VI, chloroplast, putative / PSI-H, putative (PSAH2)	3.00E-48
Contig864	At1g52690	LEA3	4.00E-21
Contig900	At1g53540	LMW heat shock protein	1.00E-48
Contig710	At1g53540	LMW heat shock protein	6.00E-54
Contig424	At1g54290	eukaryotic translation initiation factor SUI1, putative	5.00E-49
Contig88	At1g54290	Protein translation factor SUI1 homolog (GOS2 protein)	2.00E-45
Contig479	At1g54350	ABC transporter family protein	6.00E-11
Contig690	At1g54870	Glucose and ribitol dehydrogenase	1.00E-113
Contig1131	At1g55530	zinc finger (C3HC4-type RING finger) family protein	7.00E-41
Contig358	At1g55670	Photosystem I reaction center subunit V; PSI-G/psaG	4.00E-50

Contig585	At1g55960	unknown protein	2.00E-07
Contig213	At1g56420	unknown protein	3.00E-36
Contig916	At1g56600	Galactinol synthase	9.00E-146
Contig496	At1g59610	ADL3 (ARABIDOPSIS DYNAMIN-LIKE 3)	2.00E-07
Contig627	At1g60200	splicing factor PWI domain-containing protein / RNA recognition motif (RRM)-containing protein	1.00E-36
Contig656	At1g60470	galactinol synthase, isoform GolS-1	4.00E-146
Contig31	At1g61070	protease inhibitor	4.00E-10
Contig557	At1g61260	unknown protein	5.00E-32
Contig1233	At1g61850	XhPLAI, similar to putative calcium-independent phospholipase A2	2.00E-70
Contig830	At1g64720	membrane related protein	2.00E-47
Contig1124	At1g65980	PEROXIREDOXIN TPX1; THIOREDOXIN-DEPENDENT PEROXIDASE 1, antioxidant	3.00E-66
Contig499	At1g66240	XhATX1,copper chaperone	7.00E-09
Contig1184	At1g67090	Ribulose bisphosphate carboxylase small chain A	2.00E-23
Contig131	At1g67090	Ribulose bisphosphate carboxylase small chain A	3.00E-72
Contig348	At1g67740	XhpsbY, Photosystem II core complex proteins psbY, chloroplast precursor	3.00E-26
Contig836	At1g68490	unknown protein	2.00E-06
Contig141	At1g72020	XhSLL1 protein	2.00E-21
Contig624	At1g72100	LEA4	4.00E-03
Contig827	At1g72640	unknown protein	3.00E-31
Contig1566	At1g74040	IMS1; 2-isopropylmalate synthase	1.00E-48
Contig1070	At1g76650	calcium-binding EF hand family protein; putative regulator of gene silencing	3.00E-21
Contig661	At1g76650	calcium-binding EF hand family protein; putative regulator of gene silencing	6.00E-25
Contig483	At1g77330	1-aminocyclopropane-1-carboxylate oxidase, putative / ACC oxidase, putative	2.00E-114
Contig1519	At1g77510	protein disulfide isomerase (XhPDIL1-1)	3.00E-50
Contig1519	At1g77510	protein disulfide isomerase (XhPDIL1-1)	6.00E-52
Contig468	At1g78060	putative beta-xylosidase; glycosyl hydrolase family 3 protein	4.00E-39
Contig1036	At1g78380	Glutathione transferase	4.00E-37
Contig157	At1g79110	protein binding / zinc ion binding	5.00E-10
Contig1309	At1g80090	SNF4 protein, CBS domain-containing protein,	3.00E-10
Contig1104	At1g80160	glyoxalase-I like protein	1.00E-54
Contig913	At1g80160	glyoxalase-I like protein	3.00E-62
Contig1162	At1g80660	H(+)-transporting ATPase	2.00E-56
Contig1383	At2g01320	ABC transporter family protein	2.00E-60
Contig1051	At2g02120	protease inhibitor	3.00E-03
Contig536	At2g02120	protease inhibitor	2.00E-06
Contig17	At2g02760	UBC1or2 ubiquitin-protein ligase	6.00E-78
Contig1458	At2g03350	unknown protein	2.00E-58
Contig904	At2g04240	Zn-finger, RING domain containing protein	9.00E-28
Contig1326	At2g05710	Aconitate hydratase, cytoplasmic, putative, expressed	2.00E-65
Contig867	At2g05840	PAA2 (20S proteasome alpha subunit A2); peptidase	7.00E-123

Contig1386	At2g07050	XhCAS1 (CYCLOARTENOL SYNTHASE 1)	5.00E-71
Contig1010	At2g11890	adenylate cyclase	5.00E-09
Contig34	At2g13360	AGT (ALANINE:GLYOXYLATE AMINOTRANSFERASE);serine glyoxylate aminotransferase	7.00E-97
Contig426	At2g16400	BELL1-LIKE HOMEODOMAIN 7	8.00E-75
Contig216	At2g16500	XhADC1 (ARGININE DECARBOXYLASE 1)	3.00E-95
Contig952	At2g16600	ROC3 (rotamase CyP 3); peptidyl-prolyl cis-trans isomerase	1.00E-78
Contig493	At2g16850	aquaporin; PIP2 (plasma membrane intrinsic protein 2;8); water channel	4.00E-30
Contig1152	At2g17500	auxin efflux	5.00E-89
Contig1299	At2g17840	ERD7 (Early responsive to dehydration)	7.00E-31
Contig606	At2g17840	ERD7 (Early responsive to dehydration)	2.00E-84
Contig1106	At2g21660	RNA binding	2.00E-31
Contig722	At2g21820	unknown protein	7.00E-14
Contig579	At2g22292	Glycine-rich RNA-binding protein 7	2.00E-28
Contig1287	At2g24090	putative chloroplast ribosomal protein L35	4.00E-13
Contig1535	At2g24520	plasma membrane H+-ATPase	5.00E-71
Contig761	At2g25110	MIR domain-containing protein	3.00E-87
Contig972	At2g25625	unknown protein	5.00E-07
Contig59	At2g26730	leucine-rich repeat transmembrane protein kinase, putative	1.00E-104
Contig724	At2g26900	bile acid:sodium symporter family protein	2.00E-153
Contig969	At2g27500	glycosyl hydrolase family 17 protein	7.00E-85
Contig411	At2g28740	HIS4 (Histone H4)	3.00E-33
Contig909	At2g30570	photosystem II reaction center W (PsbW) protein-related	2.00E-29
Contig715	At2g30950	VAR2 (VARIEGATED 2); ATP-dependent peptidase/ ATPase/ metallopeptidase/ zinc ion binding	3.00E-155
Contig597	At2g31400	pentatricopeptide (PPR) repeat-containing protein	2.00E-75
Contig1239	At2g33150	Acetyl-CoA C-acyltransferase (3-ketoacyl-coa thiolase b)	8.00E-81
Contig787	At2g33150	Acetyl-CoA C-acyltransferase (3-ketoacyl-coa thiolase b)	1.00E-95
Contig574	At2g33380	calcium binding, EF hand	2.00E-85
Contig221	At2g33390	unknown protein	2.00E-11
Contig1459	At2g35120	putative glycine decarboxylase complex H-protein	3.00E-55
Contig1199	At2g35410	33 kDa ribonucleoprotein, chloroplast, putative / RNA-binding protein cp33, putative	2.00E-29
Contig435	At2g35980	Harpin-induced 1 domain containing protein	4.00E-34
Contig101	At2g36060	MMZ3 (MMS ZWEI HOMOLOGE 3); ubiquitin-protein ligase	6.00E-74
Contig838	At2g36460	fructose-bisphosphate aldolase, putative	5.00E-98
Contig1061	At2g36680	unknown protein	3.00E-56
Contig1122	At2g36830	Tonoplast intrinsic protein 1.1;Gamma-tonoplast intrinsic proteinTonoplast intrinsic protein, root-specific RB7 (TIP1.1)	2.00E-60
Contig15	At2g37130	Xhperox P21; Peroxidase 21 precursor; putative peroxidase ATP2a	2.00E-50
Contig463	At2g37170	PIP2B (plasma membrane intrinsic protein 2;2); water channel	1.00E-24
Contig301	At2g37170	PIP2B (plasma membrane intrinsic protein 2;2); water channel	2.00E-56
Contig727	At2g37220	RNA-binding region RNP-1 (RNA recognition motif) domain containing protein	3.00E-67
Contig1328	At2g37650	scarecrow-like transcription factor 9 (SCL9)	1.00E-41

Contig438	At2g38090	myb family transcription factor	7.00E-69
Contig1292	At2g38470	putative WRKY transcription factor	1.00E-27
Contig87	At2g39780	RNS2 (RIBONUCLEASE 2); endoribonuclease	3.00E-76
Contig897	At2g39840	Protein phosphatase 1, catalytic gsmms subunit	2.00E-13
Contig441	At2g40000	putative nematode-resistance protein	2.00E-46
Contig984	At2g40010	60S acidic ribosomal protein P0	3.00E-106
Contig889	At2g40170	LEA1	4.00E-12
Contig857	At2g40590	40S ribosomal protein S26 (RPS26B)	3.00E-38
Contig592	At2g42560	LEA3	4.00E-04
Contig998	At2g43320	unknown protein	2.00E-89
Contig776	At2g43350	glutathione peroxidase	3.00E-69
Contig590	At2g43630	unknown protein	4.00E-11
Contig1000	At2g44750	thiamin pyrophosphokinase, putative	6.00E-13
Contig171	At2g45180	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein, localized in thylakoid membrane	5.00E-17
Contig443	At2g45290	transketolase, putative , chloroplast	2.00E-131
Contig955	At2g47180	Galactinol synthase	1.00E-43
Contig918	At2g47770	benzodiazepine receptor-related	2.00E-30
Contig1068	At2g48130	protease inhibitor	1.00E-04
Contig862	At3g01570	glycine-rich protein / oleosin /oleosin5	4.00E-33
Contig1059	At3g01570	glycine-rich protein / oleosin /oleosin5	5.00E-37
Contig1038	At3g02520	GRF7 (General regulatory factor 7)	2.00E-111
Contig1147	At3g02555	unknown protein	2.00E-07
Contig1099	At3g02560	40S ribosomal protein S7	2.00E-74
Contig1544	At3g02630	acyl-(acyl-carrier-protein) desaturase, putative / stearoyl-ACP desaturase, putative	8.00E-15
Contig885	At3g02720	protease-related	6.00E-109
Contig856	At3g03070	NADH-ubiquinone oxidoreductase-related	3.00E-33
Contig1270	At3g04720	XhPR4, pathogenesis-related 4, similar to basic chitinase	3.00E-16
Contig1074	At3g04870	Zeta-carotene desaturase (Fragment)	2.00E-56
Contig920	At3g05890	Hydrophobic protein RCI2A (Low temperature and salt responsive protein LTI6A)	8.00E-11
Contig68	At3g05910	pectinacetylesterase, putative	5.00E-91
Contig355	At3g07100	protein transport protein Sec24, putative Sec24-like COPII protein	7.00E-65
Contig983	At3g07250	nuclear transport factor 2 (NTF2) family protein / RNA recognition motif (RRM)-containing protein	5.00E-17
Contig983	At3g07250	nuclear transport factor 2 (NTF2) family protein / RNA recognition motif (RRM)-containing protein	9.00E-19
Contig832	At3g08920	rhodanese-like family protein	6.00E-51
Contig1003	At3g10020	unknown protein	3.00E-08
Contig448	At3g10260	reticulon family protein	3.00E-78
Contig584	At3g10360	XhPUM1 (PUMILIO-domain containing protein); RNA binding	1.00E-78
Contig881	At3g12390	nascent polypeptide associated complex alpha chain protein, putative / alpha-NAC, putative	2.00E-44
Contig227	At3g12490	cysteine protease inhibitor, putative / cystatin, putative	3.00E-28
Contig290	At3g12500	basic chitinase	2.00E-06

Contig1480	At3g12500	basic chitinase	3.00E-07
Contig404	At3g12500	basic chitinase	7.00E-60
Contig1411	At3g12760	unknown protein, putative leucine zipper protein	6.00E-08
Contig345	At3g13062	unknown protein	7.00E-06
Contig452	At3g14420	(S)-2-hydroxy-acid oxidase, peroxisomal, putative / glycolate oxidase, putative / short chain alpha-hydroxy acid oxidase, putative; glycolate oxidase	3.00E-92
Contig197	At3g15210	Ethylene-responsive transcription factor (XhERF)	2.00E-21
Contig1224	At3g15220	protein kinase, putative	8.00E-31
Contig685	At3g15353	metallothionein type I	6.00E-04
Contig829	At3g15630	unknown protein	5.00E-10
Contig133	At3g15780	unknown protein	4.00E-04
Contig153	At3g16480	MPPALPHA (mitochondrial processing peptidase alpha subunit); metalloendopeptidase	2.00E-81
Contig1241	At3g18060	transducin family protein / WD-40 repeat family protein	2.00E-35
Contig398	At3g18280	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	1.00E-03
Contig939	At3g18830	Sorbitol transporter	3.00E-63
Contig739	At3g19240	unknown protein	1.00E-122
Contig425	At3g20770	XhEIN3 (ETHYLENE-INSENSITIVE3); transcription factor	2.00E-09
Contig1093	At3g22110	PAC1 (20S proteasome alpha subunit C1); peptidase	5.00E-122
Contig721	At3g22490	LEA6	2.00E-21
Contig734	At3g22490	LEA6	4.00E-54
Contig737	At3g22850	unknown protein	9.00E-87
Contig842	At3g24520	Heat shock transcription factor	3.00E-29
Contig801	At3g24800	PRT1 (PROTEOLYSIS 1); ubiquitin-protein ligase	3.00E-09
Contig990	At3g25570	S-adenosylmethionine decarboxylase family protein	3.00E-07
Contig176	At3g27700	LHCB2:4 (Photosystem II light harvesting complex gene 2.3); chlorophyll binding	1.00E-133
Contig956	At3g29270	ubiquitin-protein ligase	2.00E-79
Contig794	At3g44330	unknown protein	5.00E-67
Contig1128	At3g47340	glutamine-dependent asparagine synthetase , XhASN1 (Xhdin6)	2.00E-67
Contig37	At3g47470	LHCA4 (Photosystem I light harvesting complex gene 4); Chlorophyll a-b binding protein 4, chloroplast precursor	2.00E-96
Contig1146	At3g48710	GTP binding / RNA binding	5.00E-05
Contig94	At3g48940	remorin family protein	9.00E-26
Contig755	At3g49430	splicing factor SR1D	2.00E-63
Contig1018	At3g50830	COR413-PM2 (cold regulated 413 plasma membrane 2)	9.00E-52
Contig646	At3g50910	unknown protein	8.00E-27
Contig899	At3g51250	Senescence-associated family protein	3.00E-31
Contig417	At3g51550	protein kinase family protein	6.00E-45
Contig181	At3g52300	ATP synthase D chain, mitochondrial	4.00E-06
Contig610	At3g52960	Type 2 peroxiredoxin	4.00E-62
Contig1007	At3g53040	LEA3	2.00E-06
Contig740	At3g53040	LEA3	1.00E-06
Contig823	At3g53770	LEA7	9.00E-03
Contig1286	At3g54040	photoassimilate-responsive protein-related	7.00E-18

Contig529	At3g54420	class IV chitinase	1.00E-83
Contig306	At3g54890	XhLHCA1; chlorophyll binding, chlorophyll A/B-binding protein	4.00E-104
Contig943	At3g55440	Triosephosphate isomerase, cytosolic (EC 5.3.1.1) (TIM) (Triose- phosphate isomerase)	1.00E-73
Contig437	At3g55980	zinc finger (CCCH-type) family protein	3.00E-12
Contig728	At3g56010	hypothetical protein	3.00E-35
Contig608	At3g56400	WRKY transcripton factor	6.00E-27
Contig1366	At3g56740	ubiquitin-associated (UBA)/TS-N domain-containing protein	1.00E-69
Contig1180	At3g56940	Xh103 (DICARBOXYLATE DIIRON 1),Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclaseCrd1, a plastid-localized putative diiron protein	4.00E-35
Contig667	At3g57340	Heat shock protein DnaJ family protein	2.00E-35
Contig30	At3g58680	XhMBF1B/MBF1B (MULTIPROTEIN BRIDGING FACTOR 1B); DNA binding / transcription coactivator	2.00E-46
Contig1181	At3g61470	Photosystem related	8.00E-74
Contig571	At3g61790	E3 ubiquitin-protein ligase SINAT3 (Seven in absentia homolog 3)	3.00E-126
Contig1138	At3g62250	XhUBQ5, ubiquitin extension protein	2.00E-53
Contig1455	At3g62420	bZIP transcription factor	1.00E-23
Contig847	At3g62730	unknown protein	7.00E-85
Contig843	At3g63530	Zn-finger, RING domain containing protein	1.00E-29
Contig768	At4g00170	vesicle-associated membrane family protein / VAMP family protein	1.00E-71
Contig541	At4g00430	PIP1 aquaporin	3.00E-139
Contig61	At4g01050	hydroxyproline-rich glycoprotein family protein	9.00E-47
Contig413	At4g01995	unknown protein	3.00E-64
Contig32	At4g02020	EZA1 (SWINGER); transcription factor;Probable Polycomb group protein EZA1 (CURLY LEAF-like 1) (Protein SET DOMAIN GROUP 10), Putative methyl transferase	6.00E-50
Contig796	At4g02510	chloroplast protein import component Toc159-like	8.00E-55
Contig22	At4g03020	transducin family protein / WD-40 repeat family protein	5.00E-16
Contig1393	At4g03230	S-locus lectin protein kinase family protein	1.00E-43
Contig1097	At4g03280	PETC (PHOTOSYNTHETIC ELECTRON TRANSFER C)	2.00E-67
Contig385	At4g03520	XhTHM2 (thioredoxin M-type 2); thiol-disulfide exchange intermediate	1.00E-35
Contig461	At4g05320	UBQ10 (POLYUBIQUITIN 10); protein binding	6.00E-111
Contig1141	At4g05320	UBQ10 (POLYUBIQUITIN 10); protein binding	1.00E-150
Contig891	At4g05320	UBQ10 (POLYUBIQUITIN 10); protein binding	6.00E-150
Contig944	At4g05320	UBQ10 (POLYUBIQUITIN 10); protein binding	5.00E-119
Contig444	At4g05420	DDB1A (UV-damaged DNA-binding protein 1A); DNA binding ;DDB1 interacts with histone acetyltransferase complexes	3.00E-111
Contig1029	At4g10490	oxidoreductase, 2OG-Fe(II) oxygenase family protein, putative flavanone 3-beta-hydroxylase	1.00E-59
Contig941	At4g11410	short-chain dehydrogenase/reductase (SDR) family protein	6.00E-81
Contig670	At4g11600	XhGPX6 (GLUTATHIONE PEROXIDASE 6); glutathione peroxidase	1.00E-76
Contig948	At4g11650	Thaumatin-like protein precursor	8.00E-43

Contig573	At4g11910	STAY-GREEN2 protein	2.00E-68
Contig276	At4g12800	PSAL (photosystem I subunit L) Photosystem I reaction center subunit XI, chloroplast precursor (PSI-L) (PSI subunit V)	2.00E-67
Contig544	At4g13010	oxidoreductase, zinc-binding dehydrogenase family protein ,Putative chloroplastic quinone-oxidoreductase homolog	3.00E-115
Contig1470	At4g13170	60S ribosomal protein L13A (RPL13aC)	2.00E-76
Contig520	At4g13250	short-chain dehydrogenase/reductase (SDR) family protein	5.00E-137
Contig1166	At4g15093	catalytic LigB subunit of aromatic ring-opening dioxygenase family	2.00E-12
Contig478	At4g15160	protease inhibitor/seed storage/lipid transfer protein (LTP);cell wall protein like	2.00E-21
Contig564	At4g15630	integral membrane family protein	1.00E-11
Contig572	At4g15910	LEA3	1.00E-07
Contig1046	At4g15910	LEA3	9.00E-08
Contig522	At4g16160	protein translocase	5.00E-48
Contig1047	At4g16160	protein translocase	2.00E-48
Contig1092	At4g16210	enoyl-CoA hydratase/isomerase family protein	7.00E-103
Contig744	At4g17750	Heat shock factor 1	2.00E-32
Contig436	At4g17880	XhbHLH transcription , MYC7E	7.00E-62
Contig665	At4g18060	Neutrophil cytosol factor 2 family protein	9.00E-08
Contig738	At4g18100	60S ribosomal protein L32A	5.00E-52
Contig853	At4g18220	purine permease family protein	2.00E-16
Contig962	At4g18690	hypothetical protein	2.00E-06
Contig1098	At4g19210	RNase L inhibitor protein 2	8.00E-95
Contig253	At4g19230	Abscisic acid 8'-hydroxylase 1 (ABA 8'-hydroxylase 1) (Cytochrome P450 707A1)	6.00E-35
Contig623	At4g19810	hydrolyse	7.00E-68
Contig1246	At4g20360	translational elongation factor Tu	3.00E-100
Contig434	At4g20780	calcium binding protein	6.00E-47
Contig810	At4g21020	LEA3	1.00E-05
Contig173	At4g21280	XhpsbQ, photosystem II oxygen-evolving complex protein 3-like	3.00E-69
Contig206	At4g21760	putative beta-glucosidase	3.00E-39
Contig733	At4g21870	26.5 kDa class P-related heat shock protein (HSP26.5-P)	8.00E-24
Contig472	At4g24090	unknown protein	6.00E-50
Contig756	At4g24380	hydrolase, acting on ester bonds	3.00E-83
Contig139	At4g24690	putative ubiquitin-associated (UBA) protein	2.00E-26
Contig989	At4g25140	OLEO1 (OLEOSIN1)	2.00E-28
Contig1178	At4g25570	XhCYB-2 (X humilis cytochrome b561 -2); carbon-monoxide oxygenase	5.00E-71
Contig1136	At4g26740	XhS1 (Xerophyta humilis SEED GENE 1); calcium ion binding , embryo-specific protein 1 (XhS1); caleosin-related family protein	1.00E-38
Contig876	At4g26810	SWIB complex BAF60b domain-containing protein	1.00E-31
Contig232	At4g27130	Protein translation factor SUI1 homolog/(Translational initiation factor 1) (Protein eIF1)	4.00E-48
Contig683	At4g27350	unknown protein	4.00E-79
Contig291	At4g27390	unknown protein	4.00E-24

Contig1084	At4g27690	Vacuolar protein sorting-associated protein 26 XhVPS26	2.00E-100
Contig380	At4g27960	UBC9 (UBIQUITIN CONJUGATING ENZYME 9); ubiquitin-protein ligase	5.00E-73
Contig338	At4g27960	UBC9 (UBIQUITIN CONJUGATING ENZYME 9); ubiquitin-protein ligase	2.00E-44
Contig230	At4g27960	UBC9 (UBIQUITIN CONJUGATING ENZYME 9); ubiquitin-protein ligase	7.00E-80
Contig736	At4g28390	ATP/ADP carrier protein	2.00E-148
Contig502	At4g29120	6-phosphogluconate dehydrogenase NAD-binding domain-containing protein, putative D-threonine dehydrogenase	4.00E-119
Contig1090	At4g29190	zinc finger (CCCH-type) family protein	2.00E-74
Contig637	At4g29480	mitochondrial ATP synthase g subunit family protein	3.00E-37
Contig785	At4g31290	ChaC-like family protein	5.00E-88
Contig621	At4g31530	catalytic/ coenzyme binding	8.00E-57
Contig659	At4g31870	Glutathione peroxidase	2.00E-76
Contig1076	At4g32020	unknown protein	7.00E-06
Contig387	At4g32940	GAMMA-VPE (Vacuolar processing enzyme gamma); cysteine-type endopeptidase	1.00E-64
Contig477	At4g33050	EDA39 (embryo sac development arrest 39); calmodulin binding	3.00E-140
Contig603	At4g33540	metallo-beta-lactamase family protein	1.00E-76
Contig1158	At4g34040	Zn finger family protein	3.00E-59
Contig1123	At4g34050	caffeoyl-CoA 3-O-methyltransferase, putative	6.00E-73
Contig349	At4g34350	XhClb6, (CHLOROPLAST BIOGENESIS 6); 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase	4.00E-97
Contig848	At4g34450	coatomer gamma-2 subunit, putative	5.00E-67
Contig89	At4g34660	SH3 domain-containing protein 2 (SH3P2)	4.00E-73
Contig482	At4g35090	Catalase-2	5.00E-32
Contig235	At4g35450	XhAKR2, Ankyrin repeat domain-containing protein 2 (AtAKR2) (AKR2); similar to TGB12K interacting protein 2 [Nicotiana tabacum	5.00E-97
Contig344	At4g36910	LEJ2 (LOSS OF THE TIMING OF ET AND JA BIOSYNTHESIS 2)	2.00E-07
Contig953	At4g37220	stress-responsive protein, putative	5.00E-51
Contig675	At4g37870	phosphoenolpyruvate carboxykinase (ATP), putative / PEP carboxykinase, putative / PEPCK, putative	1.00E-154
Contig677	At4g38580	XhFP6 (FARNESYLATED PROTEIN 6); metal ion binding	7.00E-43
Contig533	At4g39090	stress inducible cysteine proteinase	5.00E-44
Contig773	At4g39150	DNAJ heat shock N-terminal domain-containing protein	1.00E-40
Contig462	At4g39710	Peptidylprolyl isomerase, FKBP-type domain containing protein	6.00E-61
Contig554	At4g39730	lipid-associated family protein, dehydration stress-induced protein	5.00E-47
Contig1016	At5g01300	phosphatidylethanolamine-binding family protein	7.00E-45
Contig716	At5g01300	phosphatidylethanolamine-binding family protein	2.00E-47
Contig1035	At5g01670	aldose reductase	5.00E-124
Contig545	At5g01670	aldose reductase	7.00E-132
Contig1021	At5g01750	Unknown Protein	2.00E-26
Contig188	At5g02560	histone H2A, putative	1.00E-46

Contig825	At5g02610	60S ribosomal protein L35 (RPL35D)	3.00E-30
Contig228	At5g03900	unknown protein	7.00E-32
Contig609	At5g04440	unknown protein	6.00E-45
Contig607	At5g05220	unknown protein	4.00E-03
Contig152	At5g05340	Xhperox P52; Peroxidase 52 precursor	5.00E-19
Contig808	At5g05440	unknown protein	2.00E-51
Contig1027	At5g06370	NC domain-containing protein	1.00E-44
Contig855	At5g06760	LEA1	5.00E-04
Contig697	At5g07330	unknown protein	1.00E-21
Contig940	At5g07330	unknown protein	1.00E-21
Contig65	At5g09570	unknown protein	5.00E-27
Contig746	At5g09650	inorganic pyrophosphatase family protein	8.00E-101
Contig858	At5g10360	40S ribosomal protein S6	3.00E-104
Contig1096	At5g10695	unknown protein	4.00E-14
Contig451	At5g11280	unknown protein	7.00E-73
Contig1140	At5g11770	NADH dehydrogenase / ubiquinone?	2.00E-85
Contig293	At5g12010	unknown protein	9.00E-115
Contig816	At5g12110	elongation factor 1B alpha-subunit 1 (eEF1Balpha1)	3.00E-78
Contig947	At5g13800	hydrolase, alpha/beta fold family protein	8.00E-111
Contig324	At5g14320	30S ribosomal protein S13	7.00E-30
Contig1080	At5g16450	dimethylmenaquinone methyltransferase family protein;Regulator of ribonuclease-like protein 2	1.00E-75
Contig988	At5g17330	GAD (Glutamate decarboxylase 1); calmodulin binding	2.00E-09
Contig1025	At5g17460	unknown protein	8.00E-20
Contig1048	At5g17460	unknown protein	8.00E-129
Contig589	At5g19855	unknown protein	3.00E-52
Contig817	At5g19855	unknown protein	5.00E-58
Contig1529	At5g20150	SPX (SYG1/Pho81/XPR1) domain-containing protein (molecular function not known)	9.00E-33
Contig1509	At5g20700	senescence-associated protein-related	2.00E-12
Contig364	At5g21090	leucine-rich repeat protein, putative	4.00E-87
Contig1151	At5g22250	CCR4-NOT transcription complex protein	7.00E-16
Contig445	At5g23540	26S proteasome non-ATPase regulatory subunit 14 (26S proteasome regulatory subunit rpn11)	2.00E-117
Contig74	At5g23690	polynucleotide adenylyltransferase family protein	6.00E-42
Contig930	At5g23710	DNA binding / DNA-directed RNA polymerase	1.00E-16
Contig450	At5g24030	C4-dicarboxylate transporter/malic acid transport family protein	2.00E-50
Contig551	At5g24130	unknown protein	5.00E-18
Contig602	At5g24130	unknown protein	4.00E-57
Contig692	At5g24400	6-phosphogluconolactonase-like protein	9.00E-44
Contig1534	At5g24510	60s acidic ribosomal protein P1, putative	5.00E-24
Contig1467	At5g24650	mitochondrial import inner membrane translocase subunit Tim17/Tim22 family protein	4.00E-30
Contig1290	At5g25560	zinc finger (C3HC4-type RING finger) family protein	5.00E-51
Contig1110	At5g25880	XhNADP-ME3 (NADP-MALIC ENZYME 3)	4.00E-78
Contig922	At5g35735	auxin-responsive family protein	3.00E-19
Contig269	At5g36230	eIF4-gamma/eIF5/eIF2-epsilon domain-containing protein	2.00E-49

Contig29	At5g36790	phosphoglycolate phosphatase, putative	9.00E-150
Contig311	At5g37600	XhGSR1 (glutamine synthase clone R1); glutamate-ammonia ligase	6.00E-34
Contig1001	At5g37770	XhTCH2 (TOUCH 2); CALMODULIN-RELATED PROTEIN (potential Ca(2+) sensor that is responsive to ABA signalling)	2.00E-13
Contig735	At5g39250	F-box family protein	1.00E-40
Contig51	At5g39670	calcium-binding EF hand family protein	7.00E-16
Contig466	At5g40250	zinc finger (C3HC4-type RING finger) family protein	9.00E-66
Contig130	At5g40370	Glutaredoxin-C2 (XhGrxC2)	5.00E-37
Contig974	At5g40760	Glucose-6-phosphate dehydrogenase	2.00E-97
Contig1175	At5g41700	UBC8 (UBIQUITIN CONJUGATING ENZYME 8); ubiquitin-protein ligase	9.00E-40
Contig576	At5g42050	unknown protein	6.00E-70
Contig1034	At5g42190	XhSK2 (X humilis SKP1-LIKE 2); ubiquitin-protein ligase	8.00E-54
Contig686	At5g42300	Ubiquitin-like protein 5	1.00E-28
Contig488	At5g43060	cysteine proteinase, putative / thiol protease, similar to senescence-associated cysteine protease	6.00E-29
Contig392	At5g43060	cysteine proteinase, putative / thiol protease, similar to senescence-associated cysteine protease	9.00E-69
Contig184	At5g43430	ETFBETA; electron carrier	8.00E-24
Contig1567	At5g44250	Unknown Protein	2.00E-05
Contig778	At5g44310	LEA3	2.00E-05
Contig818	At5g44310	LEA3	6.00E-06
Contig599	At5g44310	LEA3	3.00E-06
Contig701	At5g44310	LEA3	2.00E-07
Contig863	At5g44310	LEA3	4.00E-08
Contig875	At5g44310	LEA3	4.00E-04
Contig752	At5g45930	magnesium chelatase subunit of protochlorophyllide reductase	2.00E-168
Contig42	At5g46020	unknown protein	1.00E-15
Contig1028	At5g46410	NLI interacting factor (NIF) family protein	2.00E-83
Contig446	At5g48520	unknown protein	1.00E-58
Contig105	At5g48545	histidine triad family protein / HIT family protein	5.00E-24
Contig682	At5g49560	unknown protein	4.00E-30
Contig1155	At5g50250	RNA binding protein	2.00E-77
Contig771	At5g50360	unknown protein	1.00E-27
Contig850	At5g50600	short-chain dehydrogenase/reductase (SDR) family protein	3.00E-32
Contig1477	At5g50720	XhHVA22e; abscisic acid-induced-like protein	5.00E-33
Contig332	At5g50740	Heavy metal transport/detoxification protein, metal ion binding	2.00E-27
Contig1133	At5g51440	23.5 kDa mitochondrial small heat shock protein (HSP23.5-M)	1.00E-29
Contig1083	At5g51970	putative sorbitol dehydrogenase	5.00E-50
Contig1073	At5g52740	heavy-metal-associated domain-containing protein	2.00E-08
Contig14	At5g53000	TXhAP46 (2A PHOSPHATASE ASSOCIATED PROTEIN OF 46 KD); PP2A regulatory subunit TAP46	5.00E-120
Contig747	At5g54190	PORA (Protochlorophyllide reductase A); oxidoreductase/ protochlorophyllide reductase	5.00E-166
Contig343	At5g54270	LHCB3 (LIGHT-HARVESTING CHLOROPHYLL BINDING PROTEIN 3)	9.00E-107
Contig93	At5g54680	DNA binding / transcription factor	5.00E-61

Contig555	At5g55120	unknown protein	3.00E-26
Contig1044	At5g55120	unknown protein	8.00E-113
Contig1118	At5g56550	unknown protein	2.00E-05
Contig1112	At5g56550	unknown protein	1.00E-07
Contig71	At5g57550	XhXTR3 (XYLOGLUCAN ENDOTRANSGLYCOSYLASE 3); hydrolase, acting on glycosyl bonds	1.00E-104
Contig402	At5g57890	anthranilate synthase beta subunit	4.00E-86
Contig613	At5g57950	26S proteasome regulatory subunit, putative	2.00E-22
Contig465	At5g58330	malate dehydrogenase (NADP), chloroplast, putative	2.00E-75
Contig565	At5g59320	LTP3 (LIPID TRANSFER PROTEIN 3); lipid binding	2.00E-20
Contig1085	At5g59550	zinc finger (C3HC4-type RING finger) family protein; ubiquitin-interacting factor 1b	2.00E-54
Contig676	At5g59820	Zn finger transcription factor	1.00E-24
Contig432	At5g60360	XhALP (X. humilis ALEURAIN-LIKE PROTEASE); Senescence- associated gene product 2;cysteine-type peptidase	9.00E-11
Contig730	At5g60460	sec61beta family protein; putative protein transport protein subunit	2.00E-14
Contig1533	At5g60600	GcpE (CHLOROPLAST BIOGENESIS 4); 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, csb3 constitutive subtilisin 3	2.00E-50
Contig421	At5g61410	Ribulose-phosphate 3-epimerase, chloroplast precursor (Pentose-5-phosphate 3-epimerase) (PPE) (RPE) (R5P3E)	2.00E-121
Contig55	At5g62740	band 7 family protein, hypersensitive-induced response protein	5.00E-95
Contig979	At5g63660	LCR74/PDF2.5 (Low-molecular-weight cysteine-rich 74)	3.00E-05
Contig174	At5g64040	PSAN (photosystem I reaction center subunit PSI-N); calmodulin binding	2.00E-48
Contig203	At5g64180	unknown protein	6.00E-23
Contig416	At5g64300	XhGCH (X humilis GTP cyclohydrolase II);	3.00E-10
Contig1257	At5g64460	Phosphoglycerate mutase-like protein	1.00E-65
Contig1139	At5g65165	SDH2-3 (Succinate dehydrogenase 2-3); mitochondrial succinate dehydrogenase iron-sulphur subunit	4.00E-88
Contig1296	At5g65570	putative pentatricopeptide	5.00E-79
Contig985	At5g66110	metal ion binding	3.00E-46
Contig812	At5g66140	Proteasome subunit alpha type 7	4.00E-114
Contig693	At5g66400	RAB18	2.00E-04
Contig993	At5g66460	(1-4)-beta-mannan endohydrolase, putative	9.00E-30
Contig1561	At5g66570	XhpsbO-1, OEC-33, photosystem II,Oxygen-evolving enhancer protein 1-1, chloroplast precursor (OEE1)	2.00E-52
Contig914	At5g66780	unknown protein	6.00E-09
Contig97	At5g67500	porin, putative	4.00E-88
Contig933	At5g55190	putative small Ras GTP-binding protein	2.00E-39
Contig280		no match	
Contig645		no match	
Contig67		no match	
Contig1130		no match	
Contig772		no match	
Contig1126		no match	
Contig1006		no match	

Contig655		no match	
Contig809		no match	
Contig811		no match	
Contig102		no match	
Contig1002		no match	
Contig1005		no match	
Contig1011		no match	
Contig1015		no match	
Contig1026		no match	
Contig1030		no match	
Contig1033		no match	
Contig1040		no match	
Contig1042		no match	
Contig1043		no match	
Contig1057		no match	
Contig1065		no match	
Contig1075		no match	
Contig109		no match	
Contig1095		no match	
Contig1100		no match	
Contig1102		no match	
Contig1105		no match	
Contig1114		no match	
Contig1132		no match	
Contig1135		no match	
Contig115		no match	
Contig1156		no match	
Contig1160		no match	
Contig1161		no match	
Contig1165		no match	
Contig1167		no match	
Contig1169		no match	
Contig1170		no match	
Contig1171		no match	
Contig1185		no match	
Contig1194		no match	
Contig1195		no match	
Contig1197		no match	
Contig1208		no match	
Contig123		no match	
Contig125		no match	
Contig1252		no match	
Contig127		no match	
Contig1284		no match	
Contig1293		no match	
Contig1294		no match	

Contig1311		no match	
Contig1316		no match	
Contig1317		no match	
Contig132		no match	
Contig1320		no match	
Contig1333		no match	
Contig1337		no match	
Contig1341		no match	
Contig1358		no match	
Contig1363		no match	
Contig1364		no match	
Contig1378		no match	
Contig1398		no match	
Contig1410		no match	
Contig1420		no match	
Contig1427		no match	
Contig143		no match	
Contig1433		no match	
Contig1441		no match	
Contig1442		no match	
Contig1444		no match	
Contig1446		no match	
Contig1450		no match	
Contig1462		no match	
Contig1464		no match	
Contig1465		no match	
Contig1471		no match	
Contig1482		no match	
Contig1485		no match	
Contig1486		no match	
Contig1488		no match	
Contig1488		no match	
Contig151		no match	
Contig154		no match	
Contig1569		no match	
Contig182		no match	
Contig191		no match	
Contig195		no match	
Contig199		no match	
Contig207		no match	
Contig214		no match	
Contig249		no match	
Contig25		no match	
Contig26		no match	
Contig260		no match	
Contig279		no match	

Contig292		no match	
Contig302		no match	
Contig303		no match	
Contig323		no match	
Contig328		no match	
Contig337		no match	
Contig341		no match	
Contig352		no match	
Contig354		no match	
Contig36		no match	
Contig371		no match	
Contig373		no match	
Contig393		no match	
Contig401		no match	
Contig410		no match	
Contig429		no match	
Contig439		no match	
Contig442		no match	
Contig447		no match	
Contig457		no match	
Contig460		no match	
Contig467		no match	
Contig47		no match	
Contig474		no match	
Contig476		no match	
Contig49		no match	
Contig492		no match	
Contig503		no match	
Contig510		no match	
Contig521		no match	
Contig540		no match	
Contig542		no match	
Contig548		no match	
Contig563		no match	
Contig581		no match	
Contig593		no match	
Contig594		no match	
Contig600		no match	
Contig605		no match	
Contig617		no match	
Contig618		no match	
Contig629		no match	
Contig664		no match	
Contig678		no match	
Contig699		no match	
Contig700		no match	

Contig719		no match	
Contig723		no match	
Contig725		no match	
Contig731		no match	
Contig748		no match	
Contig751		no match	
Contig758		no match	
Contig760		no match	
Contig770		no match	
Contig800		no match	
Contig804		no match	
Contig837		no match	
Contig846		no match	
Contig859		no match	
Contig878		no match	
Contig894		no match	
Contig895		no match	
Contig90		no match	
Contig905		no match	
Contig925		no match	
Contig937		no match	
Contig945		no match	
Contig949		no match	
Contig957		no match	
Contig958		no match	
Contig959		no match	
Contig963		no match	
Contig977		no match	
Contig982		no match	
Contig987		no match	
Contig999		no match	
Contig1449		no match	
Contig408		no match	

A.4.1.
Cluster 1

Contig1480

ACATGCTCTGCTGCAGCAAGTGGGGTTACTGTGGCATGACGAACGCTTGGTTCGGCGAAGGTTGCCAGAGCCAGTGTCAAGGTCCG
AAACAATGCGGAGTCGAAGCCGGCAACGTGACGTGCCCGAGCGGCCTCTGCTGCAGCAAGCATGGTTACTGCGGCACCTCTGGTGA
TTACTGCGCGATCAGTCAAGGGTGCCAGAGCCAGTGCGCCGGCGAAACGGAGATACGATGTGGCCGTGACGCCGGCAATGCAGTGT
GCCCGGACGGGCTCTGTTGCAGCCCGTATGGTTATTGCGGTAGCTCAGAACGCTACTGCGGCAACGGCTGCCAGAGCCAATGCTCA
AACGAACTCAGGTGCGGCCTGCAGGCCGGCGGAGCAATGTGCCCGGAAGGTCTCTGCTGCAGCAAATCTGGTTATTGCGGTAAAAC
GAAAGAACACTGCGGCGTAGGCTGCCAAAACCAGTGCACTTCATCATCTGAACTATCGAACGTGAATCAGGTGATAAAAGAGGTTA
TATTCGGAAACAAATGATGCATTTAGAAGTTCTAGAGATGTCGATCGCGGCCCGTAGATACACCAATTTTATAGTTGGTGTAGATA
GAACTAGTCGAAGTGAATAACTTGGTCGGGAGCGCTTTACGCTTCGACTTTGTGTTGTAGGTTCGTATGTACGTGCTAGAGATGAA
GTAATAAACCGGGCCTTTCGTGTGT

LEFT PRIMER: CTGCAGCAAGTGGGGTTACT
RIGHT PRIMER: CATTGTTTCGGACCTTGACA
PRODUCT SIZE: 84
SEQUENCE SIZE: 713

Contig408

TCCTCGGAACATGGCTGAAGCTGTGAGGTAACGTCCGTGGCGTGGATCAGCTGCACACATCATGTTCTTAGCGTCCCACATTTGCT
GTGTCAGCTCCGGGACTGTGAGAGAGATGTACTGTTGAGAGCCACGGGAAGTGAGAGGGGCGAAACCAACCATGAAGAAATGTAGA
CGAGGGAATGGGATAAGGTTCACGGCGAGTTTCCTAAGGTCAGAGTTGAGTTGTCCAGGGAATCTTAAAGAGCAAGTCACACCACT
CATAGTTGCGGAGATCAAGTGGTTCAAGTCTCCAAAGCTAGGAGTGCTGAGTTTGAGAGTGCGGAAACAGATATCGTAAAGAGCTT
CGTTGTCAAGGACCATGCATTCATCAGCATTCTCTACAAGTTGGTGGACTGAGAGAGTTGCATTGTACGGCTCAACGACTGTGTCT
GAGACCTTTGGAGATGGGAAAACAGAGAATGTGAGCATCATTCTGTCTGGGTATTCCTCACGGATCTTTGATATCAAGAGAGTTCC
CATTCCTGAACCCGTGCCTCCTCCAAGAGAGTGACATACTTGAAACCCTTGAAGGCAATCACAGTTCTCAGCTTCCTTACGAACAA
CATCGAGAACAGCATCAATAAGTTCAGCACCTTCAGTATAATGACCTTTCGCCCAATTATTTCCGGCACCGTATTGACCAAAGACG
AAGTTATCAGG

LEFT PRIMER: TCCGGGACTGTGAGAGAGAT
RIGHT PRIMER: GAAACTCGCCGTGAACCTTA
PRODUCT SIZE: 111
SEQUENCE SIZE: 699

Cluster 2

Contig1567

CTTCTCGAAGTTACCGGTGGTTATATTTACCACGGCAACTCAGCAATTACTGTCAAATGGGAACTTGCGGAGGCTACGGGTTGGAC
TAGATGATGATCCTTTCTTACCAAGATCCTTAGAGTATAACGATGCTAAAGAAGGTGAACTATATATGAATTCGTTCCAGGTACAG
AATCCACCAATGATCAATGCTCATAGTGTACTGGGCCAATTCCTCTTTGATGTGTGTGTTCCAAAGAACATTGACGAGTGGGACAT
AAAATCATCGATGAATGGAAGGCAAACATTCAGGTCCGCTCGCAGTTACGGGTCTCTTAATCCTATCAGACGTATGAGGCGTTCAC
GATTGTAAAACTTGAAAGGGATGCTGGTTATCTTAACAGAAGAATGATCTATACCCACAAAAATCCATCCTGAATTCCTGTAACTA
TGTAATATGATGTTTGATAGGAATATGTTTAGAAATATCTAGCATTACTACACTTTGTTAAAAAAGGCAAAAGCTTGAGGGATGTA
TATAATATGTTGAGCAATGTTTACCAGTATATATTGAGCTGACCGATGATACTGAATAAATGTAAAGCATGATCTATCACCAGCCT
AACATTTGTTTAAAAAT

LEFT PRIMER: AGGCTACGGGTTGGACTAGA
RIGHT PRIMER: AGAGGAATTGGCCCAGTACA
SEQUENCE SIZE: 619
PRODUCT SIZE: 148

Contig23

ATATGACCACCTGAGCTCACTATATACTCCATAACTTGCACCATAAAGAAGAAGAAGAAGAAGAAGAAGCAGAAGAAGAAGAAGAA
GACGAAAACATGAACTTCAGAACTCTAACCAGCGACGACTGCACGGACGACGCTCTGCGCCTCCGCCACGGCAGTGGCGCCAGACT
TTTTGTTGTTGTTCTTCTGCTTCTGCGACGTCGGTGTCGAGAGTTTGATCCTCTTCACGCCGGCGTCGTCCAAATCAATCTGCTCC
TGCCGACACTCCTTGCTGCAGAATGGGGTATCACCTCTGCAGATCCACAAATTAGTTTCTAAACGGCTGAAAGATGAATAATAAAT
AATATACAAGATGTCAAGAACAGACCTGTACATGAAGATGTCGCTGTTGGGAGGGAGAGGTTTGGAGCAGAGTTTGCAGGAGTCGA
GGAAGTGGCGGGGATTCTCGTCGCCACCGTCACCGACGAAGAATAGCCTCACAACCATCGGCGAGGAGTAGAAGCCGGCGCCGCCA
CGAGCACCCATAGAAGAAAATGCTTCCCTCCTTCTCTTCTTCTTGCACAAAGTTGTACGGTGGCTTTGCTAATTAGTTTGTACTTG
GCAGTGAAAATGAAAATATTAATGGGAGGTTATAAAAAAAG

LEFT PRIMER: CTGTTGGGAGGGAGAGGTTT
RIGHT PRIMER: CCGATGGTTGTGAGGCTATT
SEQUENCE SIZE: 643
PRODUCT SIZE: 104

Contig531

TCGGCACGAGGGAGAATTTGAGAGGAGAGGAGACGGAGAGAGAGGCTCGTACGAAGCAGACTGCTCGTAAATCCACTGGTGGGAAG
GCACCGAGGAAACAGCTTGCTACCAAGGCTGCGCGTAAGTCTGCCCCCACAACTGGCGGTGTTAAGAAGCCTCATCGCTACAGGCC
CGGTACTGTCGCTCTTCGTGAGATCCGTAAGTACCAGAAGAGTACTGAGCTGCTGATCAGGAAGTTGCCCTTCCAGAGGCTTGTTA

GGGAAATTGCCCAGGACTTCAAGACTGATTTGAGGTTCCAGAGCCATGCTGTGCTAGCGCTGCAGGAGGCTGCTGAAGCCTACCTT
GTGGGTCTGTTTGAGGACACCAACTTGTGTGCAATCCATGCCAAGCGTGTGACAATTATGCCAAAAGACATCCAGCTTGCTAGGAG
GATTCGTGGTGAGAGGGCGTGAGTGTCTATCTTGATGATATGGATGATATTTTAGTGGATGCTGATGGTTATATTAATTTGTTAAG
TGTAGTCTGTAGTTGTCGACCAAGACTTGTTGGTCTAATGGGTTAGAACTTTATGAATTGACCTATCCTTCGCCATTTACTTACTA
GCTTAATTATCTGTTTAATGTATATGACTGGAACTCCATTGTGCGTGCTGTGTTTTCAGTAATGTAGCTGAAACTAATATGCAGAT
CATGTCAATCTCTGGAACTGTGTTTCAATTTGGTTTATTTTGAACTTTGCAGTAATTTGGTTGGATCC

LEFT PRIMER: TTGCCCAGGACTTCAAGACT
RIGHT PRIMER: TGTCCTCAAACAGACCCACA
SEQUENCE SIZE: 756
PRODUCT SIZE: 99

Cluster 3

Contig1046

GCACGAGGGTCGTCAGCTGCGGAGTTCTGGAGGGACGAGAGAGAAGCTTCTTAAAGTTCCTCTCAAATCTCTGATCTGGTTCTTCG
CGGTGCCGGCGGCGAGGAGCTTGCGATCCTTTATCTATAAATAAAACGAAGTTGAGATGGCGGCCCGCGTTTTCTTTGGCTATGCT
TCCCTTGCTGCTGCCGTCTCCGACAGCATCGCTCTCGTTTCCAGCAGGAGGGGATACTTTGCTGCCTCTGCAGCTGCAACTGGGGT
GAGAAAGGTAGAGGAGAGGATGGTGACCATGATGAAGGGGACCAGAGAGGTCGCAGATCCGGCGGCGGCAACGTCGTGGGTGCCGG
ATCCTGTAACCGGCTATTACCGGCCGGCCAACCACGCTGTCGAGGTTGATGCGGCAGAGCTCCGTCAGATGTTGCTCTCTAGCAAA
GCGTCTAGGGGCTAGAGGGGATCCTTGGATGAGGGTGATGTACTTTTAGTTCGACGTATCCACTATGGTGATGAACTGTTTTAGCT
TGGTTACAGTGATCAGTTCGGTTTGGTTGGTAATAACAGCTTGCTTGGTTATATTTTTGTAATAAATATATGATGTTTTGATGATT
C

LEFT PRIMER: CAGCAGGAGGGGATACTTTG
RIGHT PRIMER: ACCTCTCTGGTCCCCTTCAT
SEQUENCE SIZE: 603
INCLUDED REGION SIZE: 603
PRODUCT SIZE: 96

Contig1169

GCACGAGGGCGGCGGTCTCTGGACTCGACAAAACCAAAGCTGTTGTCGAGGAGAAGGTAACGGAAAAGGAATAAGAAGAGAAGAAA
GCTTTAGAACAGCAGAAGGAAAACAGGATACTCGAGGCAGAAGCTAGCAAACAAGCAGCCATGTCACAGAACGCAGCTACCCGCGA
TCAAACCACGGGAGCCCCACGAATCAACCCACCAAAGAATCCACTGACCCCCGCAGCCAACGGACCCAACGTCGGCACTGCCCGGG
TTATGGCACTGGCCAGCCTGCCGGCGGCCATATTGAAGGGGGTAAGGCTGAGAGTCATCCCATTTCAAAGGCTACTGGCACCGGGA
GGCCATCCACCGCTCATAACCCACATCTCGGCAGCCAGCCCGGCACCATCACTGGAACTGGTGGAGCTTACACTTGAAGAACGACT
CCTTATTAGCTACATAATTATAATTGTCTTCATGCTTGTCTTGTTGATTTGAACTGTAAGTTGGCAAGTAACAGTCCTTTAAACGT
AATACGATGTATGAATAAACGTGGCCGGTGCATTTCAGGAAGTTTTTTCGATCGCCCTATCCTTTCCCAT

LEFT PRIMER: GCTGTTGTCGAGGAGAAGGTAA
RIGHT PRIMER: CTGTGACATGGCTGCTTGTT
PRODUCT SIZE: 117
SEQUENCE SIZE: 586

Contig1468

GCACGAGGGTTAATTTCCAAGAACCAATTAAGTGAAAACCCTAGAAGTCCAAAATCAGCTAAAAATGTCGGGTTTCTTTCGCAAGG
TGGAGCAGAAAGTTCAGTTGGAGAAGGTGGACGGAGGCAAGCTCCGCAAGGAAAAAGTGCTGCAGTTCGAAGAGACGGGCGGGAAG
CACCACAAAGAGAAAGTGGTTCACTACAAAGAAGATCTCCTCGATGGAAATTACGGTGGCGAAGTTGGATACGGTAATCAGATGAT
CAGGGCAGATCAATACGACAGCGGGTTGGTGAAGAAGATAAAACAAAAGATGAACAAGAGGAAGGATAAGATTAATGGCAGGAAGG
GCGCCGGCGACCAGCGAGATGACGACAGCGACAGCGACTACTGATGTATGTGTATATGTCTACCTAGCTATAGCAATGAGGGAGGA
GAGAGAGCAGGGCAATAGCTCATGAGGAAGGATGTTCAATGTTATATATAGT

LEFT PRIMER: GGGAAGCACCACAAAGAGAA
RIGHT PRIMER: ACCCGCTGTCGTATTGATCT
PRODUCT SIZE: 118
SEQUENCE SIZE: 482

Contig721

GACGCGCCGGCGATCAATGACGGCGGAAGTCTGGCGCCCGAGGCTCCACCGACACCAAGACCGGTGCCATCGGGGCGATCGGCTCA
GTCTGCGGCTGCCGATGACAAGAGCAGCACTAAGCTCGGGGATATTTAATCGGACGCGACGGAGAAGATGGAGAGAGACAAGCCCG
CGACGAGGGATGACGTGGCGTGGATGGCGGCAGCGGAGATGAGGAACGATCCGGCGGCGGAGACTCGCCGCGGCGGCGTTACCGAC
GGCGTTAGGGCGGCGGTGGAGATCAATCAAGGCTGAGTTCTAATCTATTGACGTTGTTGCTTGGGCATGTTTAGCATTAGGGTAGA
GGGTGGTAGCAGTTATCCCAATAATACTTACCACCTCCTAAAGTGTTTTTGAGTTGTACATACTAGTGTCCACGTGGCTTGTTTAT
GTGAGTTTATGTAAAAAGATATAATAATAATAACTAGACAACCCAATGCTTCAGCTTTAAAG

LEFT PRIMER: ATCAATGACGGCGGAAGTC
RIGHT PRIMER: CGAGCTTAGTGCTGCTCTTG
PRODUCT SIZE: 112
SEQUENCE SIZE: 492

Contig760

TCCCGGCCACGTCGGAGTTTTGCCGGTGTTAGACAATGGTGTTGCAGTAACATCACAGAATGTGTATTGCCGGCGGAGGCCACGGG
AATTGAAGATATTCCCGAACAGAGTCGTTCGGGAAACCTGCGTGTTCGCCGGCGGTGGTGCCGGAAGTGACGACGTCCCTGTGCCG
GCGGGCGGTGCTATTGGCAACCCCGCGGATGCGTCGAACGTGTTCGCTGGCGGAGGCCCCGGAACTGCCGCAGGCATGGTTCCTGG
GCAAGCTGGACCAGTTGCACCGGAGTCGAACGTGTTCGCCGGAGGCGGCAGCGGCACAGACAGAGTTTGAGTTGTGTCGTGTGTGC
TTGCGTAAGTTTAATTAGGGTTTAATATGTAATTAGATATTAAGGTTTGATTAGTATTTTGCTGGTGTCATATTCATAAATGTTGG
TCAATAGAAACTTAAAAATACTCGA

LEFT PRIMER: GGAGTTTTGCCGGTGTTAGA
RIGHT PRIMER: CGACTCTGTTCGGGAATATCT
PRODUCT SIZE: 100
SEQUENCE SIZE: 455

Contig772

GCACAGGGCAGATGGCCCCTCCCTCTACCCCGAAGCGCCTATAAAAGTAGCACCCTCATCTCCTTCTTATCCTTAAATTGAAACAG
AAACACCGTCGAGTTTGAGAAATTTTCAGGCATTCGATTCGATCTTCAAGCAGTCGATCATGGCGTCCTGCGGAGGCAACTGCGGC
TGCGGCCCGAACTGCAAGTGCGGCAGCAACTGCAACTGCGGAGGAAACAAGATGTACCCGGGATTGGCCGAGGAGAGGAGCACGAG
TACCCAGACAAACATACTCGGCGTCGCACCTCAGCAGGAGCGCCGGGAAGGATTTGAGGCCGGCCAACGTTCAGAGAACGGAGGCT
GCAAGTGCGGTCCTAACTGCAACTGCAACCCCTGCAACTGCAAATGAGCTTCAAATAATATCTAGCTTTGTTTTGTAAGGGAGAAT
ATGTAATAATAAATAAGTCTATTTAGCGGATGCTAGTTCTAGAAAATAAGTTCTTGTACTAAATATTACTATGCAAGAGTTCTTTC
CAGATTGACGATGTGAAATATTGTAACTTTTTATTACTGTATGTGACATTATTGAT

LEFT PRIMER: CAGGCATTCGATTCGATCTT
RIGHT PRIMER: TCCCGGGTACATCTTGTTTC
PRODUCT SIZE: 123
SEQUENCE SIZE: 572

Contig813

CTAGTTTGCAACAGGTGGAGGAAGTGAAATGTCTGCTAAGGATCATCCCAATATGGTCCACAGGCATCATCTACTACATTGCTATA
GTTCAACAGCAAACCTATGTTGTCTATGAAGCCCTTCAATCCGACAGGCATCTCGGCAACACATCGTTTTTAATTCCGGCAGCTTC
GTACTCCGTGTTCGCGATGCTGGCCATGACAATCTGGATACCGATCTACGACCGAATAATCGTGCCTCGCTTCCGAAAACTAACGG
GCCAAGAAGGTGGCATAACACTATTACAAAGAATGGGAGTAGGGCTAGTCCTAGCCATAATTGGCATGGTTGTCTCCGCCATCATT
GAGATTCAAAGAAGAAGTTATGCGATTCACCGCCCGACTATTGGTACCGCCCCGAGCGGTGGAGCAATTTCTTCTATGTCGAGTTT
ATGGCTTATCCCTCAACTAGTAATATTAGGCCTCTCGGAAGCTTTCAACGCAATAGGACAAATAGAGTTCTACTACAAGCAGTTCC
CAGAGAACATGAGGAGTGTTGCAGGGTCGCTGTTCTTTTTAGGTGCCGCGACAGCAAACTATCTCAGCGGGGCTCTCATCGGAGTC
GTCCACCGGGTAACCGGACGGGGAGCCGGGAAGAATTGGCTCGCCGGAGATTTGAATAAGGGGAGGCTGGATTACTACTACTTTTT
CATAGCAGGTTTAGGGGTTTTGAACCTTGTTTACTTCATGATGTGTGCTAAGTGGTACAAGTACAAGGAGGCAGATAGTAGTAGCA
ATGCTGGTGAGATTTCACTTGTTGATGAGATTATATCAAAGGAGCCTCTTGTTTGATTTACGATTTTTATACACTTATTTAGAGAA
TTATCAACATCACTTTTTTTTAATTTCCTTTTTATTGTATCGATATTCGAGCAAATTGTTACTATATATATGTATTTTGCTTGACA
AGTGCTAATTTGCTTATGAGAGGAGTCATACCGATCTCATAAATAGGTCATTTCGGTGATTTGAGTCTCAAATAAAAAAAAAAAAA
AAAAA

LEFT PRIMER: TGGTCCACAGGCATCATCTA
RIGHT PRIMER: GTACGAAGCTGCCGGAATTA
PRODUCT SIZE: 123

SEQUENCE SIZE: 1037

Contig864

GCTGGGGAACCGCCCTGGAAAACACCAAGGGGAGAAACCCATGGGAAAAGAAGGACCCAATCTGGAAGCTGCCAGGCAAGACAAGA
AGGATCAAGACCCGGAAACTGGATGCACGAGAAGAAGACCCAGACCGGCGGGTTCATGTCGGAGAAGAAGGATCAGAGCGGCAACG
TTCTCCAGCAGACGGGGGAGAATGTGAAGAACGCTGCCGGCTCTGCCGCCAACGCAGTGAAGAACACACTGGGAATGGGCGACGAC
ACAAGAAATTCCCAGATATAAGCCAGTTTAGTACATATATATGCACGGTATAAACGTATATGGAGTATGTTTTTGGTTGTATATTT
ACTATATTTTGGTANTAAAGATGGCTGCTGGTACAGATCGATGATGTAACTGATGAGCAGTCCACTATTACTCGAATATATATAAA
TAAAGGGTCGTATTTGAATAAACAAAC

LEFT PRIMER: CTGCCAGGCAAGACAAGAA
RIGHT PRIMER: GTTGCCGCTCTGATCCTTC
PRODUCT SIZE: 103
INCLUDED REGION SIZE: 457

Contig865

GCACGAGGGTCGCTCAGAGAAGAAGAGAACTCGCTTGCTGTTTTCTGAAGAATTTTAGTTGAATCTGTGAGAAATCACGATGGCGC
AGATCATGAACAAGGCAAAGGGCTTGGTGGCGGACAAAATCTCCAAGGTGAAGAAGCCTGAAGCCGAACTCACCGACGTCTCATTC
AAGCCACGTCAGCAAGGACGCTGCCACTGTCCTTGGCAAGCTCAACATCTCCAACCCCTACGATCATTCCATCCCCATCTGCGAGA
TTTCCTACTCTTTCCGAAGTGCTGGAAGGGAGATCGCGTCGGGGACGATGCCGGATCCTGGATCAATCAAGGGCAGTGAGACGACG
ACGCTGGACATCCCGCTCAAGGTGCCCTACAACATCCTGGTTAGCTTGATGAAGGACGTGGGAAGAGATTGGGACATCGACTACGA
GTTCCACGTGAACCTTAAGGTTGATCTTCCCATAATAAAGAAAATCACCATACCTGTGTCTAAGAAAGGAGAGTTGAAGCTGCCAA
CCCTGTCTGACATCTTCTAAGAACATCAAGTGGCTTGATTAGATAGAGCTCAGTGAGAGATAGGAGGTGGTACGTTCGATGTAAAA
CTATTATGTATGAGGAATTGAGGAGTTCATATCGACAGTCGTGTTGTCCTAGGGTTGTACTCTGCTTATATAATGAAAATAAGGAG
TAGATAAACTTGGCTTGACAAGTTTATTATAATATGTATTTCTTCAGTTGCTCAAGCATGCCGTGGTCGTTATCGGGCATGACAAG

LEFT PRIMER: GTCCTTGGCAAGCTCAACAT
RIGHT PRIMER: TCCAGCACTTCGGAAAGAGT
PRODUCT SIZE: 84
SEQUENCE SIZE: 774

Contig977

GAATTCCGCTGGGTACGGCGGTGGTCATGGCAGCGGTGCCACACATGGCTCAAGCTATGGTACCGGGGCTCTTCCTGGGGAACATG
GTACTGGGTATGGCAGCGGTCTGGGAGGCGCCGCCGGGACAGGTACTCAGGATCCAAGGTACGGCGTCCCCACACAAGGAGCAGGG
CACGGCAGCGCCGCCACACAGGGCGCTGCCGTTATGGCTTCTGGCCTTGGCACAAACGCTACGCATGGCTCAGGGGACCGCGCCCC
AACAGAAGGGGGCCATGGCGGTGCTGCCGCCACGCAGGGTGCTGGGCACGGCGGCGTCGCCACTGAAGGTGCGAAGTACGATGCAG
GGCAGCAGCATGGACATGAGAAACTAGGGGATGTGCTGCACGGATCAGGTAACCCGAGCTTGGAGAATAACGGAGAAGGGAGAAAA
AAGAAGGGAGGTTTGAAGCAGAAAATATTAGGAAGCTGTTTGGGCACCGAGTAGATTACATCTGTTATCTGTCTGCTTAATATGTT
GCACGTATAACGTACTTATTCGTGCTTGTGGTGGTATTGTGCATGTCTGATGTTTGTTTGCAGCTACGCATGTCGTGCTGCCGTAT
ATTATGTTTATAATATAATATATAATTTTACAAAAAAAAAAAAAAAAAAAA

LEFT PRIMER: CGTGCTTGTGGTGGTATTGT
RIGHT PRIMER: AATATACGGCAGCACGACAT
PRODUCT SIZE: 69
INCLUDED REGION SIZE: 653

Cluster 4

Contig580

TATTTTATATAAATTTTACACTTAAATAGGAATTTCGATTACTCGATAATCTAATATCGTTGAAACAATTCGAGGTTACAAATGGA
TCATAATGGCACAACCCAATGATCACTTGCTAGCAATCTCCAAATACATCCTGCTAATAGATAATGGCTATAAATCTTAGAAATAA
ATGCAAAAATAATCGAACAATTATATCAATTACAGTTCAACGACGATCTTGTGAACTTTCTGACCGTCTTGCGGCATGCTGTCCAA
TTCGATGATGACGAGCCCCGGTTTGTTTCAATTCGTTCTGTGACTGTTCAAGAAGAAGATCATCATCATCTCATCTGAGGAAGAAG
GAATGGCGGCGAATCTGAGGCCTGTGTACGGATACACGGTGTTCTATGTGAAGGACGTGACGGAGTCGGTGGCCTTCTACGCCAAG
GCCTTCGGCTACCGAGTCCGCCGCGTCGACAACTCCAACAGGTGGGGAGAGCTGGAGAGCGGTGAGACGACGATAGCATTCACCCC
GATGCACCAGAAGGAGACCGACGGCCGGAGCGGCGTCGATCACATACCGGAGCCCGGGGAGAAGAACCCGGTGGAAGTGTGCTTGG
CATACGAGGACGTCGATGCGGCGTACAAACGGGCGGTGGAGAACGGTGCCGTGCCGGTGAGCCCGCCGGAGGACCAGGTATGGGGA

CAGAAGGTAGGCTACGTCCGTGACATCAACGGCATCACCGTCCGCCTCGGAAGCCACGTTAAGGAACCCACCCGCCAGTAGAAGCA
TATCCGATCTATGCGATAGCAACTTAATATGTCCATGTATGTGTAGATGTGAATAAAATGATCTATATGAGCTGTTTGTGTTGATC
ATTTCTGGTTTTATGTGTAGCTCAATTTCCTAATAATGTCAGAGTTAATTTTGCGACTGTTT

LEFT PRIMER: CCGGTTTGTTTCAATTCGTT
RIGHT PRIMER: CCGTACACAGGCCTCAGATT
SEQUENCE SIZE: 922
PRODUCT SIZE 99

Contig617

GCACGAGGAACACTCAATTAATCAAGCGAAGCCATGGCTCTCGAGTGGATGGTTATCGGCAGCATAGCTGCAGCGGAAGCAGCTTT
CCTTCTGATACTGACGGTACCGGCGGCGTGGGGACTTAGCGGCGGCCTCCGGCGAGCTCTGACGAACGCCGCGAGGTCAGCGCTGC
GGCCGCTGATGGCTGTCGTTCCATTCTGTCTGTTTTTGATACTGGATATCTACTGGAAGTACGAGACTCGGCCAAACTGCGGGAGC
GGTAAGGACGGGACGATCGTGTGCTCCCCGGCGGACCATCTCCGGTTCCAAAAGTCGGCCCTCAAGAGTCAGCGGAACGCACTGCT
CGTGGCGTCGGCGCTGCTGCTCTACTGGTTCATCTTTGCAATCACGAAGATGCTCGCGCGCATCGATCTGCTCGCTCAGCGTGTCG
AAAATCTGAAGAAGCACGATGATTGAGATTTGTTATGCAATTTGGGTGTTTCGTTCGACGTGTGTATGTGAATCGAATTATAAAAT
ATGCCTAGATGGATCGTAATATGTCATTATAGAGCGCGCTTATGAGAATCTTTTATTTACTTTGTTTTCTTTTGGAGAAAACGCGA
GTACGAGCAAATTACATGCTGTTTTGATCTGGTATTTTTAGTTCTTACGGGGATTTTTAGCTAAAGTTTGCGGAAATATATGAACA
TTACGACGG

LEFT PRIMER: AAGTACGAGACTCGGCCAAA
RIGHT PRIMER: ACTCTTGAGGGCCGACTTTT
SEQUENCE SIZE: 697
INCLUDED REGION SIZE: 697
PRODUCT SIZE: 99

Contig693

GCACGAGGGATTTACGAATCAAGAGACTGCTTTTTTGTTGTTTAACTTGGAAAAGCAAGTGGAGCCTGTACGGTCACCAGACTCAC
CAGACTCAGAGGCACGAGGGGGACAGTGTACCCGGCGCCTATGGCGGTGCGCCGCCTGCTGCCGGCTACGGCGCACACGAGGGCCT
GCAACAACCACATGAGAACAAGGATCATAAGAGTCTTGGACAGAAGATGGGCGAGAAGCTACACCGCTCCAGCTCCAGCTCTAGCT
CCAGCTCTGAGAGTGATGGAGAAGGAGGAAGTGAAGAAGAAAGTGAAGGACAAGCTGAAGGAGAAGCTTCCCGGGAAGAAGCAAGG
AACCACCACCACCGGAACAACAGCTACTGGAACTACAGCCACCGGAACCACCACCACCACCGGCGCGCATGGGGAACAGAAAGGAA
TGATGGAGAAGATCAAGGACAAGCTGCCTGGCCACCACTAAATTGCTTTATTGTGTGTGTGCTTGATTGTGTGTGTCGTTAGTAAT
ATGAGGGGAACAAATAAGTGATTTTTGTTGTATTTCTCGTCTT
CCGTCGAAATTGTGGTGTGTGTCTTTATGTATAATAAATATATCTTTTATACATGAACTATTGGATGTAAATTAGGTATATAACAA
AAAACAATAAATTACTGGGTTTATACCACGCTTCCGACAAATTATAAAACAACAACCAATATA

LEFT PRIMER: CCACCGGAACAACAGCTACT
RIGHT PRIMER: GGCAGCTTGTCCTTGATCTT
SEQUENCE SIZE: 708
PRODUCT SIZE: 106

Contig701

GCACGAGGTTTTGGTTCAAGCAAACAAAGTTCAGAGAAGAAGAAGAAGAGATGGCCTCCACCCAGGATGACCTCAAATTCAAGGCC
GGCCAGACCAAAGGCACTGCTGAGGAGAAGGCTGGGCACGTCGGAGATAAGGCATCCGACGCTGCGTACTCTGCCAAGGACAAGAC
ATCGGGGACTACCCAGTCTGCACAGGACAAGAGCGCCGGGGCCATGGATTCCGCCAAGACGAAGGCCCAGGAAATGAAGGACAAGG
CCGGTGAATACATGGAGTCAGCCAAGGAGAAAGCTCAGGAGAAGAAGGACGAGACCGGCAGCTACTTGCAGGAGAAGAAGGAGAAC
ACCGGCGGCCACATCGACTGCGCGAAAGAGAAAGCCCATGAGAAGAAGGACGCCACCGGCGGCTACATGGGCGAGAAGAAGGGCCA
GACCGGGGAGTACATGCACGAGAAGAAGGACCAGAGCGGCGGCGTGCTCCAGCAGGCTGGAGACATGATCAAGGGCGCCACCGAGA
CGGTGAAAAGCAAGCTCGGGATGGCCGGCGACAAGGCCACCGATTCCACTACCACCACCAATCACTAGATCGATCCGTCGATGATC
TGAGTCCCATGTTCATCATGTAATACTCTCTGGTGTGGGTTGTGTTTTATTGTGTCCTTTCCATTATCTATTGCGACTGTAATTTG
CGGCGTTCTGGTTTATCTCGCTTGTGATCTGTGTTCCATTTACAGTTGTTAATAATACTCCTATATTATTATTATATCAATTTCAA
CATCGCGTATACTTCCAGG

LEFT PRIMER: GACAAGGCCGGTGAATACAT
RIGHT PRIMER: GCCGGTGTTCTCCTTCTTCT
PRODUCT SIZE: 99

INCLUDED REGION SIZE: 793

Contig766

GCACGAGGAGAGAGCTCTGTAGAAGAAGATTGTTAGAAGCTTTTGTCAGCCATGGGGATGCTCGGGAGAACCTTGGCGGCGCCATT
GTTGTTCTTGAACTTGGCCATGTACATCGTGGTCATTGGATTCGCCAGTTGGGCTCTTAACCACTTCATTAATGGCCAAGCTAATC
ATCCTGGTTTGGCGGGGACGAACGGTGCGACGTTCTACTTCCTGGTTTTTGCGATCCTAGCCGGAGTAGTTGGTGTGGTGTCGAAG
CTGGCCGGTGCTAGCCACCTCCGAGCTTGGAGGAGCGATAGCCTCGCCGGCGCTGCCTCTGCCTCGCTCATTGCTTGGGCTATCAC
TGCCTTGGCCATGGGATTGGCATGCAAGCAGATCCACCTGAGTGATCACAGAAGCTGGAGGCTAACGGTTCTAGAAGCCTTTATTA
TCATCTTGACATTCACTGAGCTTCTCTACCTGCTTCTGCTCCACGCTGGAATCTTCAGCAGCAAATATGGACCCGGATACAGAGAC
CCAGGATACGCCGGAGCCGGCACCGGCGTCGGCGATCCCGTCAAAGGAACCGCCAGAGTATAATATGGGGATCGAATTTAACCAAG
GAGTGAATAAATAATATGGCTGGCTTCTCTCTTCCGGCCATGAATAGTGTGGGATTTGTTGATGGTTTTTATCCATGGTTTTGTAA
AGTTGATCATCAATCTTCTCTTGTCGTTTAAGGTCGTATGGATTTCTTTTGTTAGGTAGCGTATGTTATAAAACTGTGGATTATGA
ATCTCTGTCCTTCTTTGTAATCTTTTTTATGGCTGGTTGTGATGACTTATCAAAATATGTTTTAACAAATTATAATC

LEFT PRIMER: TCAAAGGAACCGCCAGAGTA
RIGHT PRIMER: CTATTCATGGCCGGAAGAGA
PRODUCT SIZE: 93
SEQUENCE SIZE: 851

Contig859

GATCTTTCAGATAAAAGCTAGTAGTTTTAAGTTCAGTTTTCAAGACATAGAGATGGAGGGTTTTGGGAGCCAGCAGCACGACCAGC
ACCGACACCAGCAGGGCACCGACCAGTTCGGCTCCCATGTCCAGCCCGGGCACGGTGGTCAGCAGGGTGTACTCGGCGGGCAGCAG
CAACACCAGCAGCACAAGGACCAGAGTCAGGGTATTGGTTCCGGCATTAGCAGCAAGCTTCACCGCTCCAACAGTTCCAGCTCCAG
CTCCAGCTCTGAAAGTGATGGAGAAGGAGGGAGGAGGAAGAAGGGTATTAAGGACAAGATCAAGGAGAAACTGCCAGGGCAACACA
ACCAAGGACAGACCGGTCAGCATGGCATGACTGGCGGCCATCAGCAGGGCATGACCGGCATGACTGGCGGCCATCAGCAGTGGCTA
CGGAGCCACCGACCAGCATGGAGAGCAAGAGGGAATGATGGATAAGATCAAGGACAAGCTTTCCGGCAATCAGTAAACCTAAATAC
CTCCAGAATTGCATGATGAGACGCATAAATATATATATTTATG
TGTATATAAGAATATGTTTGCTGTGTTTCGGTGCTGCTAAGCACCTGTGTTGATCGGTGTGTAAGTTAAGTGTGTGTGCGTCGTGA
ATAATTCATGTGTAGCAGTGAATACACATGAACGCTATGGTTCATCTTTTATCGTACTTGAATGCGAAGTCTTTTGAGTTTTTTCT
TAAAAAAG

LEFT PRIMER: GGTGGTCAGCAGGGTGTACT
RIGHT PRIMER: CGGTGAAGCTTGCTGCTAAT
SEQUENCE SIZE: 739
PRODUCT SIZE: 98

Contig915

GCACGAGGCGAAGACAGGTTGAGCTCAAGCTACCGGAGCTAAGAAGAGCGTAAAGGAGGTGTCTTGAAGAAGCGGAGTGTTAGAAG
CCTTTTGCAGCCATGGGAATGCTCGGGAGGACCTTGGCGGCGCCATTGTTGTTCTTGAACTTGGCCATGTACATCGTCGTCATCGG
GTTCGCCAGTTGGGCTCTGAACCACTTCATTAATGGCCAAGCTAACCACCCTGGTTTGGCAGGGACAAACGGTGCAACATTCTACT
TCCTGGTGTTTGCTATCCTTGCCGGAGTAGTGGGCGTGGCGTCGAAGCTCGCTGGAGCCAGTCACCTCCGAGCTTGGAGGGGCGAC
AGCCTCGCCGGCGCTGCCTCTGCCTCGCTCATCGCTTGGGCTATCACCGCCTTGGCCAATGGGATTGGCTTGCAAGCAGATTCATC
TTAGTGACCACAGAAGCTGGAGGCTGAAGGTTCTGGAGGCCTTCATCATCATCTTAACGGTTCACCGAGCTTCTCTATCTGCTTTT
GCTCCACGCCGGAATCTTCAGCAGCAGATACGGACCCGGATACAGAGACCATGGATACGCCGGCGAGCCTGTTAAAGGAACCGCTA
GAGTATAACAAACAACAATGGAATTTAACCGAGAAGTGAATAAATAAAGATGGTTGGCTTCTCTCTTCTAGTCATGAATGGTGTGA
GATTTGTTGGTGGTGTTTATCCATATGGATCTGTAAATTAGTAGCTCTACTTTGATCATAGTTGGCGAGTCTTCGAAGTATTTGTA
TGCTATGGATTTCTTTTGCTAGTGCATGTCAAAAAATTGTGGATTATGAATTTATGTTTTGC

LEFT PRIMER: TAACGGTTCACCGAGCTTCT
RIGHT PRIMER: GGCGTATCCATGGTCTCTGT
SEQUENCE SIZE: 836
PRODUCT SIZE: 93

Contig959

CGAGGNCCATAACACCTTTCCGGGGGCATATAAACGCGGAAAANGTTTCCGATGGAAACGGGCATGTGGGAAACCATAGGGGGTGA
GCTCCATTCATGTGGGCACCCGCGGCTTAAAATTTAAGCGCTCCCGGCCATNGTGTTGGGGGAATGGGACGGAAACAATTTTCACA
CAGGGAACANGCTTAGGCCCCAGTTGATTTACCGCCCAAGGCTTCGAAATTAACCCTCATAAAAGGGGAAANCAAAAAGCTGGAGG
TCCACCCGGTGGCGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCACGAGGGCTCAGCAAAAGGAAATTGGAAA

TTTCATCGTTTGCTTTCTAAGATGGAGGGCTCCGGGAACCAAGACCAGCACTACCGCACCAGCGAGCACGCTGCTCCTGGCCAGGG
CGTGCACCCTAGCCAGCACGGAAAAGGCACCAGCGAGTTCGCCGCTACCGGCCAGGGTATGTTTGGCGGCCAGCATCACGACCAGA
ACAAGCATCAAGGACATGGAACTGCTCACGAGAGTCATGGAGAAGGGAAGAAGGAGGGAATTACGGAGAAGATCAAGGAGAAACTC
CCAGGACAGCACCACCAAGAACCCACCGGCAACCAGGGTTTGACTCACAGCCAGCAAGGCCATGGAGCCACAACCAAGGAGACACT
CCTCCCAGGACAGCACCACCAAGAAGCCACCGGCAACCAGGGTTTCACGCACAACAAGCAAGGCCATGGAGCCACAACCAAGGACA
CACTCCTGTGAATGACCTATCCTCCATCCATCATATGATATATATTTAGGCTTTCAATCCGGTGCTGCTACGTTCCCGTGGTGTTC
TTTTTTTTTTTTTTTTTTTTTGTTTTCTATTTTTTAATTTTTGTGAACTGTGTGTTTGCGTGCGAATAAGTCATGTGTGTCCGTGA
ATGCACATGAACGCTTTCTAGATATCCATCAATAAAGAGTGATTACATAGTGTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

LEFT PRIMER: AGGACAGCACCACCAAGAAG
RIGHT PRIMER: CACCGGATTGAAAGCCTAAA
SEQUENCE SIZE: 1032
PRODUCT SIZE: 144

Contig989

GCACGAGGCGCCGCTCTGACAGCCGGGATAATCATCATGGGATTTATCGTTTCGGGAGGGCTGGGTGTAGCGGCGCTGTCTGTGTT
GGCTTGGATGTACAGGTACCTGATGGGACAGCACCCAGCCGGAGCCGATCAGATCGACCAGGCTCGCCAGAAGATCGCGCAGAAGG
CTCGTGAATACAAAGAGACAGCGCAGCAACGCATCAACGAGGTCTCCCAGTAAACATCTAGAGAGAGTTTCAGATGCTACCAGCTC
TCTTATTATCTTGTCATGTGTATTACTGGTGTATTTTGTATGGCATATTAGGGTTTTGGAGTATGTAGGAGTATGTAAGTCAGGGT
GTCTTGGTGATCGGTACGAATAAGTTTGCATC

LEFT PRIMER: GTCTGTGTTGGCTTGGATGTA
RIGHT PRIMER: TTGTATTCACGAGCCTTCTGC
PRODUCT SIZE: 108

Cluster 5

Contig1561

CCCAGGAGGCGAGCGCGTACCCTTCCTCTTCACAGTCAAGCAGCTCGTCGCCTCCGGCAAGCCAGAGAGCTTCGGTGGACAGTTCC
TCGTTCCCAGCTACAGGGGCTCCTCCTTCCTCGACCCTAGGGCCGTGTTGGCTCAGCTGGATACGATAACGCAGTGGCGCTGCCGG
CAGGTGGAAGAGGAGATGAGGAGGAGCTCGTCAAGGAGAACATCAAGGACGCCTCCTCCTCCACCGGAAGCATCACCCTCAGCGTG
ACCAACAGCAAGCCAGAGGCTGGCGAGATCATCGGAGTGTTCGAGAGCATCCAGCCGTCTGATACGGATCTCGGAGCAAAGGCGCC
TAAGGATGTGAAGATCCAGGGAATCTGGTTGCTCAACTTGAGCAGTAGCGAGAACGGTGAAGCACTTTGAAGATAGATTTTTCTTC
TTGTTCTAAAAGAAAAAGGGGCTGGCGTTGGGTTGTGGATGGTTAGAATTCATTGCGCATTTCTCATCCTTGATGTATGTCTTTTC
GTTTCTTCAAACAAATATATTGTAATGTAATATAATTACTTGCCTAAAATTAATCTCATATTCTACCCTTAAAAAAAT

LEFT PRIMER: CGAGAACGGTGAAGCACTTT
RIGHT PRIMER: TGCGCAATGAATTCTAACCA
PRODUCT SIZE: 97
SEQUENCE SIZE: 594

Contig348

GCACGAGGCCGGCGCTCAACCAGCTCAATAGAATGAGGAGCAACAAGGGGATCATCGTAGGTCTTGGGATAGGTAGCGGGCTCGTG
GCAGGTGGGTTCTTGTCGACTCCTAGTGCATCAGCCGGGAGATTAGCGTGTTGGCTGAGGCTGCTTCAAACGATAACAGGGGTCTT
CTGCTCTTGTTTGTGTTGCACCAGCTATTGCGTGGGTACTCTACAACATCCTCCAGCCAGCTTTGAACCAAATCAACAGGATGAGA
TCCAATTAATATGGGACCCTTCTGCTTGAGTATAACTATCGTCTTTGATCTTTGCTCTTAATTATAGATTTGTTGCTACATATAAA
GCGGTTGTTACGTCTGCAAGAACTTATGATATTCCTTCTATGGGGTTGTATGTTGTCAGGAACTTTTATTGAACTGGAAGTTTCCA
TCACAAAG

LEFT PRIMER: CCAGCTATTGCGTGGGTACT
RIGHT PRIMER: TCAAGCAGAAGGGTCCCATA
PRODUCT SIZE: 94
SEQUENCE SIZE: 438

Contig498

TCTCTCAAAGCTTTCTCCGCCGCCTTGGCTCTCTCCTCCATCCTCGTCTCCTCCGCCGCCGCCCCGCCGCCTGCCCTTGCTGACAT
CTCCGGGCTCACTCCATGCAAGGAGTCCAAACAATTCGCCAAGCGCGAGAAACAAGCACTGAAGAAGCTCGAGTCCTCGCTTAAAA

ACTACGCGCCGGACTCCGCGCCTGCGTTGGCCATCAAGGCCACCATGGAGAAGACGAAGAAGAGGTTCGACAACTACGGGAAGTAC
GGGCTGCTGTGCGGTACCGACGGGCTGCCCCACCTCATCGTCAACGGTGATCAGAGGCACTGGGGAGAGTTCATCACCCCTGGGCT
TCTTTTCCTTTACATCGCTGGGTGGATCGGATGGGTGGGGCGTAGCTACCTCATCGCCGTTAGAGATGAGAAGAAGCCTACGCAGA
AAGAGATCATCATCGATGTCCCGCTTGCTTCTCGTCTTCTCTTCCGTGGTTTCATTTGGCCGGTCGCTGCTTACCGTGAGTTCCTC
AATGGTGAGCTCGTCGCCAAGGACGTATGAGCGAGCTGACGTGGAGGGTTGGAAAGGATTAAGAAGGAGGGGATCTCATTTTGTAA
CTTCAAACCAATTTCTACGTTTAATGTGGTGTAATTAAACATATTTGTGGAGCAATTAATTAATCATAAGTACTCTTAAA

LEFT PRIMER: GCTACCTCATCGCCGTTAGA
RIGHT PRIMER: AACCACGGAAGAGAAGACGA
PRODUCT SIZE: 93
SEQUENCE SIZE: 682

Contig591

ACGCGGGGAAGAGCAGAGTGAGGCTGAAAGACACGAACATCGGGAAGTAGGTCCGGAAGCTCTAGCAAACAATTCAATGGCAGCCA
GTATGATGTCTTCTTTGGCACTCAAACCGGCTACTCCGGTGTTCGAGAGGTCGTCGGTGAAAGGACTTCCATCTCTGGTCAGATCA
TCTTCCTCCTTCAAAGTTCAAGCTAGCGGTGGCAAGAAAATCAAGACTGCAACTCCCTATGGTCCTGGAGGAGGGATGAACTTGCC
AGCAGGTCTGGATGCATCTGGAAGAAAGCAGAAGGGAAAGGGTGTCTACCAATTTGTGGACAAATATGGTGCAAATGTTGATGGCT
ACAGCCCGATCTACTCACCAGACGAATGGTCTCCTAGTGGTGATGTTTATGTTGGAGGAACCACTGGTCTCATTGTCTGGGCAATC
ACTCTTGCTGGACTTCTGGGTGGAGGTGCTCTTCTTGTGTACAGCACCAGTGCTTTGGCTCAGTAAATCAACCAATTTCAAAACTC
TTGTACTCAAACTGCTGCTATACATGTCCTCCATCTTACTTTTTCTATATTCAAGAATTACTTTTTCTATGTAGTTTGTGAGTTTG
TTCTTACTTGTGATGAATCAGTAAACTA
TATGGGCACTGTCTGTGTATTTACTCTGTAAATCCACTTGATAATCATGTTCGTTGGTTTTTTTTTTT

LEFT PRIMER: GATGGCTACAGCCCGATCTA
RIGHT PRIMER: CAAGAGTGATTGCCCAGACA
PRODUCT SIZE: 99
SEQUENCE SIZE: 698

Contig656

GCACGAGGCTCCTCCCCAGGTCCCCGTCGATGTTAACGCTACCGCCTCACCGGAAAGCTTGCAGTGCCCAACACCGGCTACTCTAA
GCGTGCCTACGTGACTTTTCTCGCCGGCAGTGGCGACTACTGGAAGGGAGTTGTCGGGTTGGCCAAGGGGCTCAGGAAGGCTAAGA
GCGCGTACCCACTGATCGTTGCGATGCTTCCAGATGTTCCGGAAGAGCACCGCGAAATTCTTCGAAGCCAAGGGTGTATCGTCAGA
GAGATCGAACCAGTCTACCCGCCCGAGAACCAGATTCAGTTCGCCATGGCTTACTATGTTATCAACTACTCCAAGCTTCGCATCTG
GAACTTTGAGGAATACAGCAAGATGATTTACCTGGACGCGGACATCCAAGTGTACGATAACATCGACCACCTCTTCGACATGCCGG
ACGGGTATTTCTACGCAGTGATGGACTGCTTCTGTGAGAAAACATGGAGCCATTCGCGCCAGTTCTCCATCGGGTACTGCCAGCAG
TGCCCTGACAAGGTCAAGTGGCCCATGGAACTGGGAAACCCTCCTGCACTTTACTTCAACGCCGGCATGTTCGTTTACGAACCGAG
CCGCCTCACTTGCGATAGCCTCCTTGAGACCCTCCGCATCACGCCACCAACCCCCTTCGCCGAACAGGATTTCTTGAACATGTTCT
TTGAGAAGAACTACAAGCCCATCCCTCTTGATTACAATCTTGTTTTGGCCATGCTGTGGCGCCATCCCGAGAACGTCGAGCTCGAA
AAGGTCAAAGTTGTTCACTACTGTGCTGCAGGTTCAAAGCCATGGAGGTACACTGGAAAAGAAGATAACATGGACAGAGAGGACAT
CAAGATGCTAGTGGCTAAATGGTGGGACATCTACAACGACGAATCACTCGACTTCAAGCCAGAGGATGTAGTGCCCGAGGGCGACG
CCTTGACGAAGACTCCAATCATGGCAACCCTGCCGGAGCCGGCCATTTCCTTTATCCCTGCACCATCTGCAGCCTAAGAAGAAGAA
ATGTCTACATGATCGAGTGCATAGGGAAAAATAACTTTATCTACTATATAGATGGCTATTTTGTGCTCCACATGGCTAGCATCGTA
TTGCTAAGACTTGCCGCAGTAGTTATGTTGGATCATTTTCAATGAGTTAAAAAAGTGCTATGCATTATATAAAAAGTATTTCGAAG
C

LEFT PRIMER: AGCGTGCCTACGTGACTTTT
RIGHT PRIMER: CAGTGGGTACGCGCTCTTAG
PRODUCT SIZE: 101
SEQUENCE SIZE 1205

Contig727

AGTGAACAAGGAAAAGTCATGGAAGCTAAGGTTGTTTGTGATAGGGACAGCGGAAAGATCAAGAAGTTTTGGGCTTTGTCACTTAC
AGTTTCTACTAAGGAGGCAAACGATGCTATTTCATCTCTCAATGGCTTTGACTTTGATGGCAGACCCATTAGAGTAACCATGGCAG
AATCAAGGCCAAGACGCCAATTCTGATGATGGATCATGCTTCTATCATAACCAATCAAGAGTAGAGCTATTTGTCTTGTCTTCAAT
TCTTAATTAGGTTTTTTTATATGCATTGTAACTCTCGTCGAACTTCATTTCCACTCTGGTCTCCCGGTGCATGGAGACAAAACATT
TTGCCCAAGAATTTCTTAGCAGATCCAGCAGTCTCTCTTTCTGCATTAGGCTGTATTGTTCTCTTTCAGTACTTTATAGATGGGAC
TATAGGTGTCATAGATTGCTCGGAGACGTCCGACTGTATCAATGATTTTCTCTTAAGAATTTGAATGAATATGAAACCTCC

LEFT PRIMER: GGACAGCGGAAAGATCAAGA
RIGHT PRIMER: ATGGGTCTGCCATCAAAGTC
PRODUCT SIZE: 110
SEQUENCE SIZE: 511

Housekeeping genes

Contig1360

GAAAAGTTGCTCTCCAGCGATCTCTGGGCTGCCAGGATTGAAAGGTTTGTTCTTTTGTATTTATGAAATGAAATGGATGCTAATCG
TAAAACCCGAATTAAGTCGATCTTAGGACTGTTTGGAAAGTGCATTGTTTACTATGTACCATAAGTACGATTGAGTGTGTGTACTG
GAGTATTACTTGTGCTAAATGGTGTCGCATATTCACAACGAATTATTTTGTGGAGCTTTATAGTTTTATTCATCTTGACCATCTGG
GAATTTCTCCAAGTTGATGACATCGAAGCATTTTGCGCTAGGTTTCTGTAACTTATTGCTT
ACAGGCCAATTTTCTTACTCTTTGCCAGGTTTTGGAAGGAACCAGACAGTAGATACATGCTGAGGTGTAGGTGGTATATTATTCCT
GAAGAGACAGCAGCGGGAAGGCAACCGCATAATCTACGAAGAGAGCTCTATCTTACCAATGATTTTGGCGAGTTTGAGGTGTCAAA
CAAATATTTTTAATTCTTTTATTTGCATTGGTCATCTGATTTTATCACAAAGCACTCTGCTGTTATGTGATACTCGATGATTGTAT
GCCGTCGCCTTTATAATGTTAGATGTGTATTGATAGAAAGGAACGTTCAACTTCAAACAAGATGGGTTTCTTAGGTG

LEFT PRIMER: CCAATGATTTTGGCGAGTTT
RIGHT PRIMER: ATAAAGGCGACGGCATACAA
PRODUCT SIZE: 131
SEQUENCE SIZE: 654

Contig1367

AGAAATCCAACCAATTGAAGAACTCGGTTTCACACTACCCAGGAACACAGGCACAATCTCGAGCTTAAGAATTATGCTCTGTCCGT
GCAACTCAGGCATGCACAGCAGAATAATTTTTTTCCAGGTAGGTTCAATCCAGATGTATGCTGATTGTAATGTCTGGAGCAACTCT
CAAGTCAGGAACGCAGCATGTCAGAGCCATATCGAGATGAGTGATTCTGAGGAATCATTGTGTATCAATAAACTCTTAAGATTAAA
AGAGAGTCCTGGTGGAGCAGATTTTTGCTGCGATTGGCTTGGTTATTTCTTCCTCTGCGACAATGGTTGTGTTCCAGAATTTTGAT
GTCTACCCTCCAACTATGTACAGTTACTCTTTTGTAAAAAATTTCAGTAAGTTATCTACAGTTATTTTCT

LEFT PRIMER: AAGTCAGGAACGCAGCATGT
RIGHT PRIMER: TGCTCCACCAGGACTCTCTT
SEQUENCE SIZE: 414
PRODUCT SIZE: 104

Contig1463

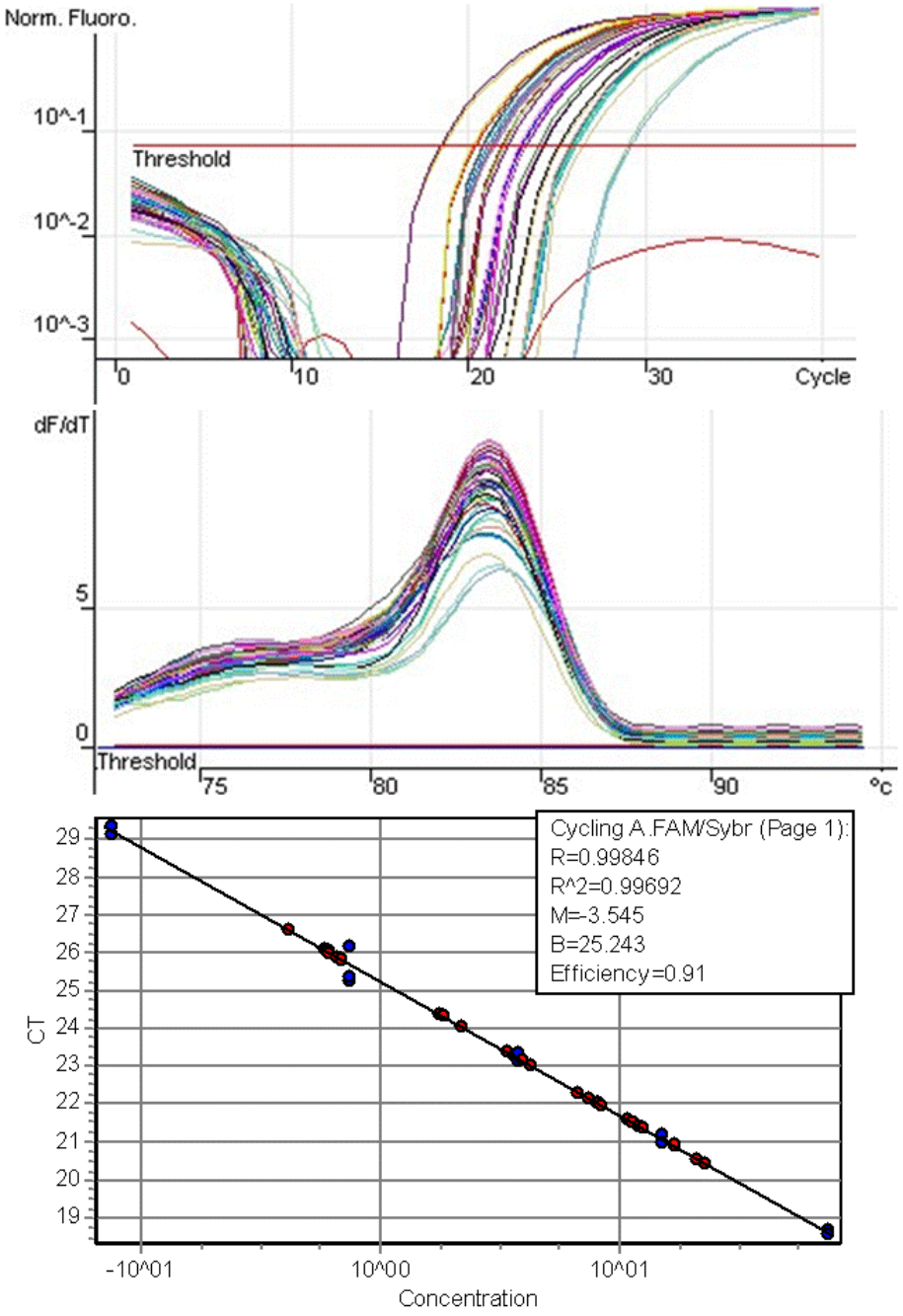
GCACGAGGCGAAGATCTGAAAAGATGACATCATCGGTTGCTATCGACACAAGTGCTGCTCCTTTCTATGTTTCCCGAGAGGCGCTA
GTGGATCAATCTGGAAATAGAACCGTTATGGAGTCTGCTGTGTTGGGATCCAACTCTACATACGATGGAGTGCATAGAGAGGTTGG
TAGTATTCCCTTTGGTATGTCAGGGATGTACTAGTCTTGGAATAGGACACTGTAACGATGTTCTATCGGTTCGTCACTCATGGTGT
ACAGAACCTTAGTATTTGGAAATGCCGAAGAGAAAACTCCATTAAAGCTCAAAGTATGTCGACTTTACTGGCTCCATGATTTTGCG
TTTTATGTAGACGCGAGCTGTAAATAGACGCTGGTTAGTTTGTTTTCGCCTTTGCTCGACTAAGAGATTACCTTGTTTTTCTCCCG
CTAGTAACGATGCTTTTACTATTGTCTCTAGTTGGATGTATTTATGAGATGATGCAAGAAGGTTTTTTTTTGTCAAAA

LEFT PRIMER: CGGTTGCTATCGACACAAGT
RIGHT PRIMER: GGATCCCAACACAGCAGACT
PRODUCT SIZE: 104
SEQUENCE SIZE: 508

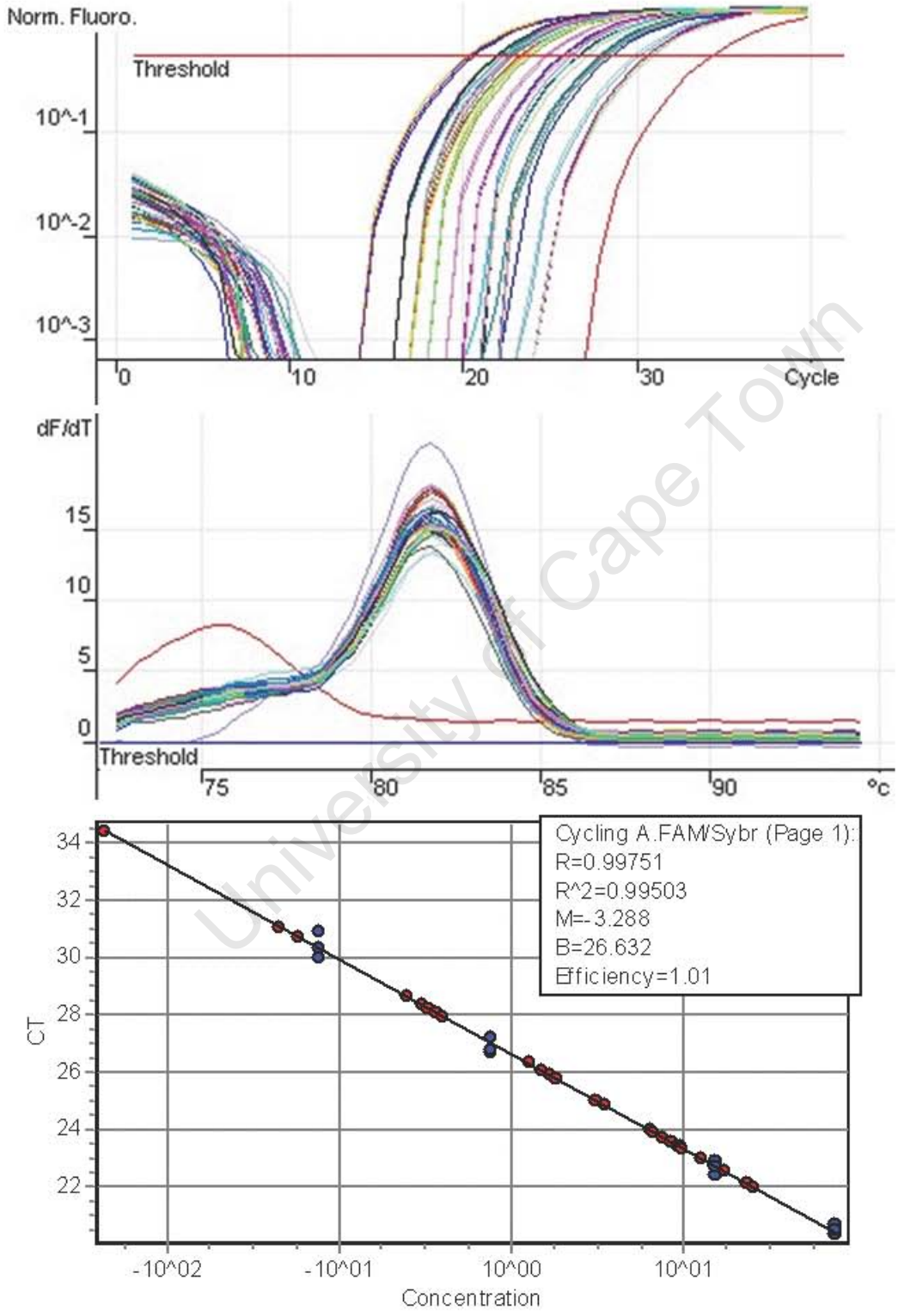
A.4.2

Amplification plots, melting curves and standard curve plots of samples of known dilution (blue dots) and unknown samples (red dots) vs the Ct values taken from the amplification plots. Data is presented in numerical order of contig number.

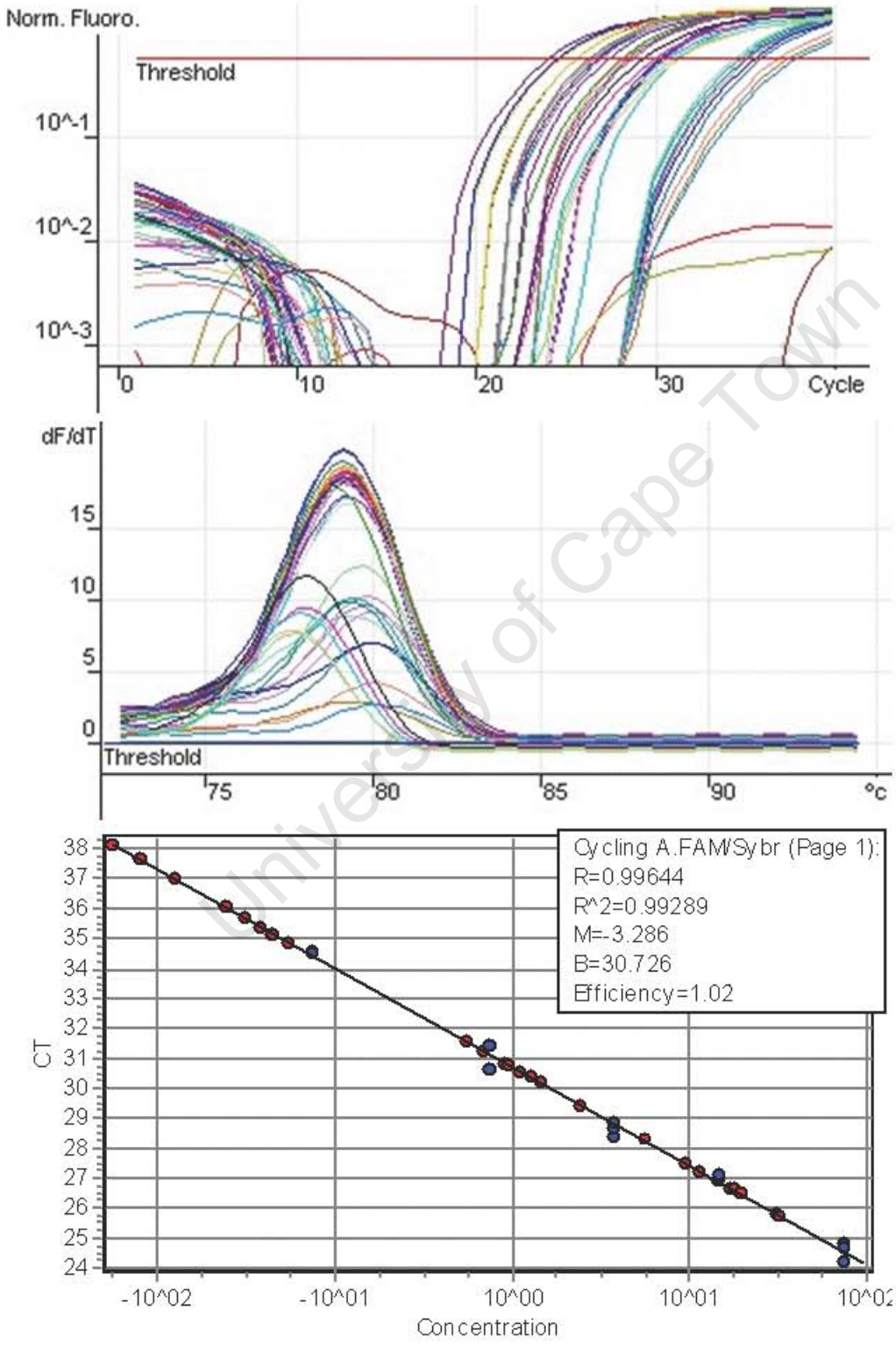
A.4.2.1.



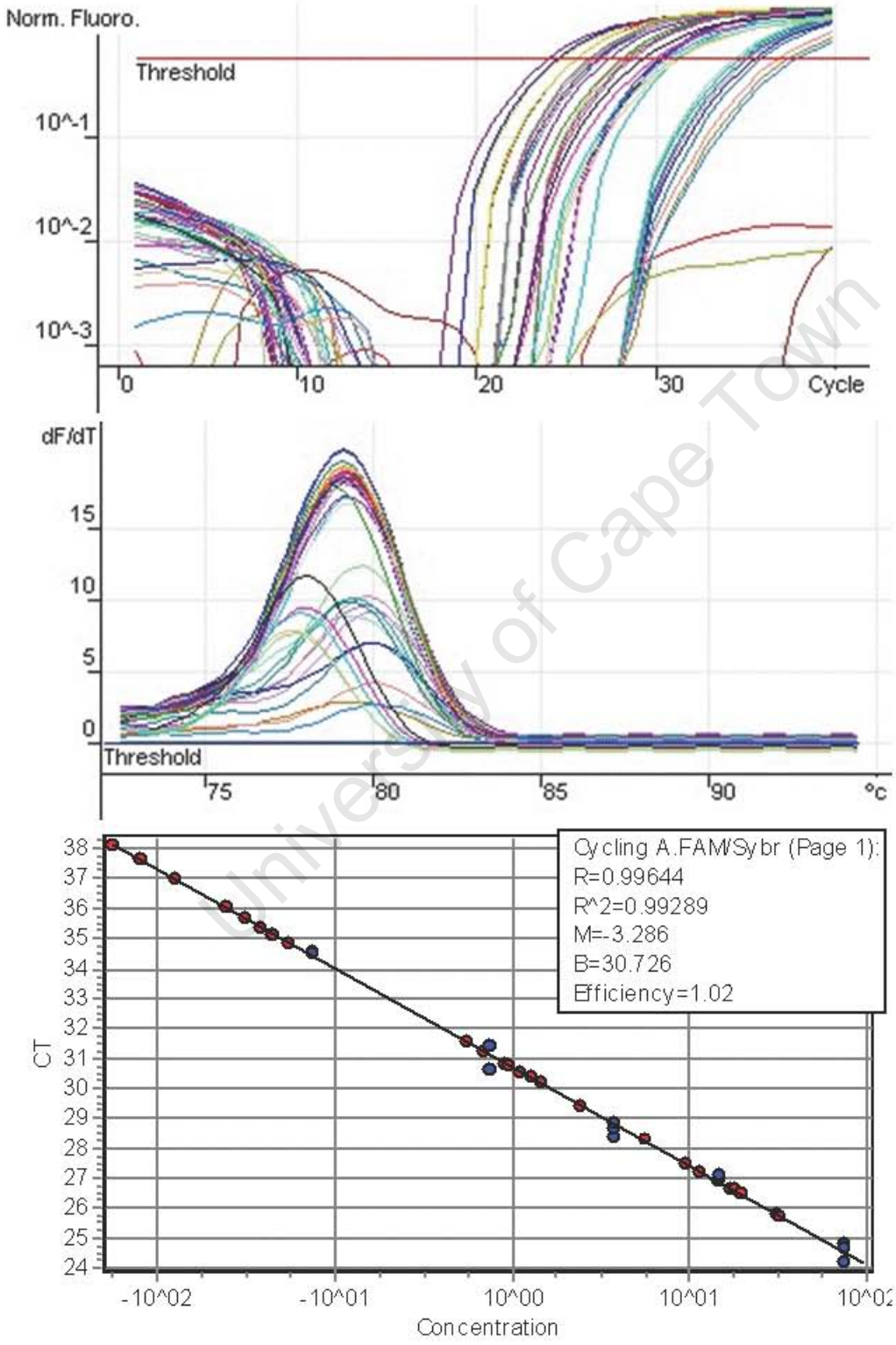
A.4.2.2.



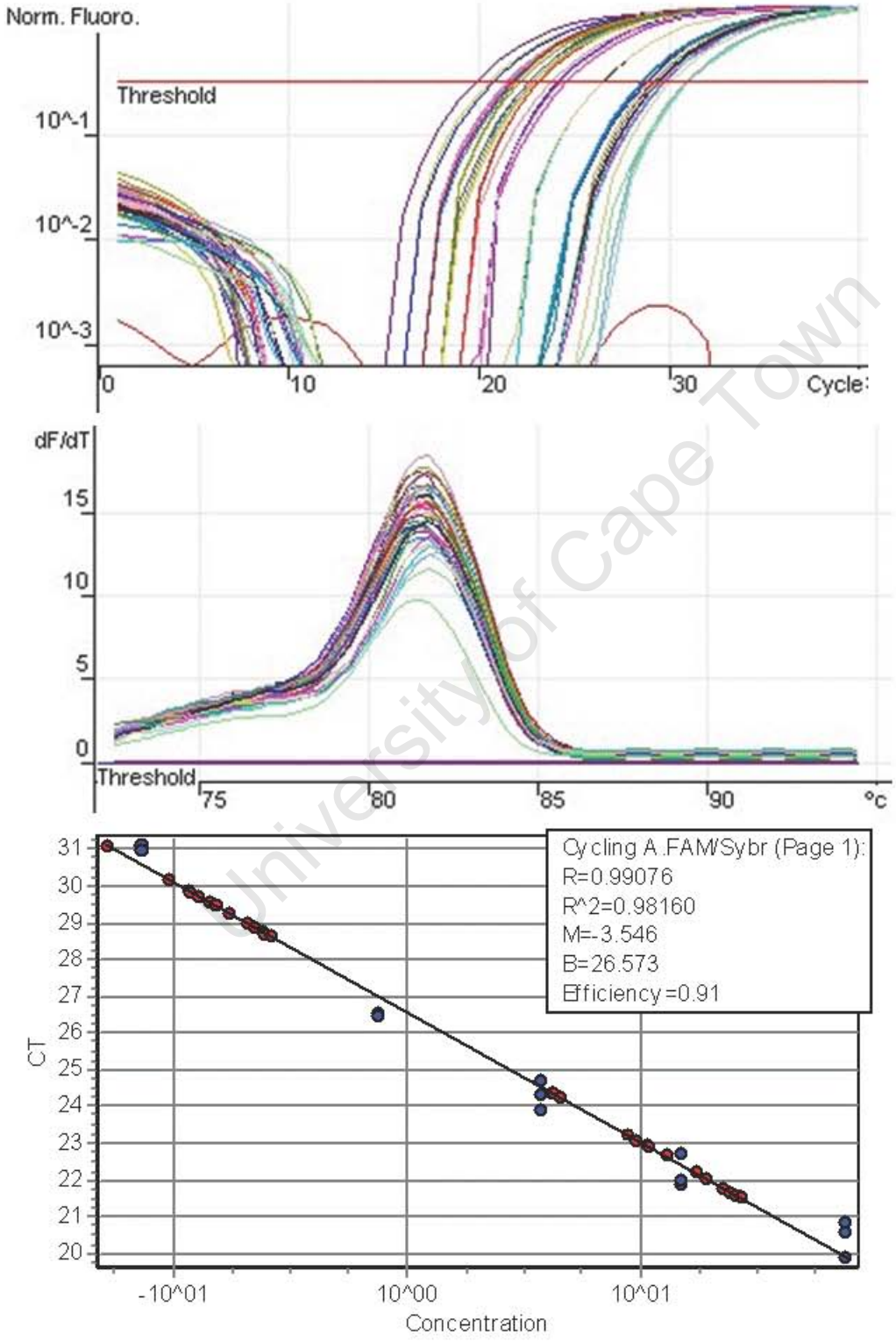
A.4.2.3.



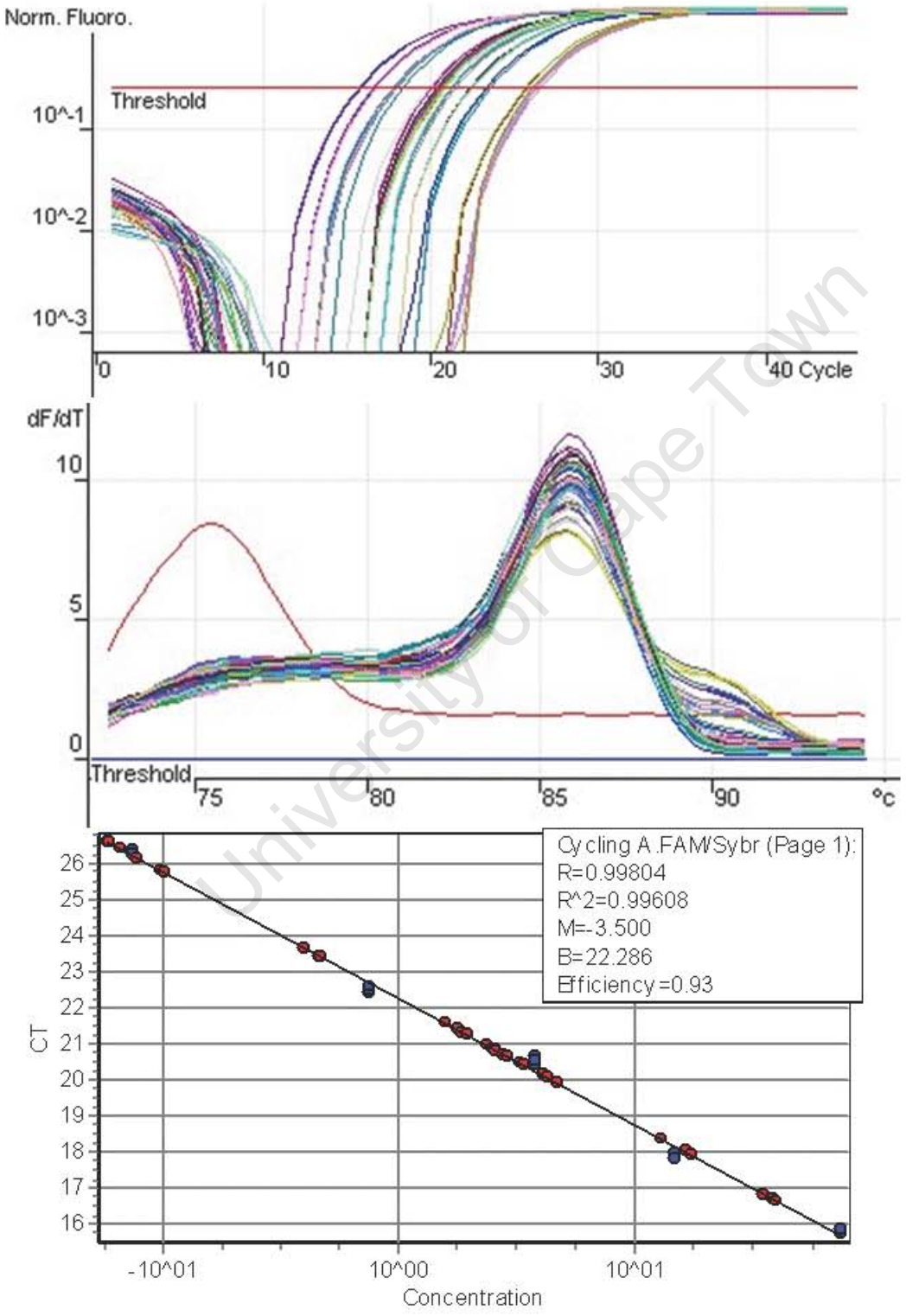
A.4.2.4.



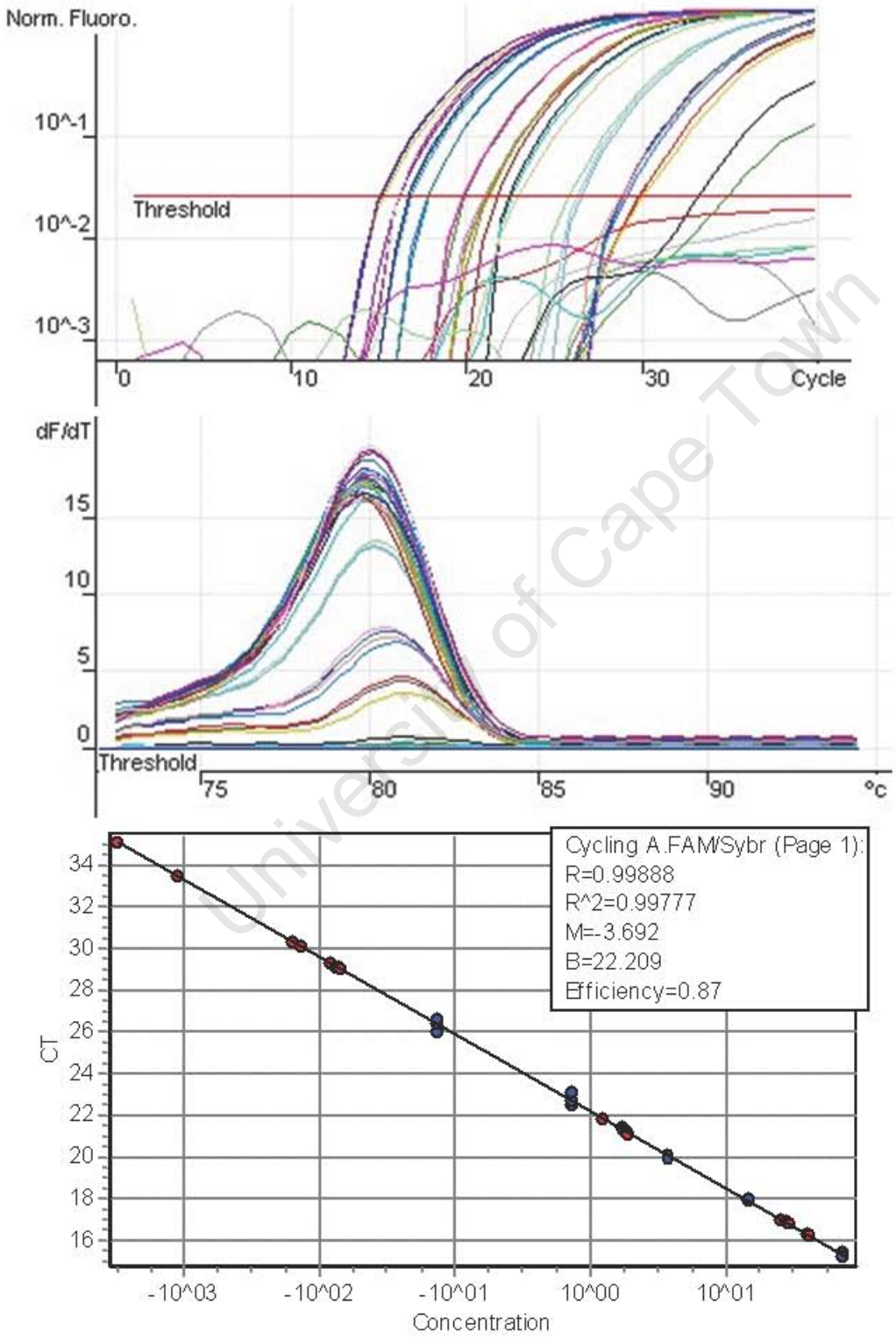
A.4.2.5.



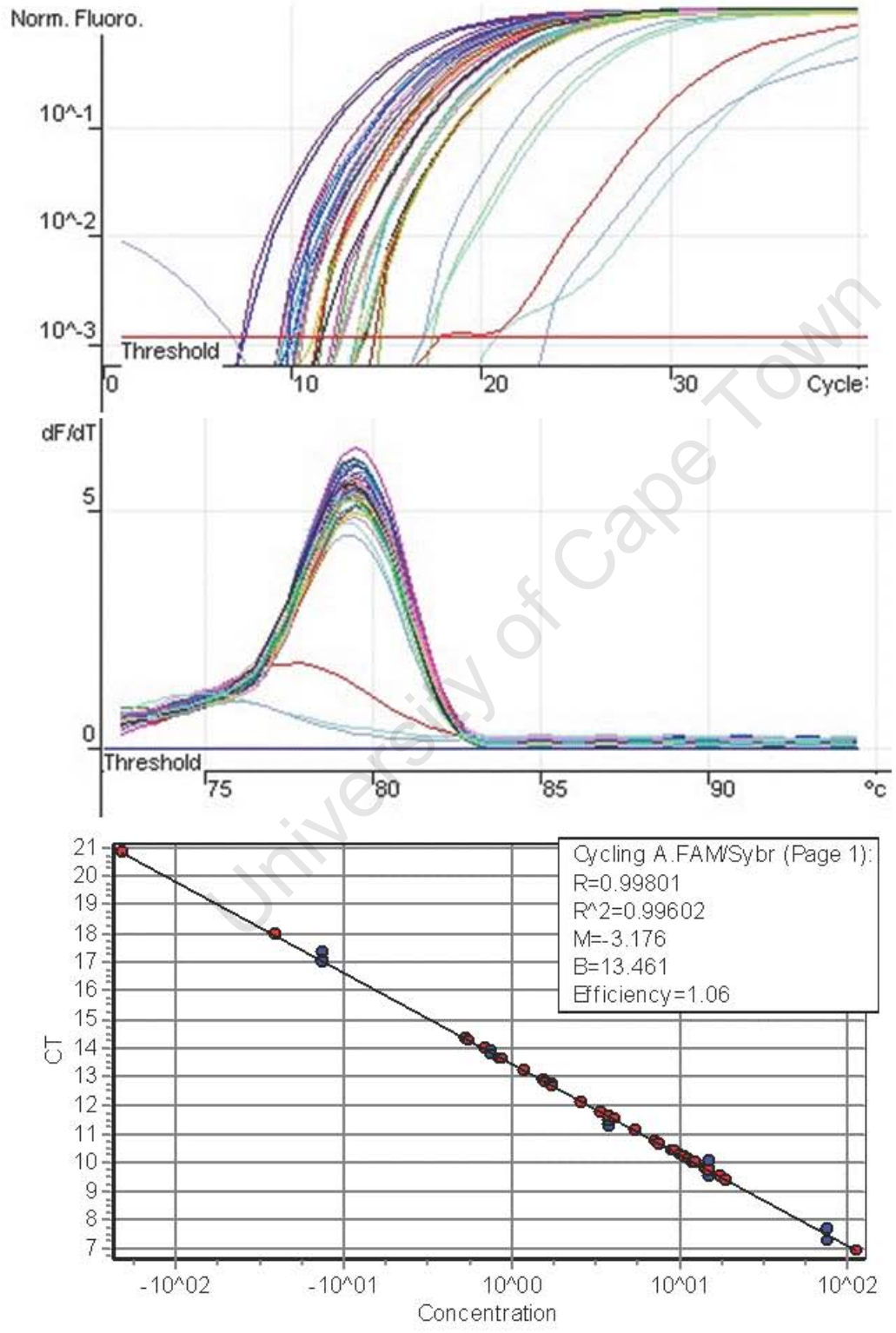
A.4.2.6.



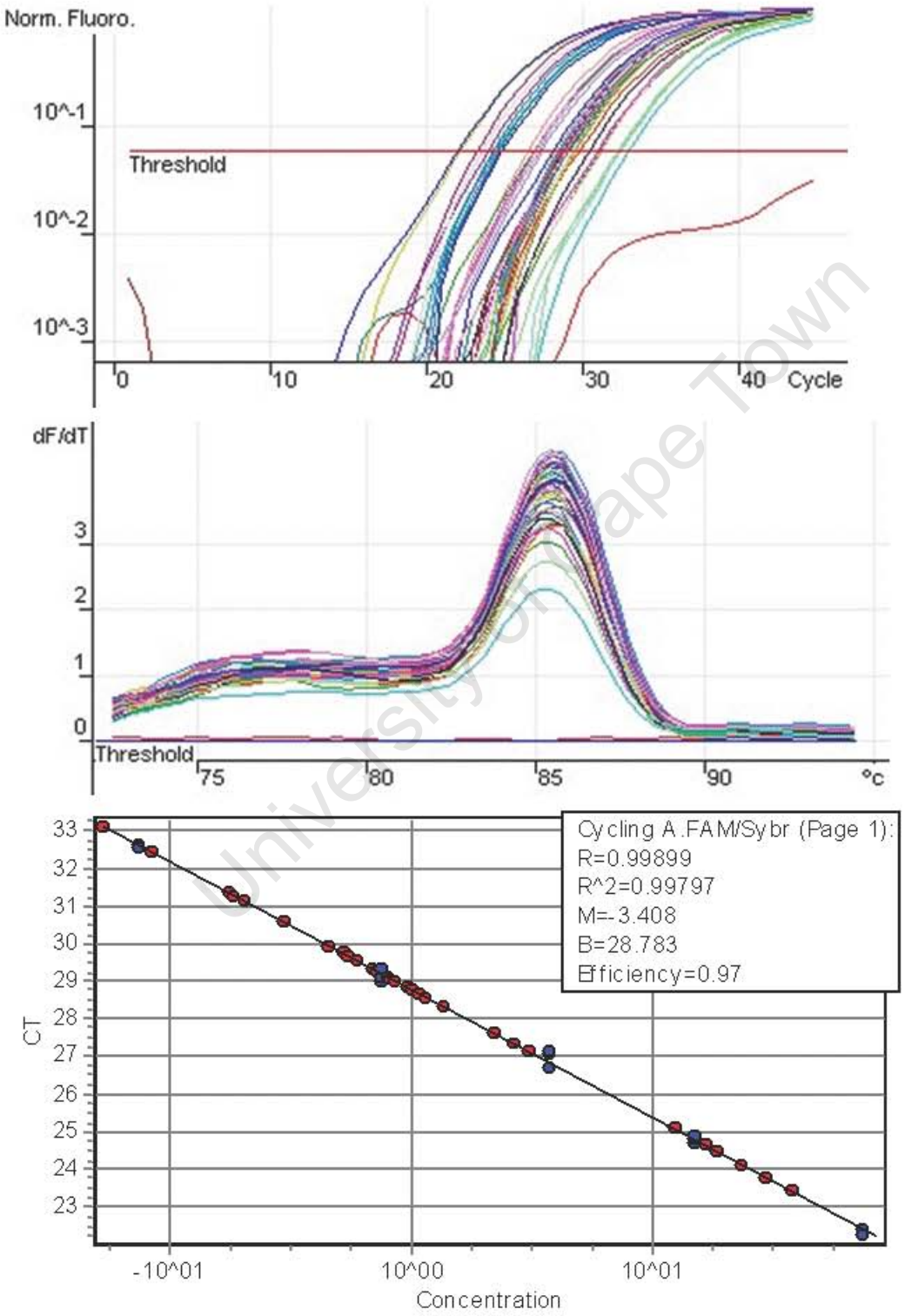
A.4.2.7.



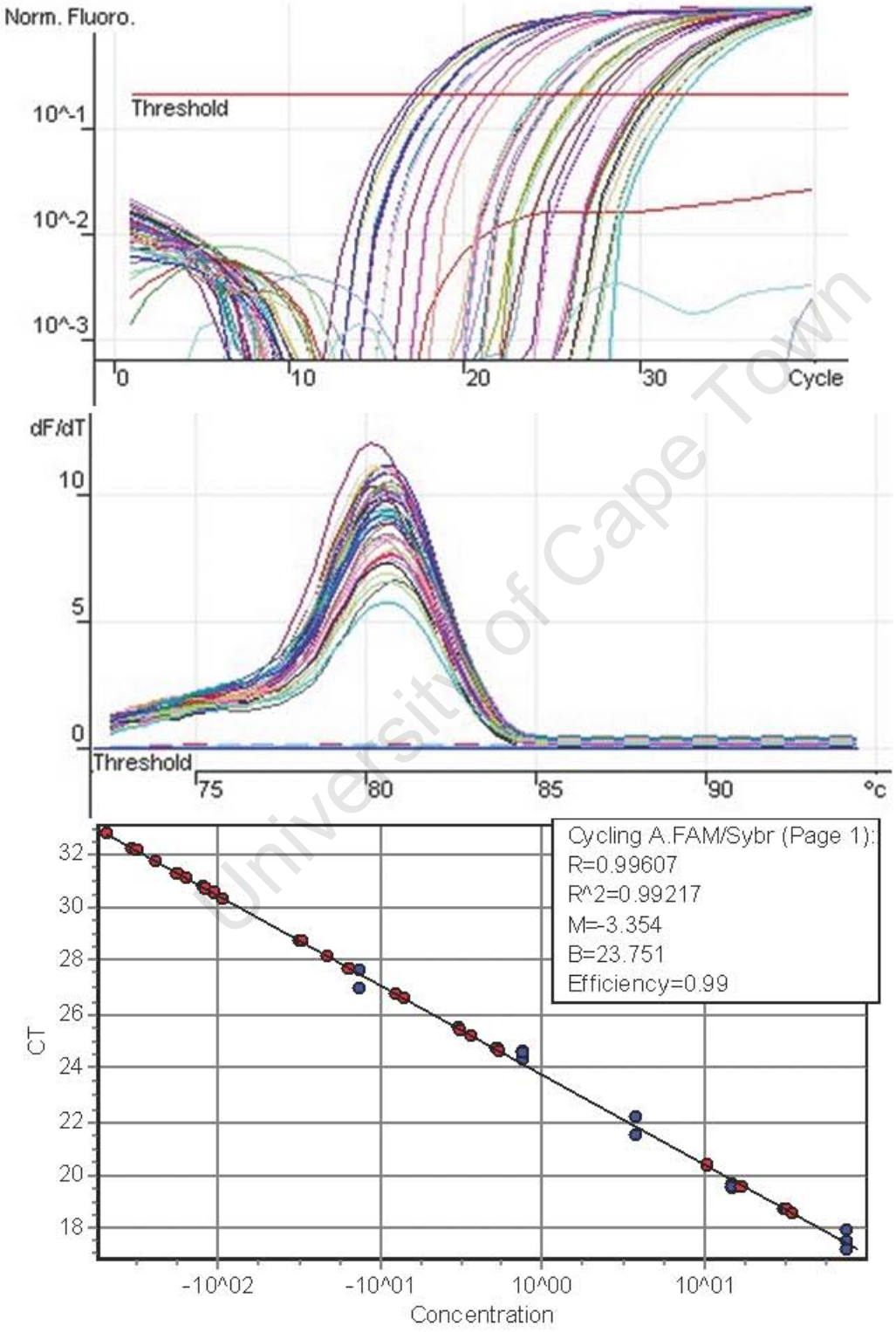
A.4.2.8.



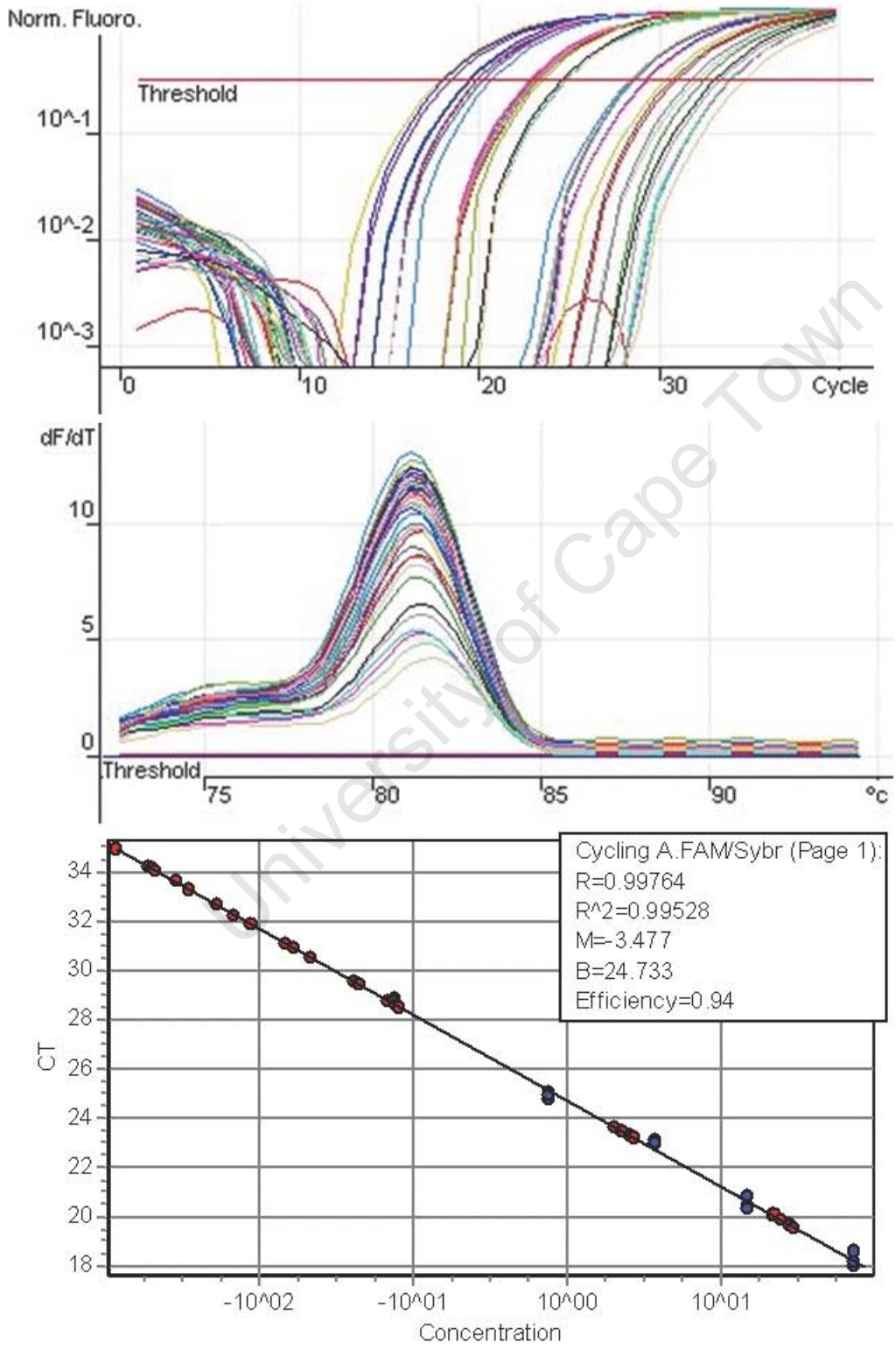
A.4.2.9.



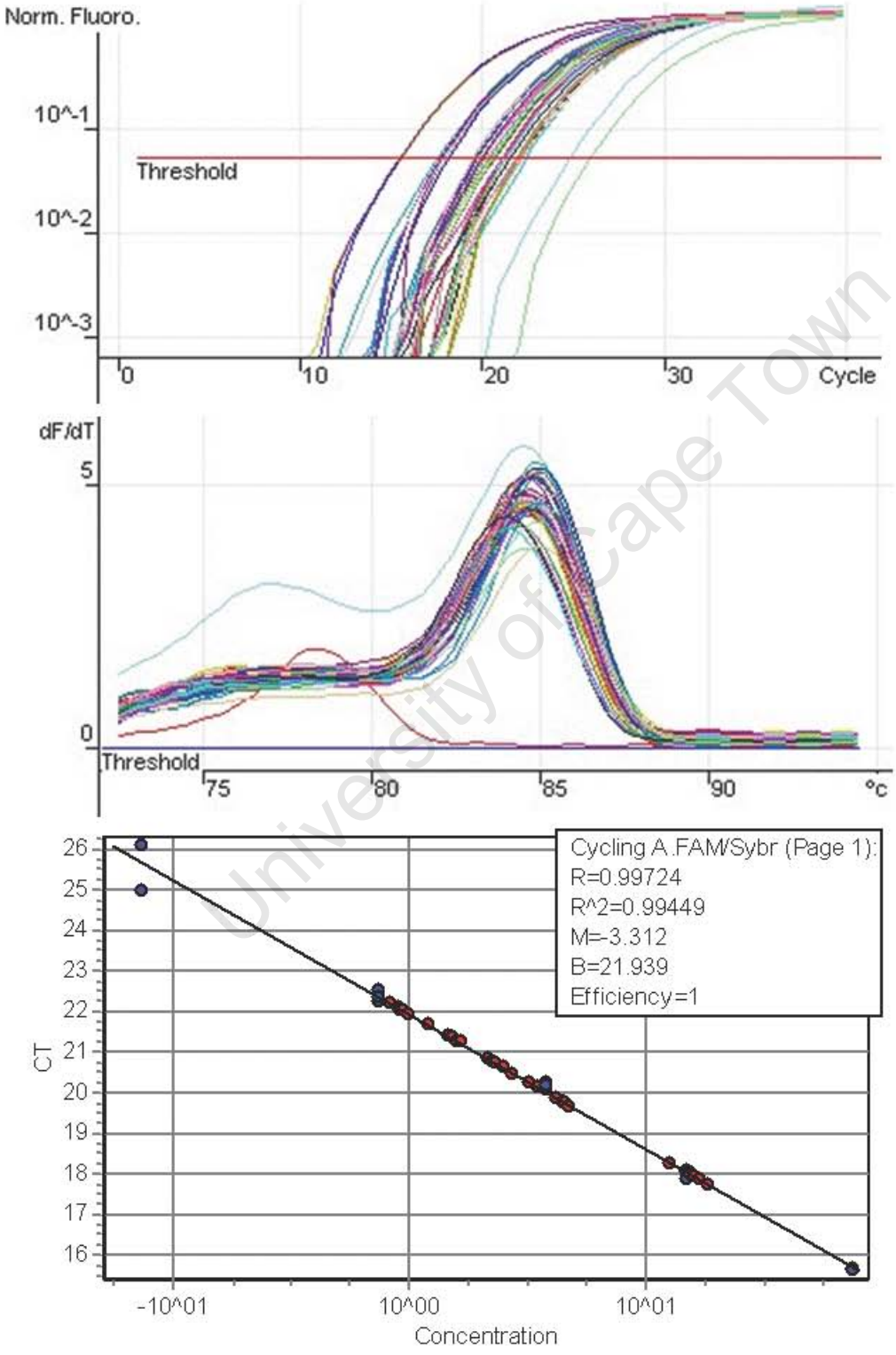
A.4.2.10.



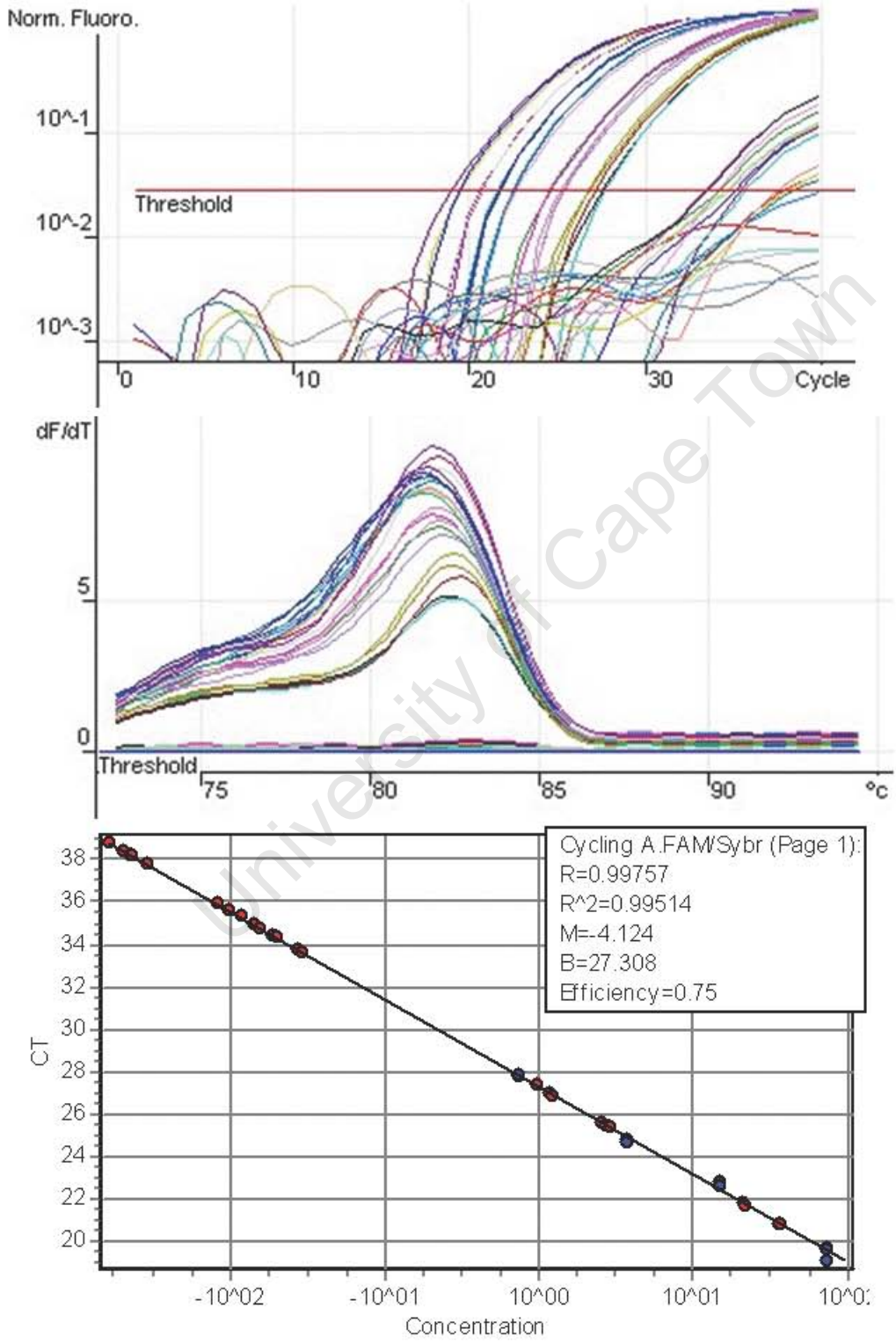
A.4.2.11.



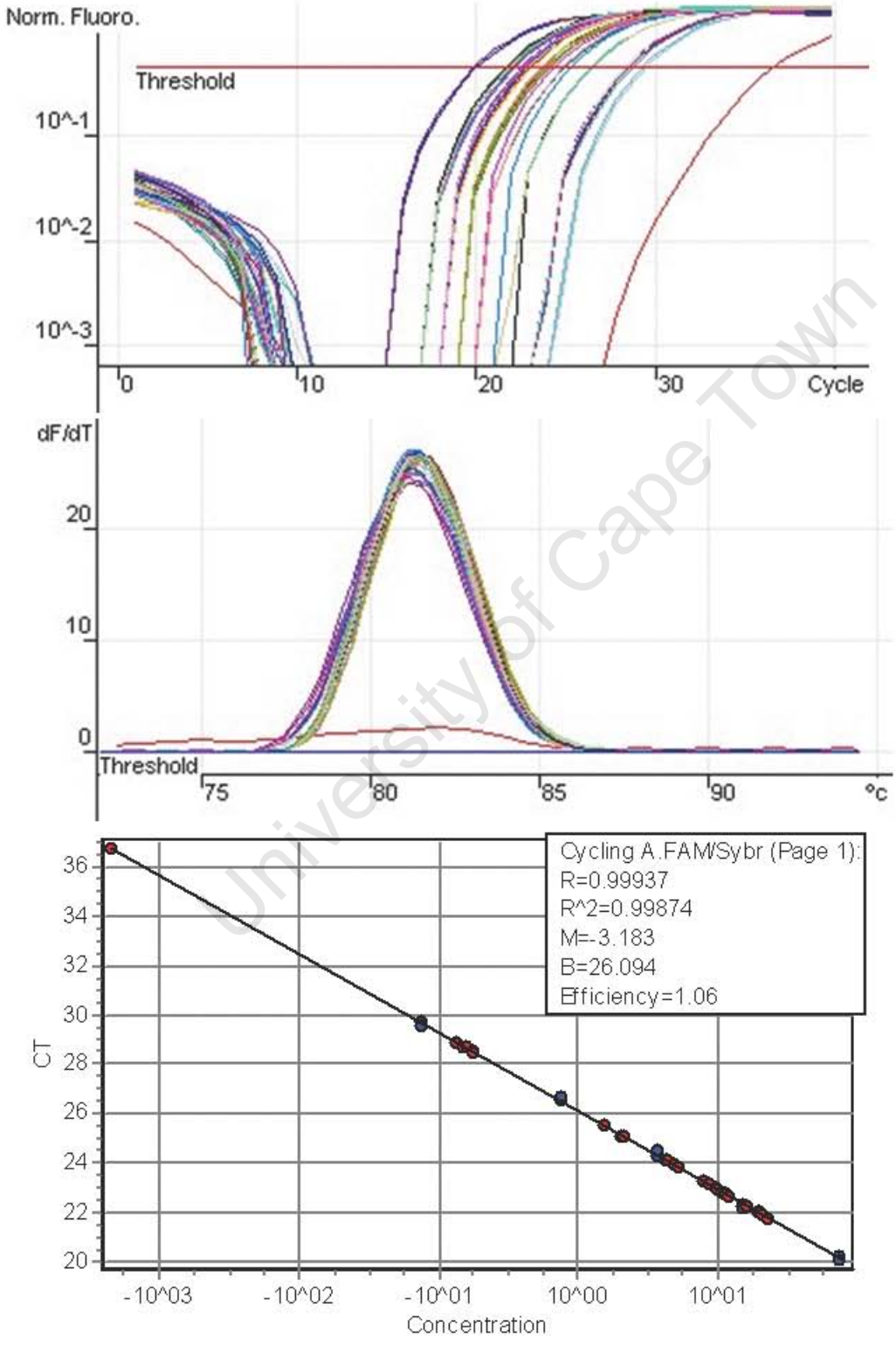
A.4.2.12.



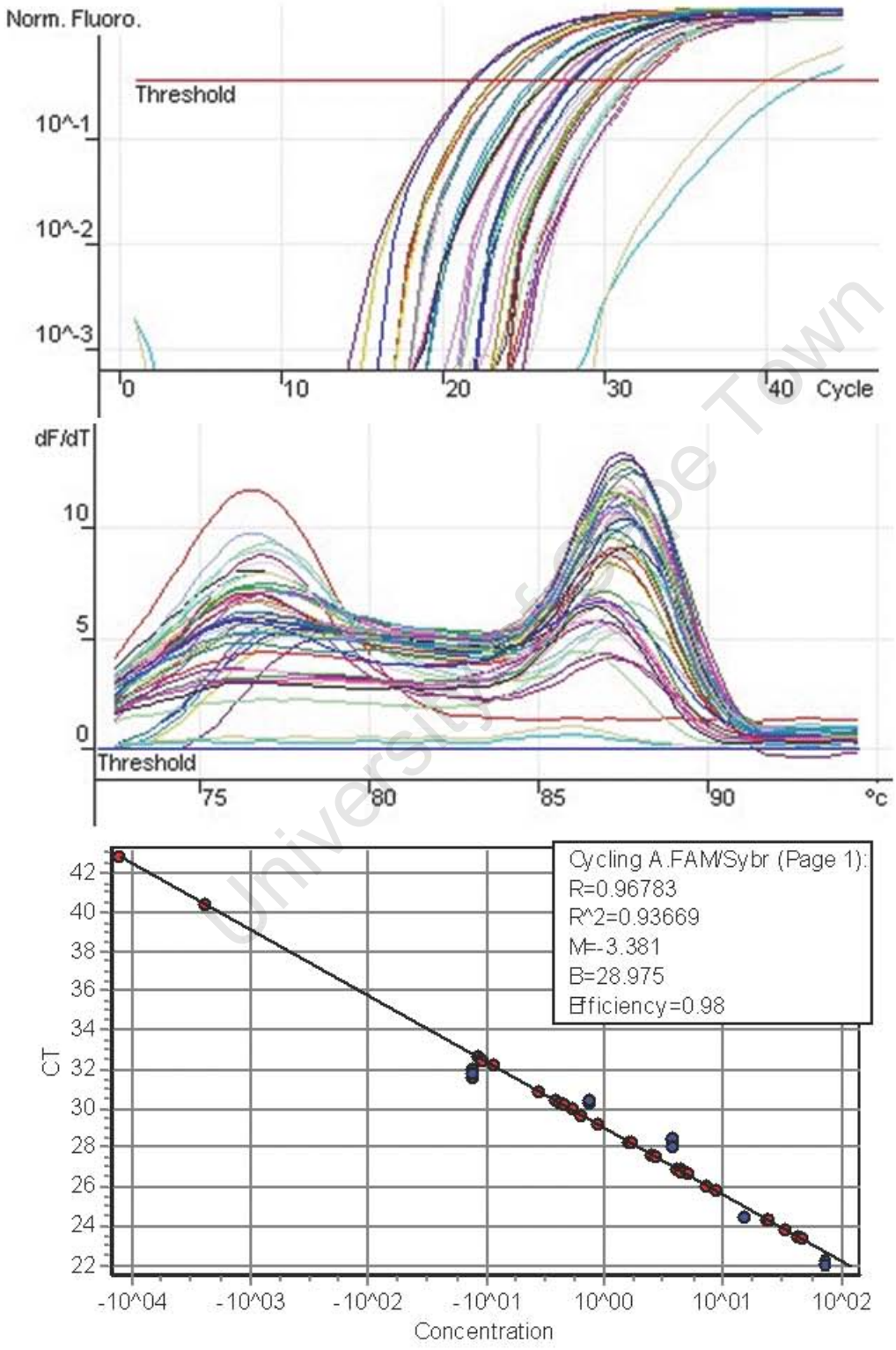
A.4.2.13.



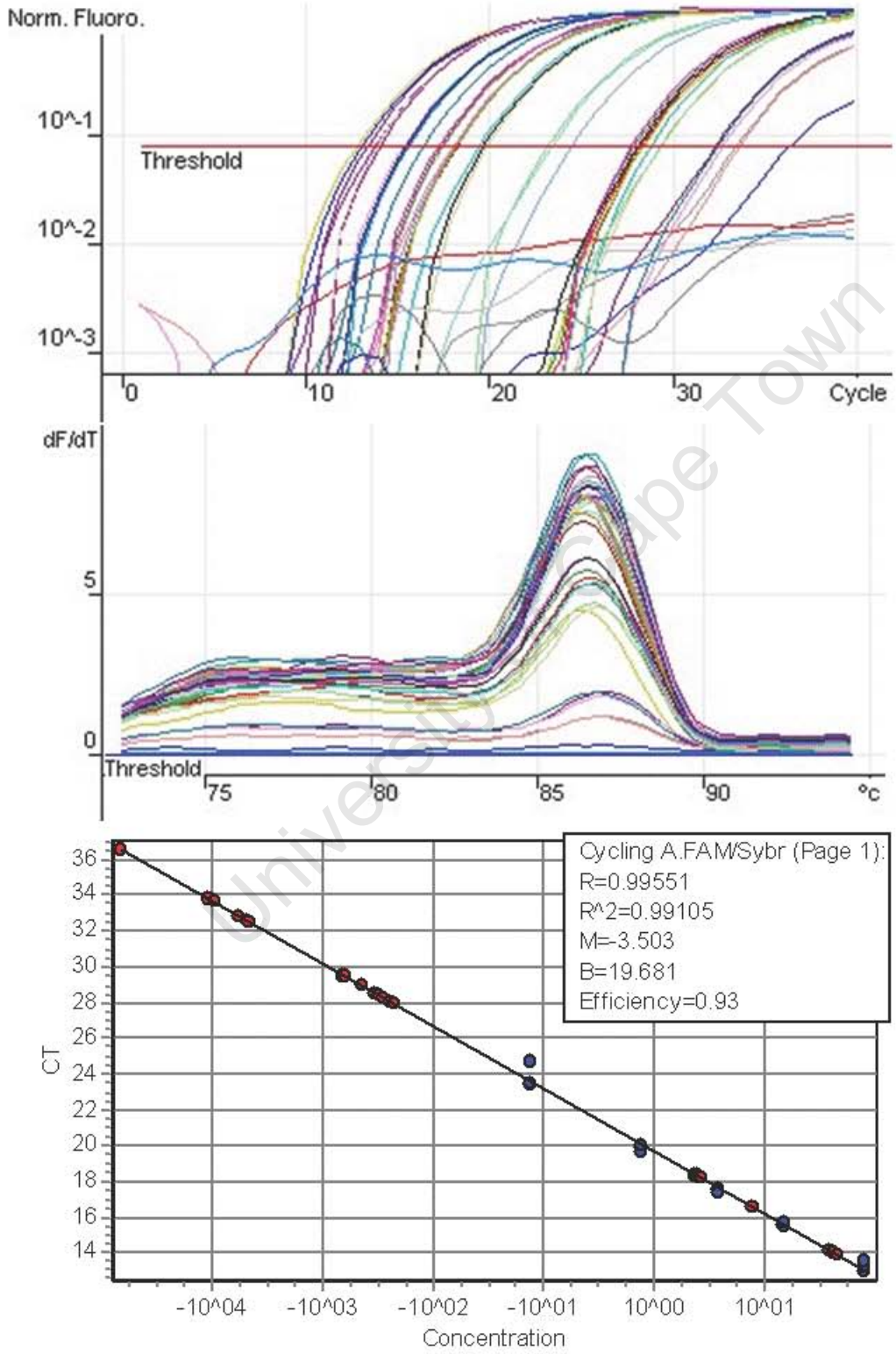
A.4.2.14.



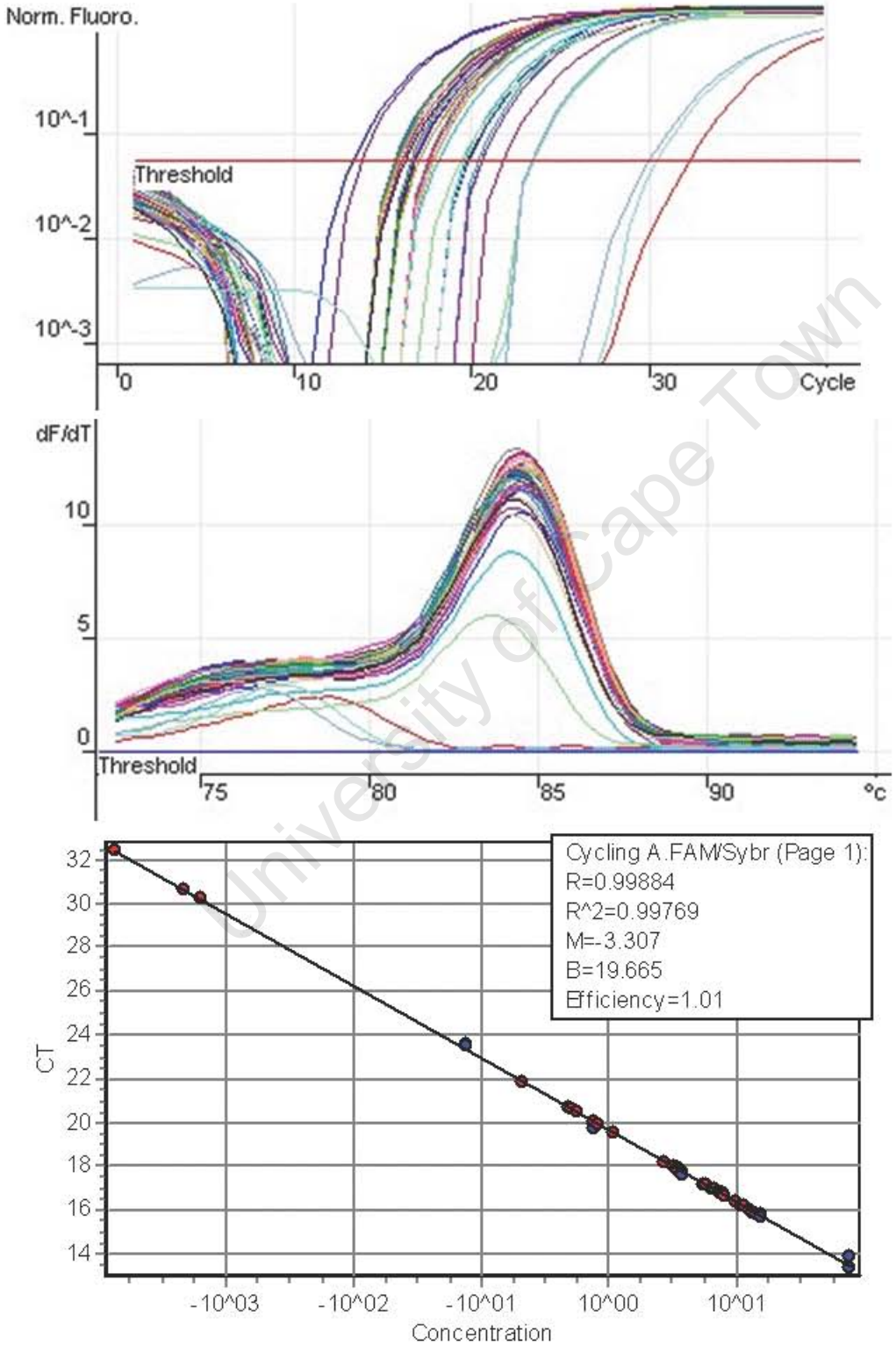
A.4.2.15.



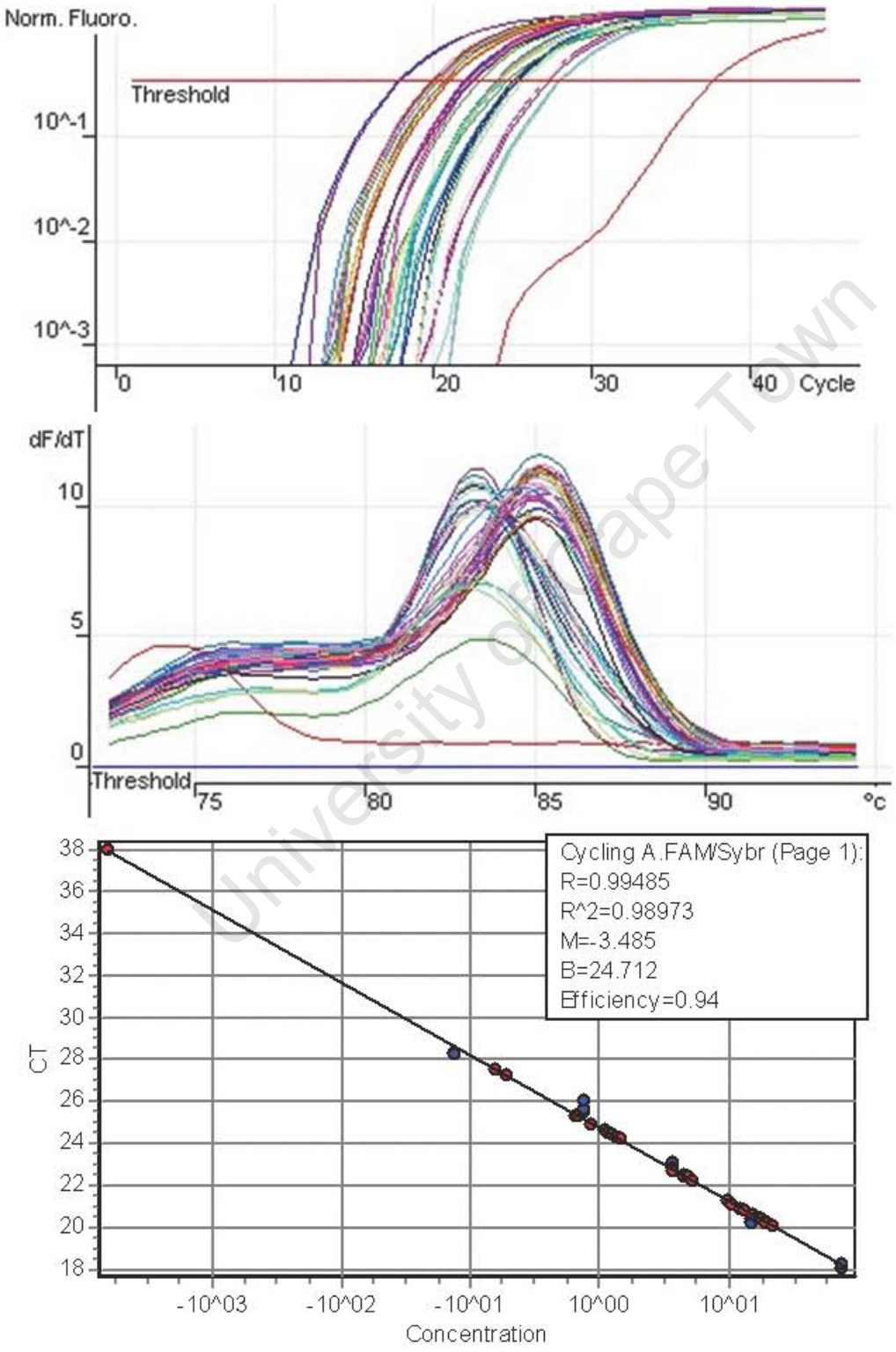
A.4.2.16.



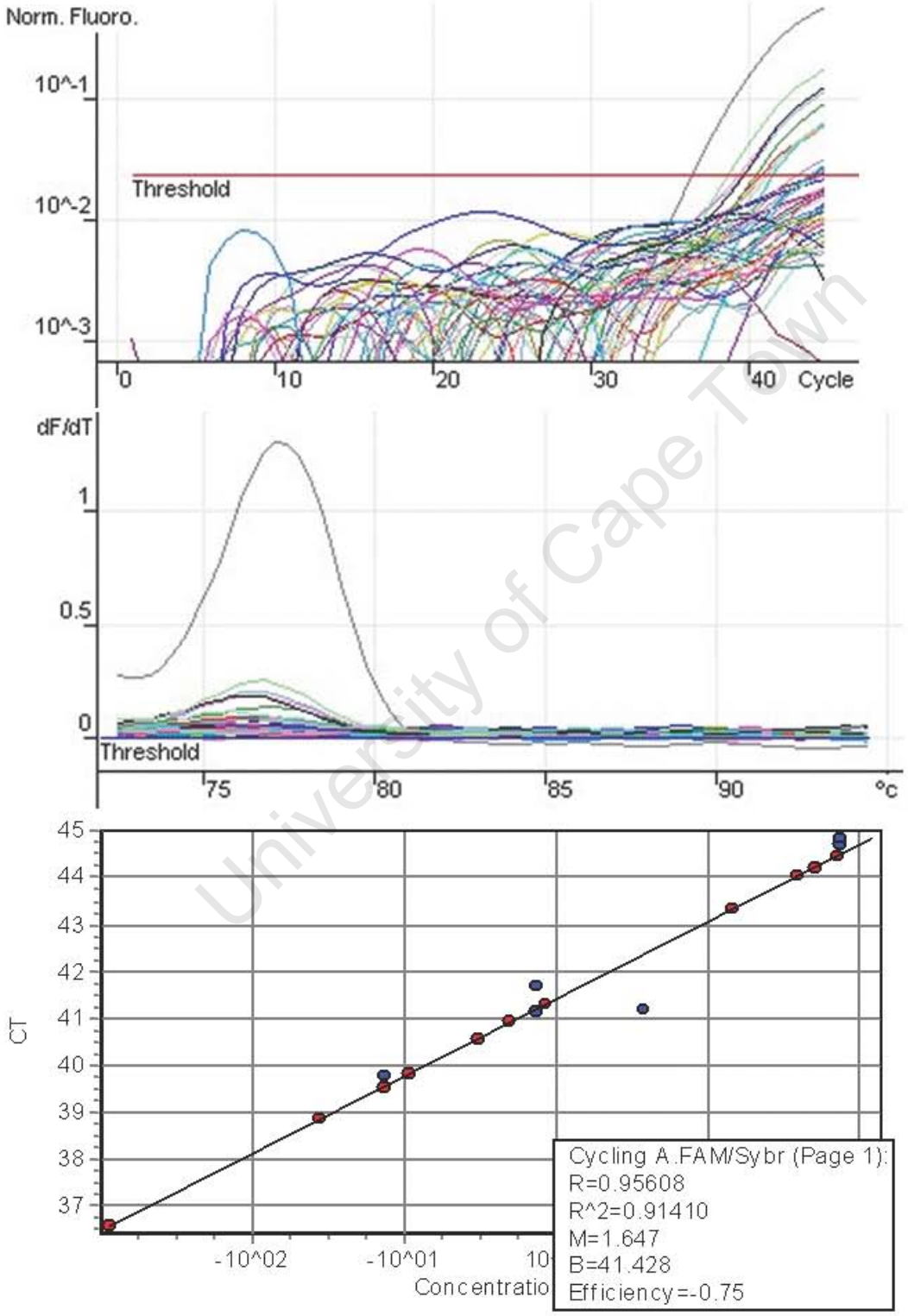
A.4.2.17.



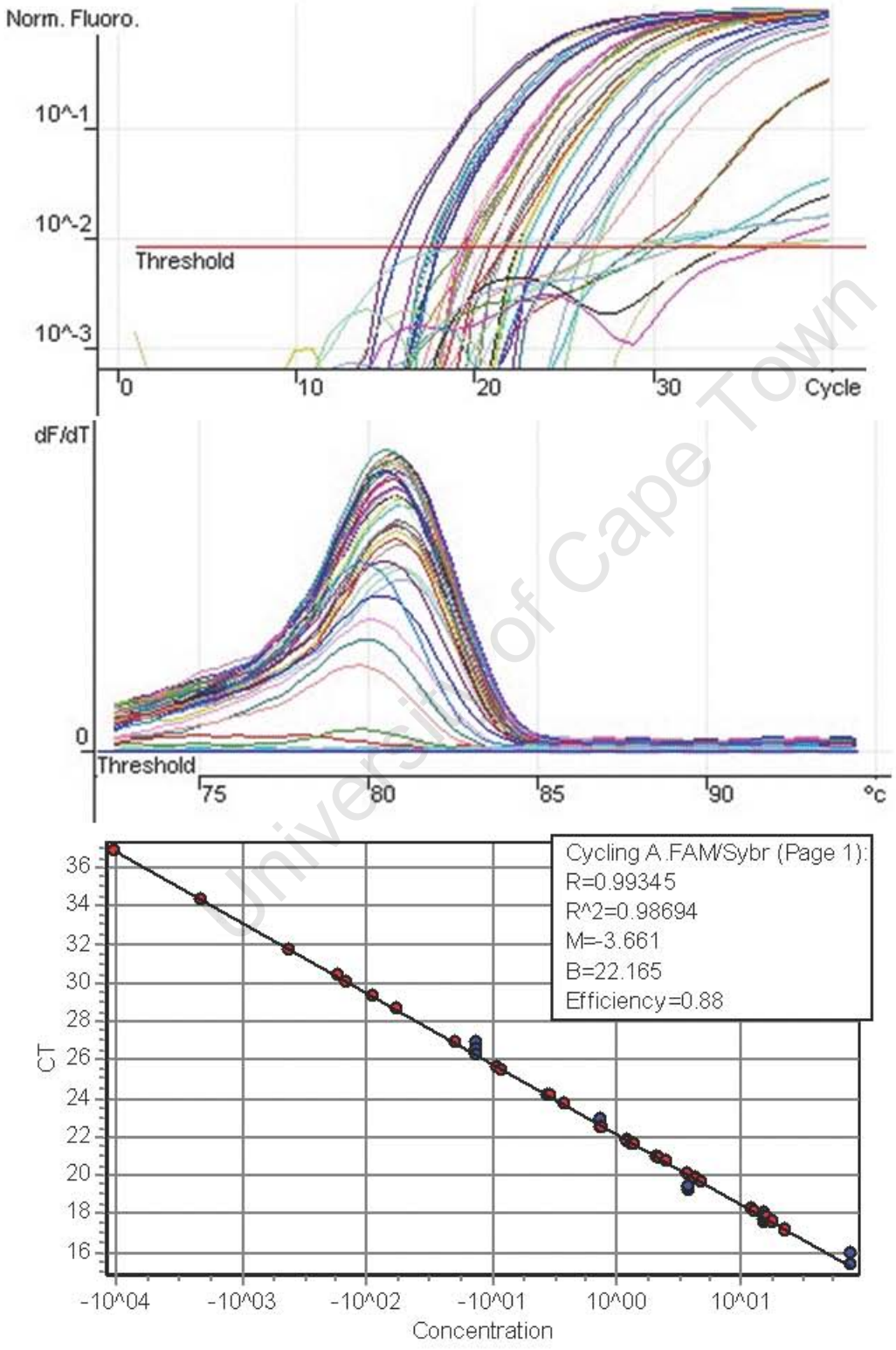
A.4.2.18.



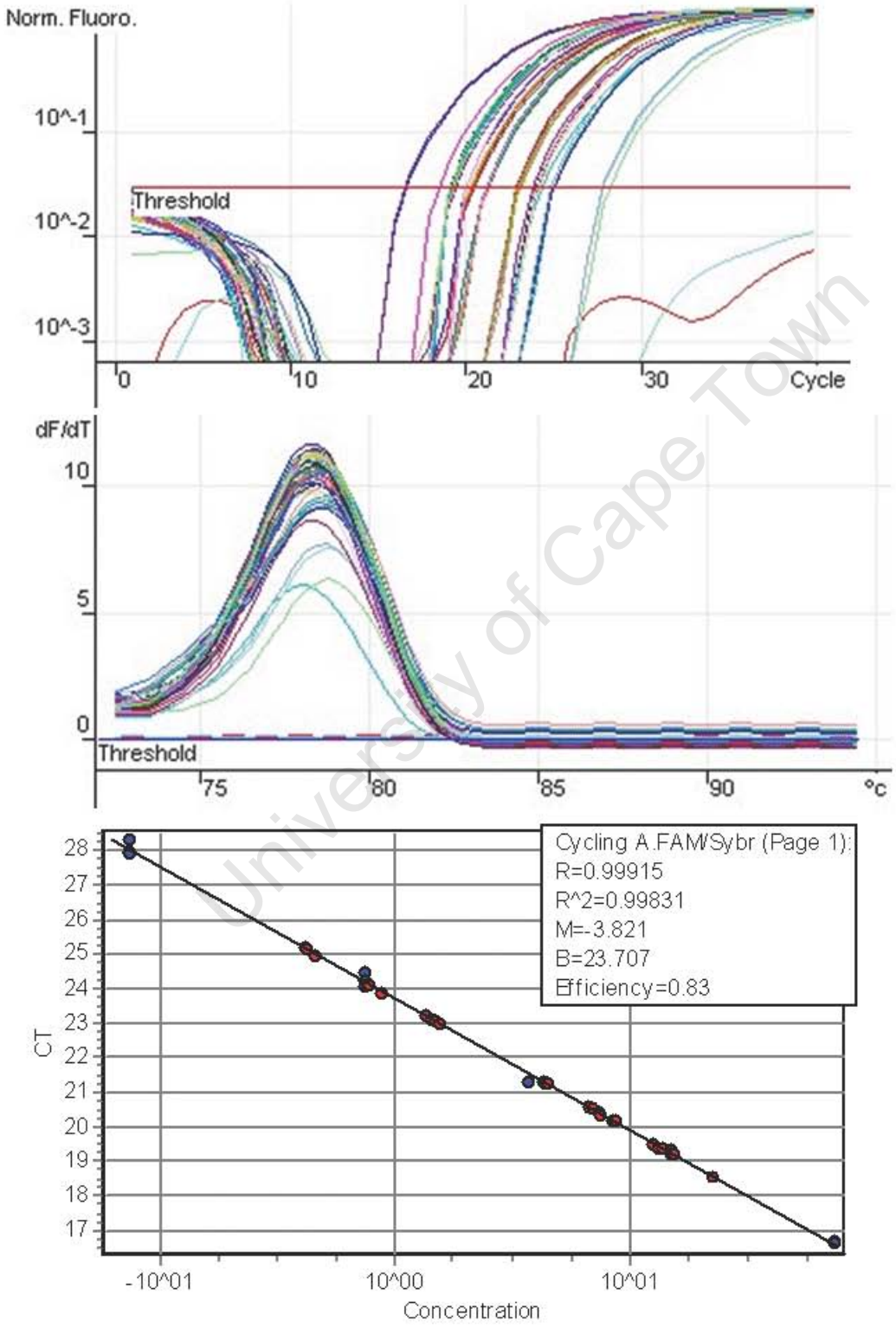
A.4.2.19.



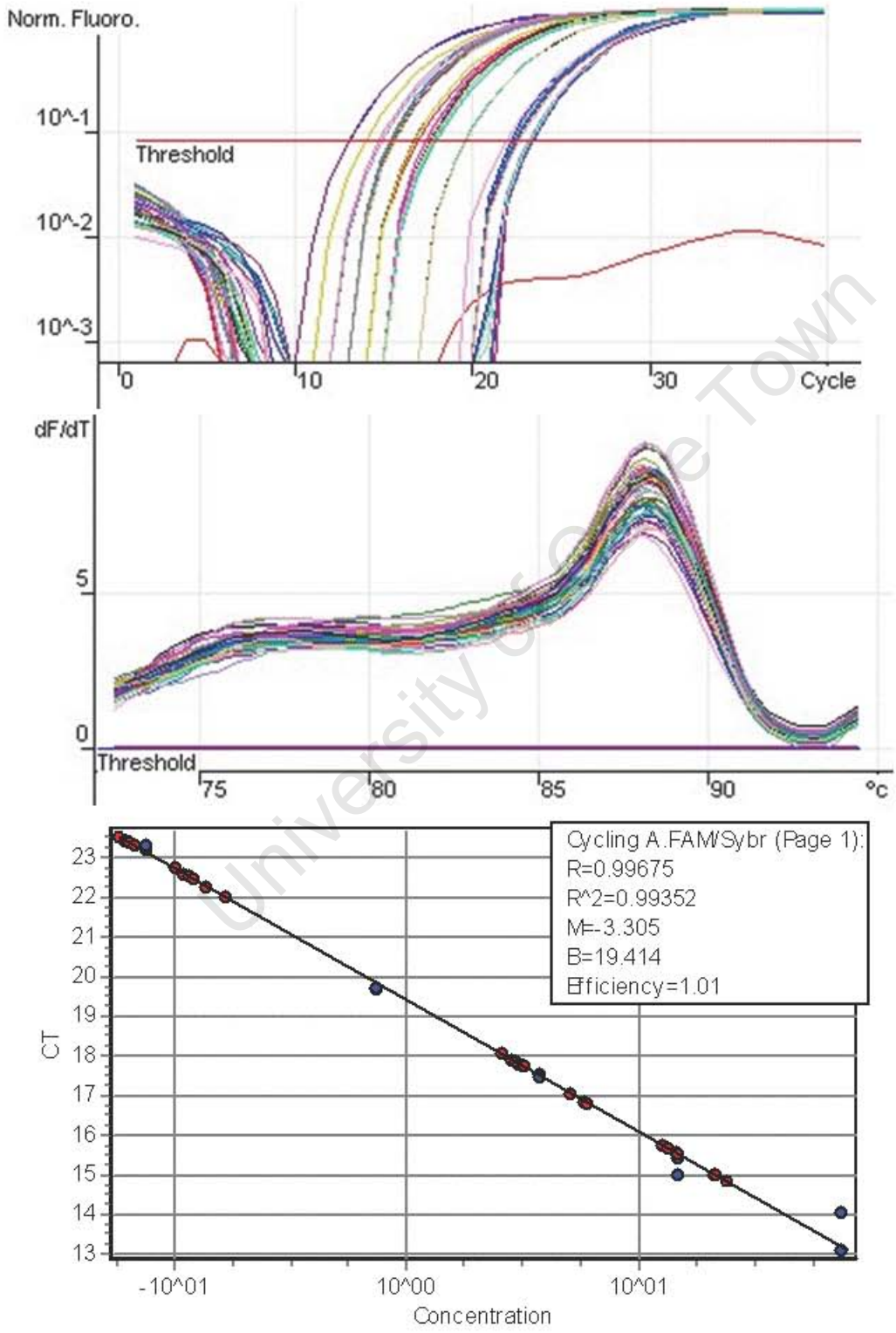
A.4.2.20.



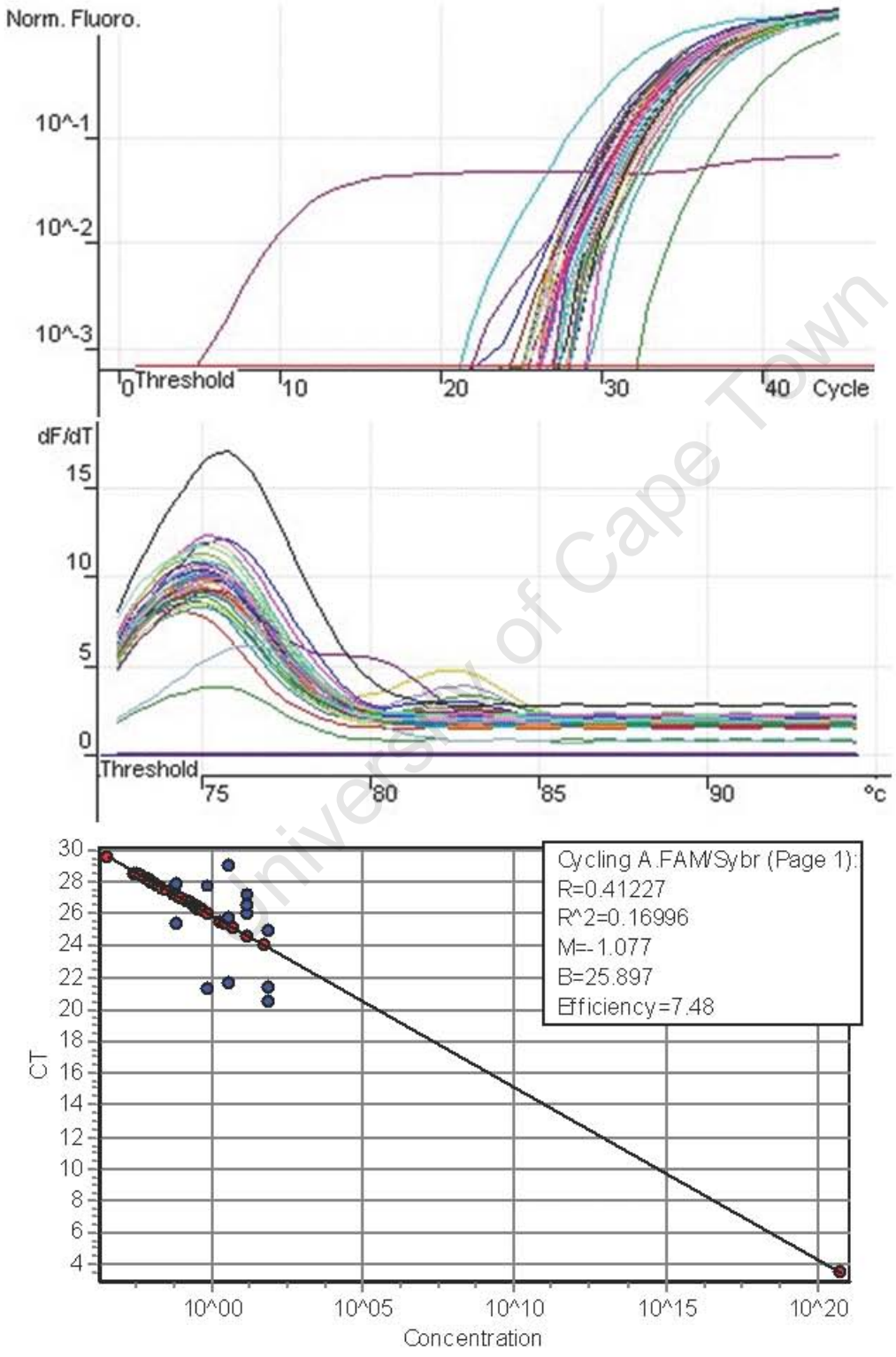
A.4.2.21.



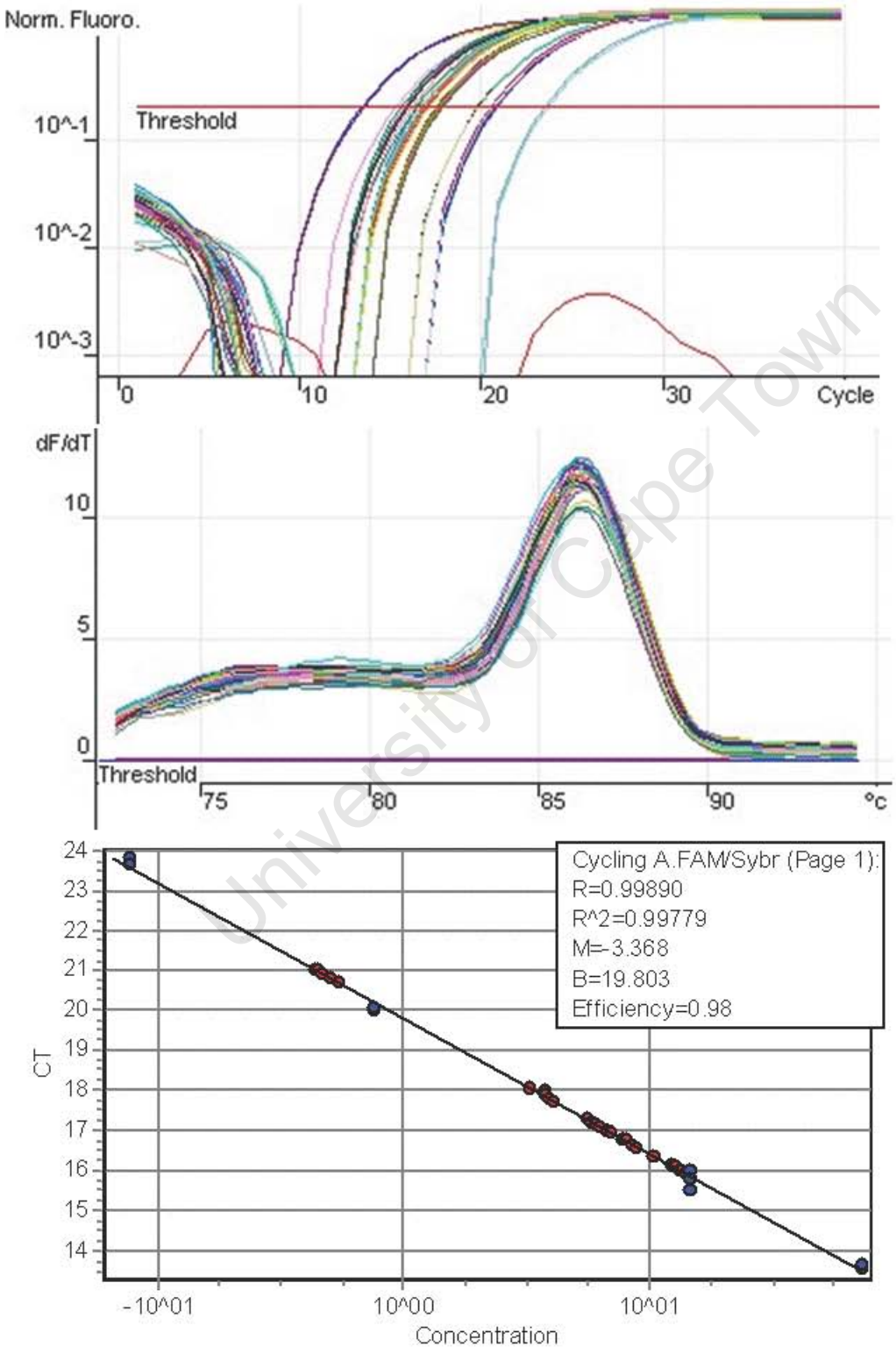
A.4.2.22.



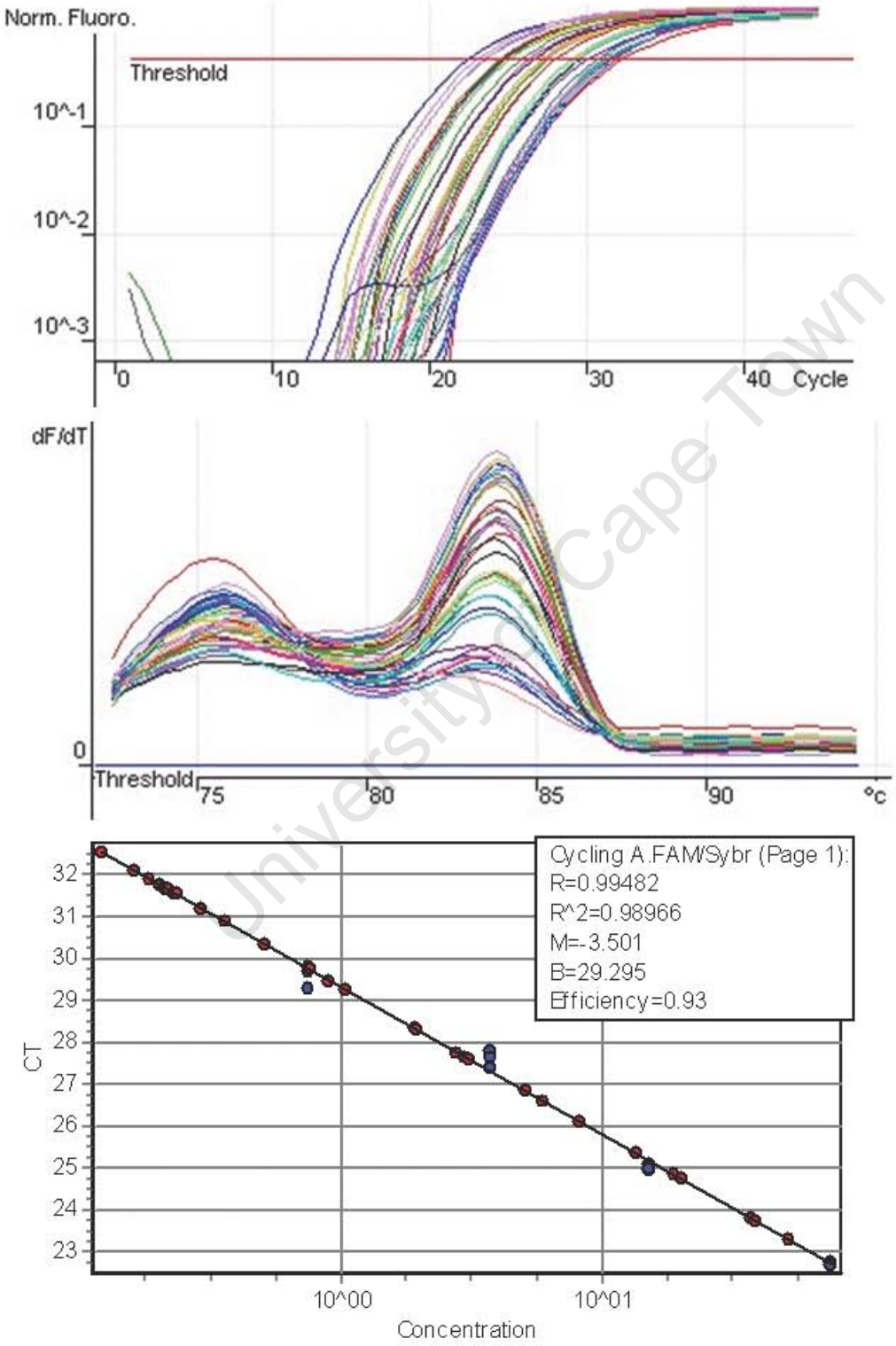
A.4.2.23.



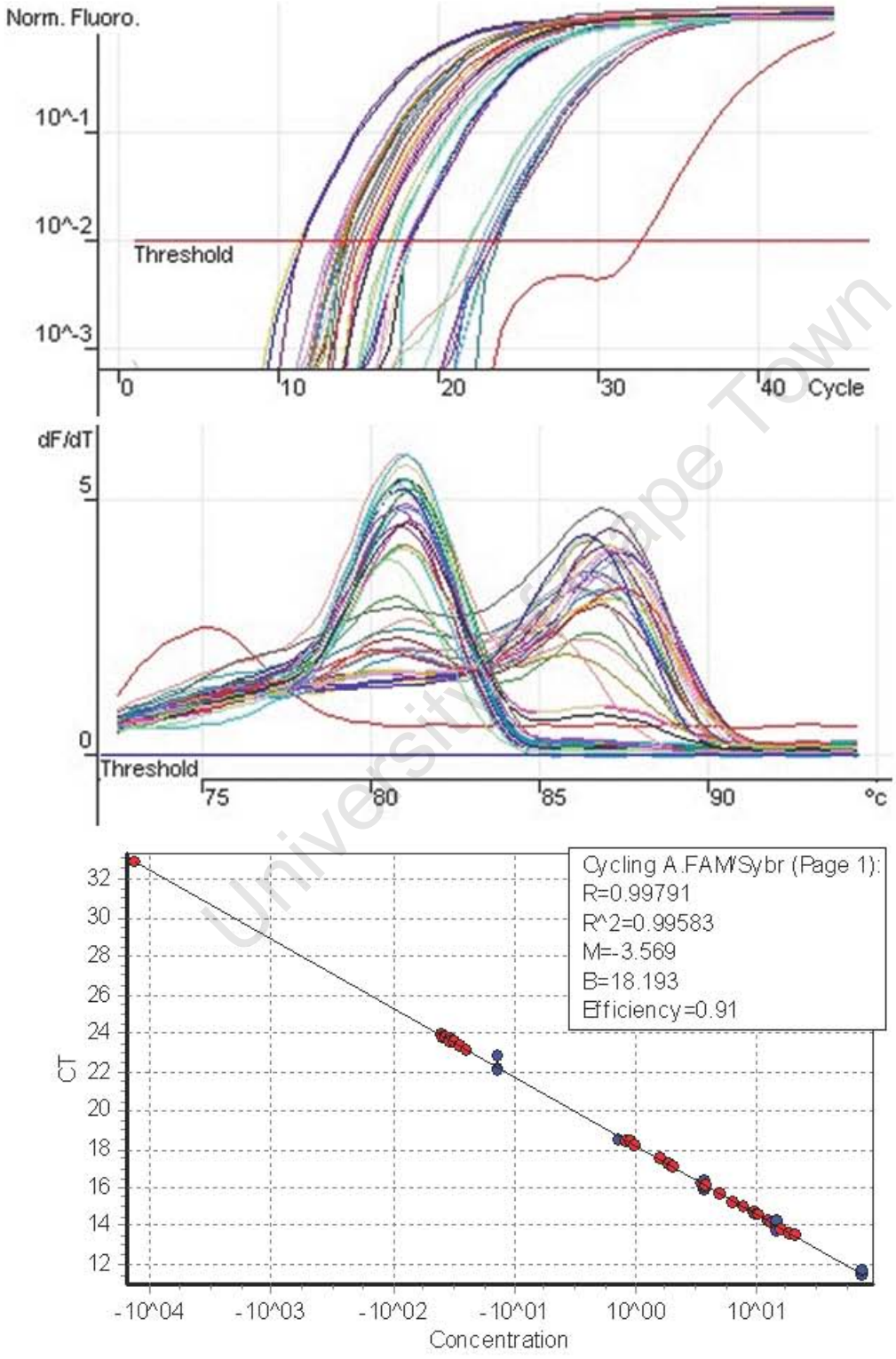
A.4.2.24.



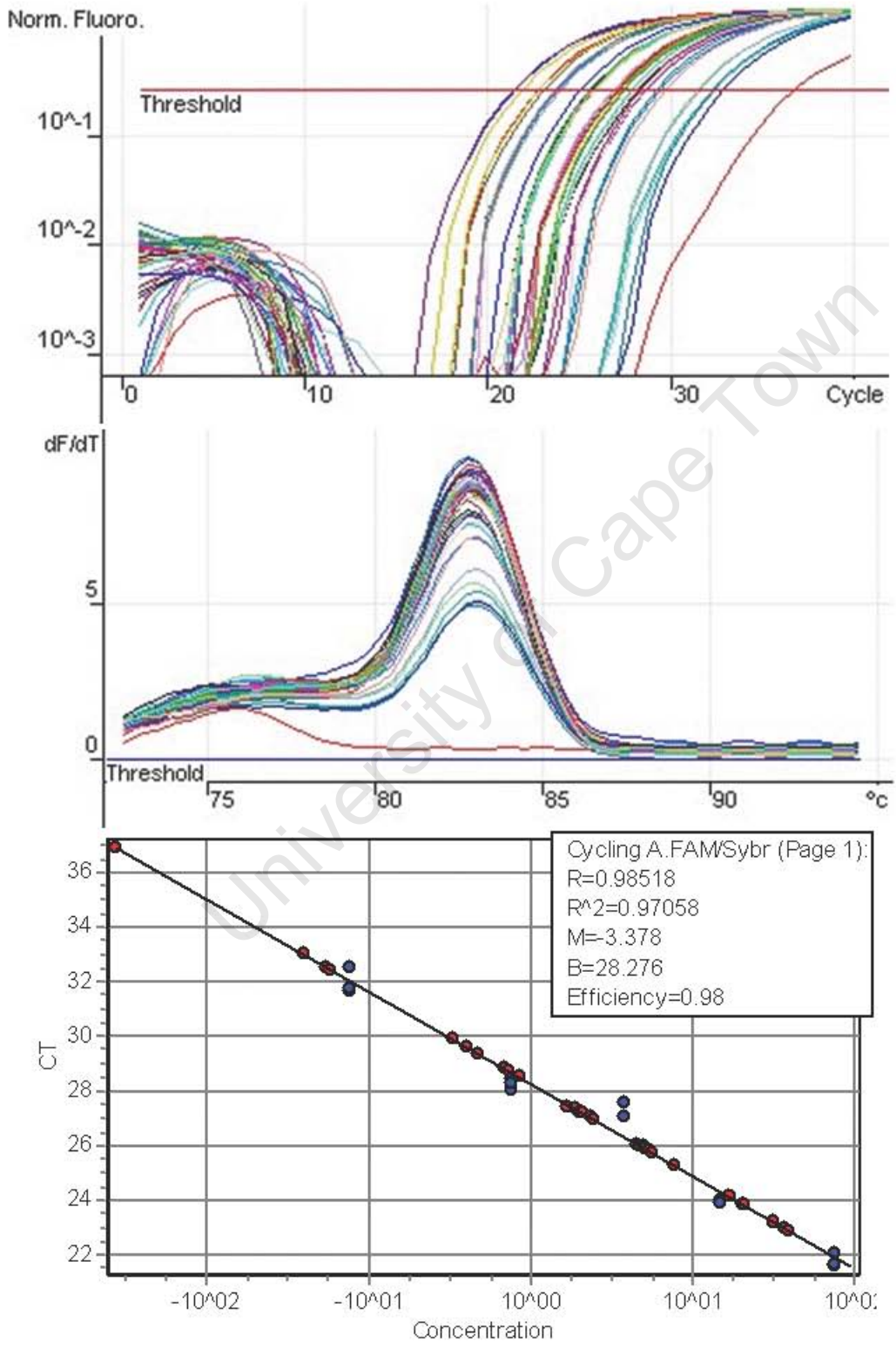
A.4.2.25.



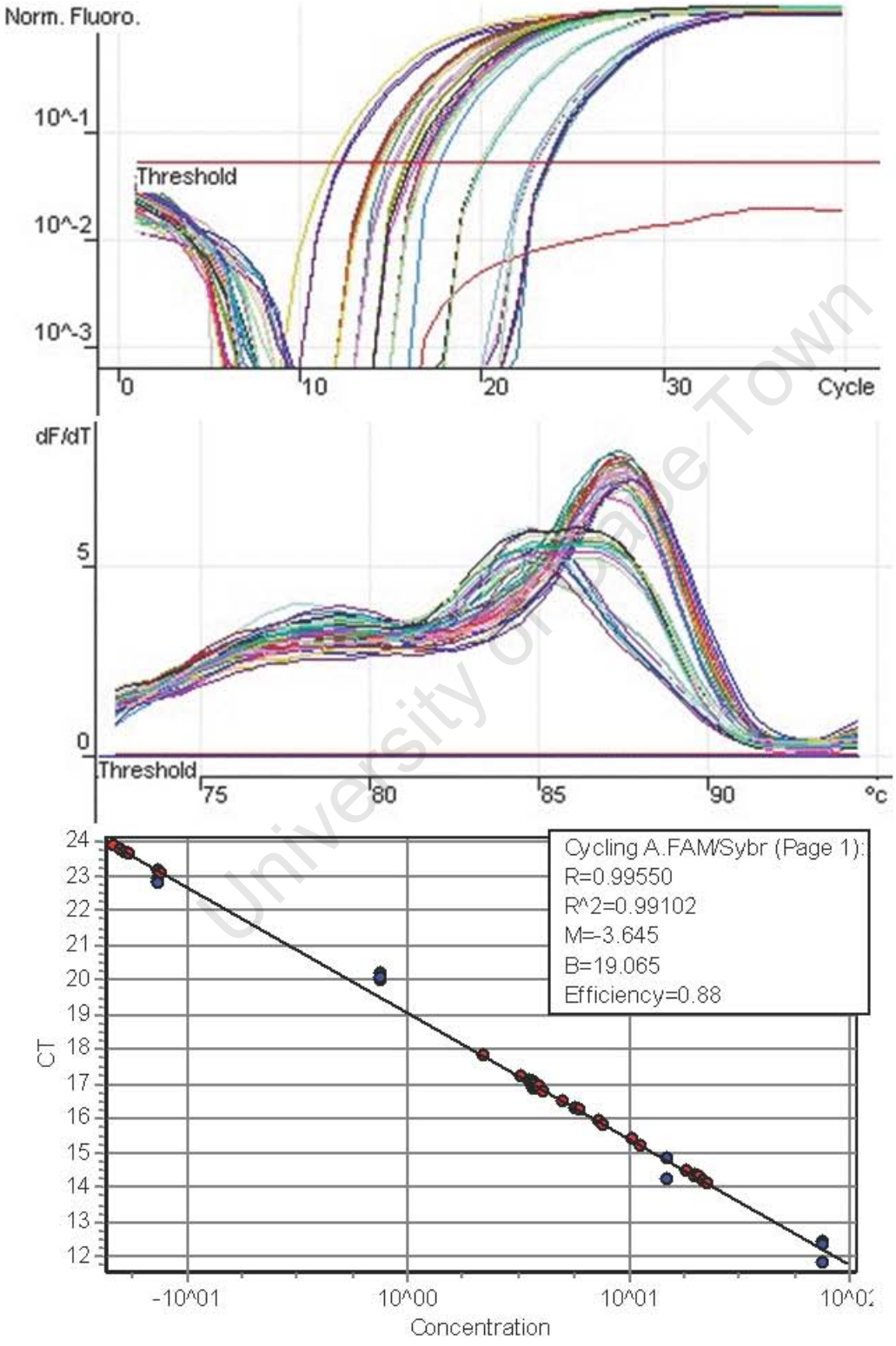
A.4.2.26.



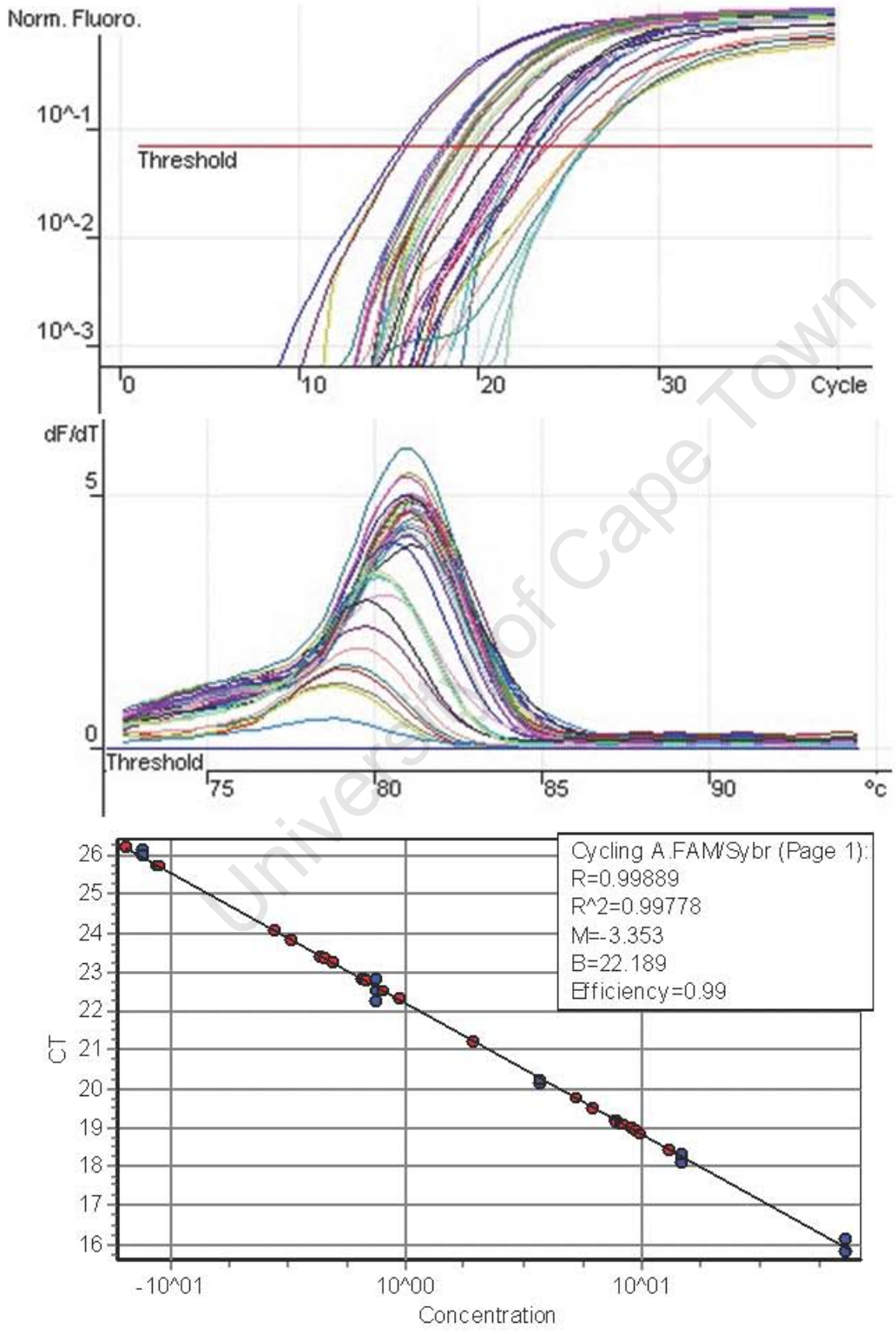
A.4.2.27.



A.4.2.28.



A.4.2.29.



A.4.2.30.