MOLECULAR CHARACTERIZATION OF TEMPERATURE-SENSITIVE GROWTH MUTANTS OF ARABIDOPSIS THALIANA

Elke Vaeck

Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor (Ph. D.) in Science (Biotechnology)

Academic year 2005-2006

Promotor: Prof. Dr. Marc Zabeau
Co-promotor: Prof. Dr. Dirk Inzé
EXAM COMMISSION

Promotor

Prof. Dr. Marc Zabeau (Secretary)
Dept. of Plant Systems Biology
Flanders Interuniversity Institute for Biotechnology (VIB) – Ghent University

Co-Promotor

Prof. Dr. Dirk Inzé
Dept. of Plant Systems Biology
Flanders Interuniversity Institute for Biotechnology (VIB) – Ghent University

Promotion Commission

Prof. Dr. Rudi Beyaert (Chairman)
Dept. of Molecular Biomedical Research
Flanders Interuniversity Institute for Biotechnology (VIB) – Ghent University

Dr. Marnik Vuylsteke
Dept. of Plant Systems Biology
Flanders Interuniversity Institute for Biotechnology (VIB) – Ghent University

Dr. Gerrit T.S. Beemster
Dept. of Plant Systems Biology
Flanders Interuniversity Institute for Biotechnology (VIB) – Ghent University

Dr. Frank Van Breusegem
Dept. of Plant Systems Biology
Flanders Interuniversity Institute for Biotechnology (VIB) – Ghent University

Prof. Dr. Wout Boerjan
Dept. of Plant Systems Biology
Flanders Interuniversity Institute for Biotechnology (VIB) – Ghent University

Dr. Janny L. Peters
Dept. of Plant Genetics
Institute for Wetland and Water Research – Radboud University Nijmegen

Dr. Sandra Ormenèse
Laboratoire de Physiologie Végétale
Université de Liège (Sart-Tilman)
If the doors of perception were cleansed
everything would appear to man as it is, infinite.

W. Blake
Acknowledgements

Spending five years preparing my PhD thesis was a challenging but rewarding experience. During this period, my colleagues, family and friends have supported me in many different ways. For this, I want to express my sincere appreciation to them.

First, I want to thank my promotor, Marc Zabeau, for giving me the opportunity to start working in the department and for expressing confidence in me. Marc, during the past five years you were always there at the crucial moments to give useful input, advice and new, inspiring insights.

Second, I’m grateful to Dirk Inzé for allowing me to work in a modern and well-equipped research environment and for his continuous support and advice.

Also, I want to express my gratitude to Pia Neyt, Janny Peters and Tom Gerats for the help and guidance they gave me during the first year of my PhD research. After Janny and Tom left the lab, Gerda Cnops further assisted me in the map-based cloning procedure, which I have sincerely appreciated.

From the second year on, Marnik Vuylsteke guided me in my work and recently helped me a lot in finalizing this thesis. Marnik, thank you for your input and help, the useful conversations, the trust, and ... the statistics.

Gerrit Beemster has taken the responsibility of being my “second advisor” during the past two years, especially when it came down to the biological stuff. He also spent quite some time to work his way through this thesis. Gerrit, thank you!

In Marnik’s lab, I met a lot of new people over the years, some of which have become real friends. Hilde, Lindy, Ruth, Ilse, Emmy, Debbie, Raphaël, Roel, Jeroen, Annelies and others, thank you for the enjoyable time together!! Especially, I want to mention Emmy Coppens, who, with her dedicated work, made a substantial contribution to the map-based cloning of the ts mutants.

I also want to thank my husband, Kurt, who has supported me fully all the way from the beginning of this work and who especially during the last six months, had to share with me all the inconvenience and stress of a race to the deadline for finalizing this thesis. I could not imagine succeeding in this work without the support of my wonderful parents, who always encouraged me to go for new challenges in life. My very sincere gratitude to my little sister, Kirsten, and brother in law, Andy, who were always there for me in difficult times and who developed the design for the cover of this thesis.

THANK YOU!
Table of Contents

Chapter One
GENERAL INTRODUCTION

1.1 The Identification of Genes with Unique and Essential Functions  2
1.2 Temperature-Sensitive Mutants: Tools to Study Essential Genes In Vivo  2
1.3 Establishing Gene Function Through Mutagenesis  19
1.4 Plant Functional Genomics Through Transcriptome Analysis  46

Chapter Two
ISOLATION OF TEMPERATURE-SENSITIVE MUTATIONS IN ARABIDOPSIS THALIANA

2.1 Introduction  58
2.2 Experimental Procedures  59
2.3 Results  62
2.4 Discussion  70
2.5 Conclusion  74

Chapter Three
MAP-BASED CLONING OF THE TEMPERATURE-SENSITIVE MUTATIONS dgr1 AND dgr2 IN ARABIDOPSIS THALIANA

3.1 Introduction  77
3.2 Experimental Procedures  78
3.3 Results  82
3.4 Discussion  90
3.5 Conclusions  101

Chapter Four
EXPRESSION ANALYSIS OF THE TEMPERATURE-SENSITIVE dgr2 MUTANT OF ARABIDOPSIS THALIANA

4.1 Introduction  104
4.2 Experimental Procedures  104
4.3 Results  107
4.4 Discussion  112
4.5 Conclusion  114
Chapter Five
GENERAL CONCLUSIONS AND PERSPECTIVES 114

Chapter Six
SAMENVATTING 117

References 120
In little more than a decade, research in plant biology has taken huge steps forward towards understanding nearly all aspects of plant growth and development. Although significant progress has been made in many species, we will focus here on Arabidopsis thaliana, which has been adopted as a model system for molecular genetic studies. One of the long-term goals of Arabidopsis research is to obtain a better understanding of all the biological processes that operate in a plant. To achieve this, it is necessary to define the minimal set of genes required to produce a viable plant with a normal phenotype under diverse conditions. Because of the complexity and the multitude of biological processes required for a plant to grow and develop, a large and diverse set of essential genes are likely to be involved. Applying forward genetics is a powerful method to identify the relevant genes. This approach usually involves a screen for mutants that are used to identify genes involved in specific processes. However, if the genes underlying these processes are essential for the viability of the plant, mutations will often result in a lethal phenotype. Therefore, conditional lethal mutants are very useful tools to study the function of essential genes in vivo.

In our own work, we isolated temperature-sensitive Arabidopsis mutants to identify novel genes that are essential for plant growth and development. Temperature-sensitivity is the most common type of conditional mutation in which the mutant phenotype only appears at certain (usually higher) restrictive temperatures. This way, gene function can be studied at the restrictive temperature, while mutant lines can be maintained and propagated at the permissive temperature, at which the wild type phenotype is expressed. Through the phenotypic analysis and molecular genetic characterization of these temperature-sensitive mutants we aimed to unravel a piece of the huge network of biochemical processes that underlies plant growth and development. This introductory chapter presents an overview of the history, use and generation of temperature-sensitive mutants and the molecular genetic methods and technologies that are currently applied for their characterization.
1.1 THE IDENTIFICATION OF GENES WITH ESSENTIAL FUNCTIONS

‘Essential’ and ‘non-essential’ are classical molecular genetic designations that relate to the functional significance of a gene with respect to its effect on organismic fitness. Generally, a gene is considered to be essential if a knockout results in (conditional) lethality or infertility. In organisms of which the genome sequence is known, the large-scale identification of essential genes with unique functions using gene inactivation technologies has become an important research topic. The long-term goal is to define the minimal gene set needed to produce a viable organism with a normal phenotype under diverse conditions. The systematic identification of essential genes has been described in several prokaryotes, including Mycoplasma genitalium (Hutchison et al., 1999), Staphylococcus aureus (Ji et al., 2001), Haemophilus influenzae (Akerley et al., 2002), Vibrio cholerae (Judson et al., 2000) and Bacillus subtilis (Kobayashi et al., 2003). Related work has also been published for two model eukaryotes, Saccharomyces cerevisiae (Giaever et al., 2002) and Caenorhabditis elegans (Kamath et al., 2003) and significant advances have been reported in the identification of essential genes associated with human diseases (Jimenez-Sanchez et al., 2001).

Only recently, a comparable dataset of essential plant genes has been constructed (Tzafrir et al., 2004). An initial collection of 250 EMB genes required for normal embryo development in Arabidopsis is described and represents the first large-scale dataset of essential genes in a flowering plant. Sixty percent of these EMB identities were derived from a forward genetic screen of T-DNA insertion lines generated at Syngenta (McElver et al., 2001). Most EMB genes are required for viability and may therefore be considered as essential in a traditional sense. When compared with 550 genes linked to other knockout phenotypes, EMB genes are enriched for basal cellular functions, deficient in transcription factors and signalling components, have fewer paralogs and are more likely to have counterparts among essential genes of yeast (Saccharomyces cerevisiae) and worm (Caenorhabditis elegans). EMB genes also represent a valuable source of plant-specific proteins with unknown functions required for growth and development. Although the large-scale identification of essential genes is possible using regular knock-out mutants, the determination and study of gene function in vivo using such mutants is often problematic due to the resulting lethal phenotypes (Despres et al., 2001).

1.2 TEMPERATURE-SENSITIVE MUTANTS: A TOOL TO STUDY ESSENTIAL GENES IN VIVO

Conditional lethal mutants have provided the means to genetically study mutations in essential genes. The potential to control the activity of individual genes reversibly and quantitatively has opened up new perspectives for the in-depth study of gene function. A decisive feature of this system is that phenotypically the state of a null-mutation can be induced without altering the target gene itself, while the possibility to subsequently return to the wild type (WT) state makes the mutant phenotype truly conditional. Equally important is the quantitative nature: gene activities can be varied by small increments, a feature that is of particular interest for the study of gene products that are involved in a variety of intracellular equilibria. Whereas in general they will participate in such interactions with different affinities, a gradual change of their intracellular concentration may reveal phenotypes and targets of interaction that remain obscure when the respective gene is simply inactivated.

Temperature-sensitive (ts) mutations are the most widely used type of conditional mutations in which there is a
marked drop in the level or activity of the gene product when the gene is expressed above (or below) a certain temperature (restrictive temperature) (Chakshusmathi et al., 2004). Below (or above) this temperature, at so-called permissive temperatures, the phenotype of the mutant is similar to that of the WT. Ts mutants provide an extremely powerful tool for studying gene function in vivo. They represent a reversible mechanism to lower the level of a specific gene product at any stage in the growth of the organism simply by changing the temperature of growth. Although there are a number of other excellent inducible systems for gene expression (Gossen and Bujard, 1992; Gossen et al., 1995) ts mutants have several unique advantages over such systems, including fast temporal response, high reversibility and the applicability to any tissue type or developmental stage of an organism. The use of ts alleles is therefore a well-established method that is applicable to any organism upon which temperature changes can be imposed. These include viruses, prokaryotes, yeasts, nematodes, insects, plants and even amphibia and fish (Zeidler et al., 2004).

1.2.1 In the Early 50’s, the First Temperature-Sensitive Mutants were Exploited to Obtain Mutations in Essential Genes

In 1946 Mitchell and Houlahan produced the first published description of a Ts mutant (Neurospora). Subsequently, Horowitz and Leupold isolated several Ts mutations in Neurospora (1951). The use of these conditional mutants contributed significantly to the establishment of the one gene-one enzyme paradigm stated by Beadle and Tatum in 1945. Further isolation of Ts mutations was performed to estimate the fraction of essential genes in Escherichia coli. This work was soon recognized by Edgar and co-workers who demonstrated in bacteriophage T4 that ts mutations could be used effectively to gain information about essential gene functions and that they can occur in many different genes in an organism (Edgar and Lielausis, 1964; Epstein et al., 1963). From that moment, extensive Ts lethal mutant collections were created in bacteriophage T4 (Wood and Revel, 1976) and Escherichia coli (Kohyama et al., 1966; Sevastopoulos et al., 1977). The mutations in these collections have allowed the genetic dissection of essential functions such as DNA replication, the cell cycle and cell division, membrane biogenesis, transcription and translation (Allen et al., 1970; Schmidt et al., 1989).

1.2.2 Temperature-Sensitive Yeast Mutants have Served as Powerful Tools for the Identification of Novel Cell Cycle Genes

Since the discovery of yeast cell division cycle mutants by Hartwell and colleagues in the early 70’s, the field of cell cycle research has largely benefited from the use of Ts mutations (Cayrol et al., 1997). When Hartwell launched the study of cell cycle genetics nothing was known about the mechanisms for achieving the orderly progression of cell cycle events. He found evidence for genes, cdc genes, whose function is required at specific time points in the cycle by identifying mutants that arrested the cell cycle at a specific stage (figure 1.1).

The organism that Hartwell chose to study was the budding yeast, Saccharomyces cerevisiae. Because the cell cycle is an essential process, mutagenized yeast cultures were screened for Ts mutants that grew and divided normally at the permissive temperature, but arrested cell division with a characteristic phenotype at the restrictive temperature (figure 1.2). The first paper to describe cdc mutants was by Hartwell, Culotti and Reid (1970), who also
coined the term ‘execution point’ - the stage in the cell cycle when the gene function is required. In this paper, three cdc genes were described, which paved the way for the identification of many more such genes, and for the discovery of the molecules and mechanisms controlling the cell cycle.

Although routine methods for cloning yeast genes were still ten years away, Hartwell’s work went much further. With thoughtful and elegant genetic analysis, he used the mutants as tools to block the cell cycle at specific stages and to ask questions about the interdependence of yeast cell-cycle events. For example, in his J. Mol. Biol. Paper (1971), Hartwell concluded that DNA synthesis is required for later events, such as nuclear division. By contrast, the formation of the bud is not dependent on DNA replication, although DNA replication does prevent further bud initiation in the same cell cycle.

Three years further and yeast cell-cycle genetics was well established, as demonstrated by the synthetic model of the cell cycle that was presented in Science (1974). Genetics had divided the cycle into two parallel pathways, which comprised two sets of dependent steps and involved a total of 19 cdc genes. At the beginning of the cell cycle, both pathways depend on the completion of a step that Hartwell and co-workers termed ‘start’. This event is defined by the famous cdc28 mutant (famous because the cdc28 gene was later shown to encode the founding member of the cyclin-dependent kinase family), and is also the event at which yeast mating factor arrests cell division to prepare cells for mating. ‘Start’ turned out to be an important control point for the cell cycle in many eukaryotes.

Three years further and yeast cell-cycle genetics was well established, as demonstrated by the synthetic model of the cell cycle that was presented in Science (1974). Genetics had divided the cycle into two parallel pathways, which comprised two sets of dependent steps and involved a total of 19 cdc genes. At the beginning of the cell cycle, both pathways depend on the completion of a step that Hartwell and co-workers termed ‘start’. This event is defined by the famous cdc28 mutant (famous because the cdc28 gene was later shown to encode the founding member of the cyclin-dependent kinase family), and is also the event at which yeast mating factor arrests cell division to prepare cells for mating. ‘Start’ turned out to be an important control point for the cell cycle in many eukaryotes.

Figure 1.1 THE BEHAVIOR OF A Ts CDC MUTANT (GRIFFITHS ET AL., 1996). (A) At the permissive (low) temperature, the cells divide normally and are found in all phases of the cycle. (B) On warming to the restrictive (high) temperature, at which the mutant gene product functions abnormally, the mutant cells continue to progress through the cycle until they come to the specific step that they are unable to complete (initiation of S phase, in this example). Because the cdc mutants still continue to grow, they become abnormally large. By contrast, non-cdc mutants, if deficient in a process that is necessary throughout the cycle for biosynthesis and growth (such as ATP production), halt haphazardly at any stage of the cycle depending on when their biochemical reserves run out.

Figure 1.2 THE MORPHOLOGY OF BUDDING YEAST CELLS ARRESTED BY A CDC MUTATION (GRIFFITHS ET AL., 1996). (A) In a normal population of proliferating yeast cells, buds vary in size according to the cell-cycle stage. (B) In a cdc15 mutant, grown at the restrictive temperature, cells complete anaphase but cannot complete the exit from mitosis and cytokinesis. As a result, they arrest uniformly with the large buds, which are characteristic of late M phase.
Later on, ts mutations have also proved invaluable for the dissection of other diverse processes in microorganisms such as Chlamydomonas reinhardtii flagellar assembly (Adams et al., 1982; Huang et al., 1977) and protein folding (Goldenberg et al., 1983; Gordon and King, 1993; Gordon and King, 1994; Smith et al., 1980).

1.2.3 VARIOUS ASPECTS OF THE DEVELOPMENT OF CAENORHABDITIS ELEGANS HAVE BEEN STUDIED USING TEMPERATURE-SENSITIVE MUTANTS

A diversity of developmental processes has been dissected in the nematode Caenorhabditis elegans, in which ts mutations have played a crucial role. Most of these studies describe ts mutations causing cell division and proliferation deficiencies. Several large-scale screens for embryonic lethal mutants have identified mutations that affect various aspects of cell division in the early embryo (Hirsh and Vanderslice, 1976; Miwa et al., 1980; Cassada et al., 1981; Kemphues et al., 1988a, Kemphues et al., 1988b), while another screen has identified mutations that specifically affect the postembryonic lineages (Horvitz and Sulston, 1980). Both sets include mutations that disrupt the temporal and spatial patterns of cell division as well as mutations that block division altogether. As not all lineages are affected by these mutations, it is not known whether these genes encode key players in the mechanisms of division that are common to all cells.

More recently, O’Connell, Leys and White (1998) designed a genetic screen for ts cell-division mutants to identify nonlineage-specific cell-division mutations in Caenorhabditis elegans. They identified genes that encode key components of the cell-division machinery, and that are thus required for most cell divisions, and studied the effects of these mutations during the early embryonic divisions. They observed defects in chromosome segregation, spindle morphogenesis and alignment, centrosome duplication and cytokinesis. These results confirm that this approach is capable of identifying genes with a broad range of essential functions including roles in cell division.

1.2.4 STUDIES WITH CONDITIONAL LETHALS IN DROSOPHILA MELANOGASTER, REVEALED THAT TEMPERATURE-SENSITIVITY IS NOT A PROPERTY CONFINED TO MUTATIONS ON CERTAIN CHROMOSOMES

Ts mutations have been exploited to study a variety of genetic, developmental and neurological phenomena in the model insect Drosophila melanogaster (reviewed by Suzuki 1970; Suzuki, Kaufman and Falk, 1976; Hall, 1978a; Hall and Greenspan, 1979) (figure 1.3).

During these studies, it seemed that temperature-sensitivity is not a property confined to mutations on certain chromosomes or to a limited number of sites within a chromosome. Indeed, recessive and dominant ts mutations on the X-chromosome (Suzuki et al., 1967; Wright, 1968), Y-chromosomes (Ayles et al., 1972), chromosome 2 (Baillie, Suzuki and Tarasoff, 1968; Suzuki and Procunier, 1969) and chromosome 3 have been analyzed (Fattig and Rickoll, 1972; Holden and Suzuki, 1968). In addition, cold-sensitive recessive lethals on the X-chromosome (Mayoh and Suzuki 1972) and dominant
lethals on the autosomes were isolated (Holden and Suzuki, 1973; Rosenbluth, Ezell and Suzuki, 1972; Suzuki and Procunier, 1969; Tasaka and Suzuki, 1972).

Recently, Barry Ganetzky and Robert Kreber created a collection of more than 100 different ts neurological Drosophila mutants, which are behaviorally normal at 25°C, but display severe locomotor defects, including uncoordinated movement, ataxia, seizures or complete paralysis, within five minutes of exposure to 38°C. These mutants are being used in screening programs aimed at identifying novel pharmaceutical agents and drug targets in an in vivo biological system. They are also used in functional genomics research to identify novel genes involved in neural function and diseases and as experimental models for human neuromuscular disorders, including epilepsies, muscular dystrophies, periodic paralyses, myotonias, dystonias and others (Palladino, Hadley and Ganetzky, 2002).

1.2.5 PLANT TEMPERATURE-SENSITIVE MUTANTS AS TOOLS TO ANALYZE DEVELOPMENTAL AND GROWTH PROCESSES

1.2.5.1 SOMATIC EMBRYOGENESIS IN DAUCUS CAROTA (CARROT) CELLS

Ts mutations have been successfully exploited in research into somatic embryogenesis of cultured cells of carrot (Breton and Sung, 1982; Giuliano et al., 1984; Lo Schiavo et al., 1988, 1990; Schnall et al., 1988). Although cultures of carrot cells provide a unique experimental system for studies of somatic embryogenesis, which is a straightforward example of the plasticity of plant cells, further genetic analysis of mutants is difficult because of the poorly characterized genetic background of carrot.

1.2.5.2 PHOTOSYNTHESIS IN MEULLOTUS ALBA (SWEETCLOVER)

In sweetclover ts mutants were isolated in which the restrictive growth temperature produces a chlorophyll deficient phenotype (Bevins et al., 1993, Markwell et al., 1986; Yang et al., 1990). In the past, several approaches have been used in attempts to define the genetic control and biochemical nature of the photosynthetic apparatus in plants. The most frequently used mutants are defective in chloroplast structure, function, and/or pigment biosynthesis. Mutations resulting in the inability to express chlorophyll A are lethal once seed reserves are depleted because this pigment is absolutely required for photochemical energy conversion. When a severe deficiency of chlorophyll a is not lethal, growth is marginal and often results in poor seed production and difficulty in maintaining mutant lines. In contrast to the situation with chlorophyll a deficient mutants, chlorophyll b functions only in a light-harvesting capacity and many healthy mutants defective in the expression of chlorophyll b have been reported. Nevertheless, the utility of mutants that are temperature-sensitive for chlorophyll expression was corroborated since this study demonstrated that the relief of photosynthetic energy limitation that accompanies temperature-induced greening is directly correlated with progressive changes in carbon isotope fractionation, a sensitive physiological indicator. Therefore, the authors advise other researchers interested in chloroplast development to consider the use of ts mutants.

Also in Arabidopsis thaliana (Markwell and Osterman, 1992), Zea mays (corn), Gossypium hirsutum (cotton) and Glycine max (soybean) (Alberte et al., 1974) ts mutants have been used to study the photosynthetic apparatus.
1.2.5.3 **CELL GROWTH ANISOTROPY IN ARABIDOPSIS THALIANA**

Cortical microtubules and cellulose microfibrils have been widely accepted as primary determinants of growth anisotropy in plants (Wasteneys, 2000; Baskin, 2001). To support anisotropic growth, it is essential that microtubules and cellulose microfibrils are oriented at right angles to the main axis of cell expansion, especially in rapidly elongating cells. The cellulose synthase constraint model proposes that cortical microtubules control the direction of cellulose deposition by guiding the cellulose synthase enzyme complexes resident in the plasma membrane as they add cellulose microfibrils to the inner layer of the cell wall (Giddings and Staehelin, 1991). This hypothesis, however, is incompatible with some experimental evidence (Emons et al., 1992; Preston et al., 1988). The alignment of cortical microtubules and cellulose microfibrils is not always similar, even in diffusely expanding cells (Sugimoto et al., 2000).

A number of ts Arabidopsis mutants with cell expansion defects have started to provide insights into the cell wall structures that regulate plant cell expansion (Williamson et al., 2001). In a recent study Sugimoto et al. (2003) tested the cellulose synthase constraint model using two of these ts mutants, rsw1-1 and mor1-1 (figure 1.4). The RADIAL SWELLING1 (RSW1) gene encodes a cellulose synthase catalytic subunit AtCesA1 (Sugimoto et al., 2001; Williamson et al., 2001) while the MICROTUBULE ORGANIZATION1 (MOR1) gene encodes a microtubule associated protein (Whittington et al., 2001; Eleftheriou et al., 2005). They analyzed cellulose microfibril patterns in cells that had developed either with microtubules completely de-polymerized with the drug oryzalin, or disorganized in the mor1-1 mutant at its restrictive temperature. Cellulose microfibrils continued to be deposited in parallel order under these conditions. This result suggested that organized cortical microtubules are not required for oriented deposition of microfibrils and regulate growth anisotropy by some other mechanism. The independence of cortical microtubule organization and cellulose microfibril alignment was supported by the observation that double mutants of mor1-1 and rsw1-1, cellulose-deficient mutants with misaligned microfibrils, had additive phenotypes. Nevertheless, this study did not rule out the possibility that the deposition of microfibrils was simply...
maintained because of pre-existing well ordered microfibrils, which, once established, might serve as a template for the continued deposition of transverse, parallel microfibrils (Himmelspach et al., 2003).

The ts mutants mor-1, rsw1-1 and others like rsw-2 (Gillmor et al., 2002; Lane et al., 2001; Nicol et al., 1998), -3 (Burn et al., 2002; Peng et al., 2000), -4 and -7 (Wiedemeier et al., 2002) provide a powerful system in which to test the causal relationship between cortical microtubules, cellulose microfibrils and anisotropic cell expansion since it is possible to reversibly alter microtubule organization or cellulose production simply by elevating the temperature by a few degrees.

1.2.5.4 ORGANOGENESIS IN ARABIDOPSIS THALIANA

ROOT DEVELOPMENT

To investigate the genetic basis of adventitious organogenesis Konishi and Sugiyama (2003) isolated nine ts mutants defective in various stages of adventitious root formation. Five root initiation defective (rid) mutants failed to initiate the formation of root primordia. In one root primordium defective (rgd) mutant, the development of root primordia was arrested. The last three mutants were root growth defective (rgd) after the establishment of the root apical meristem (Sugiyama, 2003).

The temperature-sensitivity of callus formation revealed further distinctions between the isolated mutants. The rid1 mutant was specifically defective in the reinitiation of cell proliferation from hypocotyl explants, while rid2 was also defective in the reinitiation of cell proliferation from root explants. Furthermore, these two mutants exhibited abnormalities in the formation of the root apical meristem when lateral roots were induced at the restrictive temperature.

The rgd1 and rgd2 mutants were deficient in root and callus growth, whereas the rgd3 mutation specifically affected root growth.

The rid5 mutant required higher auxin concentrations for rooting at the restrictive temperature, implying a deficiency in auxin signaling. The rid5 phenotype was found to result from a mutation in the MOR1/GEM1 gene (§1.2.5.3) encoding a microtubule associated protein, suggesting a possible function of the microtubule system in auxin response.

SHOOT DEVELOPMENT

The isolation of three ts mutants in Arabidopsis that are defective in the redifferentiation of shoots from root explants has been described by Yasutani et al. (1994). In an attempt to genetically dissect shoot redifferentiation and at the same time also root redifferentiation the responses in tissue culture of the srd mutants were characterized (Ozawa et al., 1998). Shoot and root redifferentiation can be induced at high frequency in Arabidopsis from hypocotyls and root explants by a two-step culture method. Tissues are precultured on callus-inducing medium and then transferred onto shoot- (for shoot redifferentiation) or root- (for root redifferentiation) inducing medium. The srd mutants exhibited temperature-sensitivity at different steps of organogenesis, which allowed the identification of three states associated with organogenic competence: ‘IC’ (incompetent), ‘CR’ (competent with respect to root redifferentiation), ‘CSR’ (competent with respect to shoot and root redifferentiation) (Sugiyama, 1999).

Hypocotyl explants were shown to be in the ‘IC’ state at the initiation of culture and to enter the ‘CSR’ state, via the ‘CR’ state, during preculture on callus-inducing medium. Root explants seemed
to be in the ‘CR’ state at the initiation of culture. The transition from ‘IC’ to ‘CR’ and that from ‘CR’ to ‘CSR’ appeared to require the functions of SRD2 and SRD3, respectively. Moreover, explants in the ‘CSR’ state redifferentiate shoots with the aid of SRD1 and SRD2 when transplanted onto shoot-inducing medium. Histological examination of the srd mutants revealed that the function of SRD2 is required not only for organogenesis but also for the reinitiation of cell proliferation in hypocotyls explants during culture on callus-inducing medium. Linkage analysis using RFLP markers indicated that SRD1, SRD2 and SRD3 are located at the lower region, the central region and the upper region of chromosome I, respectively.

Making use of the srd mutants a hypothetical scheme for in vitro organogenesis of Arabidopsis thaliana could be set up. The process of organogenesis was subdivided into genetically distinct phases. These phases could be related to physiologically identified phases by tissue-transfer experiments. In addition, such mutants help to find elementary processes that are common to apparently different phenomena.

In another study Pickett et al. (1996) designed a screen for ts mutations that cause a conditional arrest of early shoot development in Arabidopsis to identify genes involved in meristem function. They describe the characterization of three mutations: add (arrested development) 1, 2 and 3 (figure 1.5). Analysis of the add mutant phenotypes indicates that these mutations are sensitive to temperature shifts at various points in development, suggesting that their activity is required throughout development.

At the restrictive temperature the add1 and add2 mutations disrupt apical meristem function as assayed by leaf initiation. Furthermore, add1 and add2 exhibit defects in leaf morphogenesis following upshift from permissive to restrictive temperature. This result suggests

---

**Figure 1.5 Wild-type and ADD Mutant Seedling Development After Germination and Growth at 29°C for 6 Days (Pickett et al., 1996)**: (A) Wild-type Wassilewskija seedling with cotyledons and two true leaves, (B) the apical region of an arrested development 1 (add1) seedling lacking true leaves, (C) wild-type Landsberg erecta seedling with an epicotyl displaying the first 2 true leaves. Both leaves have developed leaf blades and trichomes, (D) arrested development 2 seedling with cotyledons, true leaf primordia are not visible, (E) wild-type Columbia glabrous1 (gl1) seedling with 3 true leaves. The gl1 mutation prevents trichome development in this line; however, all true leaves have acquired dorsoventrality, (F) the apical region of an arrested development 3 mutant. Epicotyl development is delayed in this mutant because leaf blade expansion has not occurred. However, primordial leaves are asymmetric and blade development is apparent, suggesting that these leaves have acquired dorsoventrality. Scale bar=250 mm. c, cotyledon; tp, true leaf primordium; t1, leaf from first set of true leaves; tr, trichome; m, meristematic region.
that proximity to a functional meristem is required for normal leaf morphogenesis.

The add3 mutation does not have this effect but prevents the expansion of leaf blades at high temperature. In this mutant the temperature-dependent arrest of epicotyl development is due to a failure of normal leaf development rather than new leaf initiation.

Although all add mutants show a reduced rate of root growth in comparison to WT plants, temperature-dependent arrest of root growth is not observed. Given the complicated physiological interplay between shoot and root, the available data support a specific role for the ADD genes in regulating aspects of shoot meristem function or leaf development.

A different kind of ts phenotype, which involves shoot development, was observed in homozygous topless-1 (tpl-1) mutants (Long et al., 2002). At the restrictive temperature the apical pole of the embryo develops as a root instead of a shoot. The result is a seedling with two roots joined at the base without hypocotyls, cotyledons or shoot apical meristem. This mutation could provide an entrance to molecular genetic study of this important, but previously experimentally intractable, decision in plant development.

**FLOWER DEVELOPMENT**

**Ts** Arabidopsis mutants have been useful in studying the function of floral homeotic genes, which specify the identity of floral organs. There are three classes of floral homeotic genes, designated A, B, and C (Bowman et al., 1991; Coen and Meyerowitz, 1991). Class A alone specifies sepals, A and B together cause petal formation, B and C combined specify stamens and C alone directs carpel formation. With minor variations, the ABC model has been validated in different species, including Arabidopsis thaliana, Antirrhinum majus and Petunia hybrida. In Arabidopsis, function B is performed by a pair of homeotic genes: APETALA3 (AP3) and PISTILLATA (PI) (Goto and Meyerowitz, 1994; Hill and Lord, 1989; Jack et al., 1992). The AP3 and PI proteins belong to the MADS box family of transcription factors and bind DNA in vitro only as AP3/PI heterodimers, which may explain why mutations in either class B gene cause very similar phenotypes and why the ectopic B function requires combined expression of both AP3 and PI (Jack et al., 1994; Krizek and Meyerowitz, 1996).

Ts mutants have been isolated for AP3 (ap3-1 allele) and DEF (def-101 allele) (Bowman et al., 1989). Mutant plants grown at low temperature produce nearly WT flowers, but growth at the restrictive temperature causes the development of sepals (only A function present) and the replacement of stamens by carpelloid organs (only C function present). Such conditional mutants have been useful to verify that B homeotic function is required throughout organ development and to establish the stage in organ development up to which organ identity can be redirected by activation of a different homeotic function.

Similar results were obtained using ts mutations in the DEFICIENS (DEF) (ortholog of APETALA3) and the GLOBOSA (GLO) (ortholog of PISTILLATA) gene of Antirrhinum majus (Schwarz-Sommer et al., 1992; Sommer et al., 1990; Tröbner et al., 1992; Zachgo et al., 1995).

**1.2.6 As Functions and Interactions of Identified Proteins Come under Close Scrutiny, it Becomes More Important to Understand Which Properties of Mutant Proteins Cause Temperature-Sensitive Defects**

Both heat-sensitive (not able to grow at a higher restrictive temperature) and cold-sensitive (not able to grow at a
lower restrictive temperature) mutations have been isolated in a variety of organisms. Heat-sensitive mutations, commonly referred to as ts mutations, are generally due to production of a thermolabile protein or to temperature-sensitive protein synthesis, folding or, in the case of multimeric molecules, assembly. Cold-sensitive mutations are usually found in proteins, which participate in macromolecular complexes such as ribosomes or spliceosomes.

1.2.6.1 There are two general classes of temperature-sensitive mutations: those generating thermolabile proteins and those generating defects in protein synthesis, folding or assembly

Ts mutations were initially divided into two classes by Sadler and Novick (1965).

The first class was labeled ‘TL’ (thermolabile) and represented mutants in which the native protein is destabilized as a consequence of the temperature-sensitive amino acid substitution (Goldenberg, 1988). Examples of this type are ts mutants of lambda repressor (Hecht et al., 1984), yeast mitochondrial hsp60 (heat shock protein 60) (Martin et al., 1992) and LAR, a human receptor-like protein tyrosine phosphatase (Tsai et al., 1991). Such mutants typically show phenotypic defects when grown first at the permissive temperature for a certain period and transferred subsequently to the restrictive temperature. The basis of ‘TL’ mutations lies in the heat-sensitive nature of all proteins. When heated proteins denature, their tertiary folding structure is transformed into a variable, random-coil form. Protein folding is a process, which is accompanied by an unfavourable decrease in entropy and thus, driven by enthalpy. However, elevated temperatures increase the entropy term in the free energy equation \( \Delta G = \Delta H - T \Delta S \), and hence favour the random-coil form. Each protein has a specific temperature at which it unfolds, called the transition temperature. So, ‘TL’ mutations are missense mutations that result in a destabilized protein with a lower transition temperature. If the transition temperature falls within the normal growth range of the organism, then it is possible to grow the mutant organism at a low temperature, where the protein is folded and active, and at a high temperature, where the protein is unfolded and inactive.

The second class of ts mutations was labeled ‘TSS’ (temperature-sensitive synthesis). These mutations, like the ‘TL’ class, are defective when mutants are grown at the restrictive temperature. However, if mutants are grown at the permissive temperature, and then shifted to the restrictive temperature, they do not display phenotypic perturbances. The deficiency only occurs if the mutant protein was expressed and permitted to assemble at the restrictive temperature. The defects in gene product formation caused by ‘TSS’ mutations are varied. Most of the ts mutants of this type are defective in protein synthesis, folding or assembly. However, Singer and Gold (1976) described a ‘TSS’ mutation in the rII gene of phage T4 that probably acted at the level of messenger RNA structure or function and ts mutants of tRNAs exist (Smith et al., 1970). Mutants of this class include ts mutants of more than 30 genes of bacteriophage T4 (Edgar and Lielaulis, 1964), ts mutants of T4 DNA polymerase (Dewaard et al., 1965) and ts alleles at more than 30 sites in the P22 tailspike endorhamnosidase (Smith and King, 1981).

A systematic examination of ‘TSS’ mutations in the tailspike protein of phage P22 revealed that the primary defect was in the folding of the polypeptide chain (Goldenberg and King, 1981; Goldenberg et al., 1983; Sturtevant et al., 1989; Mitra et al., 1991). The temperature-sensitive
amino acid substitutions had no effect on the stability or activity of the thermostable tailspike protein, once matured at the permissive temperature. However, if expressed at the restrictive temperature, a folding intermediate was unable to continue productive folding. This subclass of ‘TSS’ mutations was subsequently called ‘TSF’ (temperature-sensitive folding) (Yu and King, 1984; King et al., 1989). Similar ‘TSF’ phenotypes have also been reported for mutants of D-lactate dehydrogenase (Truong et al., 1991) and bacterial luciferase (Sugihara and Baldwin, 1988).

The absence of intragenic complementation is a general property of ‘TSF’ (temperature-sensitive folding) mutations in genes encoding for multimeric proteins

Many sets of missense mutants, including ts mutants, exhibit intragenic complementation. This kind of complementation generally occurs in genes encoding different subunits of multimeric proteins and therefore provides a window into polypeptide chain domains and subunit interactions (Fincham, 1966).

Two general models have been proposed to explain intragenic complementation. First, if each allele has knocked out a different enzymatic function of the protein, the presence of both alleles may restore full protein activity (McGavin, 1968). Second, if the protein is oligomeric, the two alleles may help each other to correctly fold and assemble (Crick and Orgel, 1964).

Interaction of missense polypeptide chains, following this second model, has been demonstrated for many different oligomeric proteins (Schlesinger and Levinthal, 1963; Zabin and Villarejo, 1975). Significant intragenic complementation has been reported for ts mutations within the genes for a number of phage T4 structural proteins (Bernstein et al., 1965) and ts mutations in the SV40 major coat protein (Behm et al., 1988). Therefore, it is assumed that direct (subunit) interaction corrects the mutant defects and hence provides the mechanism for intragenic complementation.

However, Gordon and King (1994) found that, despite the intimate interactions between coat protein molecules of phage P22 during the subunit assembly process, ‘TSF’ coat protein mutants did not show any intragenic complementation. They were also recessive. These genetic properties reflect the nature of the intracellular defect: defective folding of the polypeptide chain. The same recessive character and lack of intragenic complementation has been found with ‘TSF’ mutants of the trimeric P22 tailspike protein (Smith et al., 1980; Goldenberg et al., 1983). These properties are consistent with the idea that the mutant coat protein molecules are not defective in the stability (or function) of the folded molecule, but of an assembly-incompetent precursor. Subunit/subunit interaction between different alleles is expected to offer the possibility of improving the stability or function of folded assembly-competent molecules, but not of destabilized ‘TSF’ folding intermediates.

Therefore, a ‘TSF’ defect can be defined as the destabilization of a folding intermediate at the restrictive temperature, which prevents the mutant chain from reaching the conformation required for subunit/subunit recognition.

Selection of second-site suppressors of temperature-sensitive mutants has provided a general approach for identifying genes of which products interact with an initial protein of interest

For conditional-lethal mutants, restrictive conditions provide an absolute selection for further mutation, since only revertant individuals can grow. Such revertants need not be true WT, rather, they may have acquired suppressors, new mutations that act in a way that corrects, replaces, or bypasses the original defect.
The distinction between ‘TSF’ and ts mutations that perturb the function of the already folded protein (‘TL’ (thermolabile) or assembly-defective ‘TSS’ (temperature-sensitive synthesis)) is likely to be particularly important in the isolation of second-site suppressor mutations (figure 1.6). If the initial mutation impairs the function of the active intracellular conformation (‘TL’ or assembly-defective ‘TSS’), gene products interacting with the mature form of the protein can be isolated. However, if the initial mutation destabilizes a folding intermediate such selection may yield sites influencing folding and maturation pathways (Fane and King, 1991).

Global suppressors are a type of second-site suppressors that counter the deleterious effects of a wide range of protein folding and stability mutations. So, the characterization of mutants that alter folding and stability of proteins may help identify residues that specify critical steps in the folding pathway or ones that highlight key interactions in the native structure. In a selection for second-site suppressors of tailspike ‘TSF’ mutations global intragenic suppressors were obtained (Fane et al., 1991; Mitraki et al., 1991). By suppressing the pathway to inclusion body formation, these substitutions improved the folding efficiency of different ‘TSF’ mutations, as well as of the WT tailspike protein.

1.2.6.2 Cold-Sensitive Mutations Involve an Altered Protein Conformation That Is Still Able to Interact at Higher Temperatures but Fails to Do So at Lower Temperatures

Cold-sensitive mutations affect the ability of proteins to form multimeric or macromolecular complexes. Despite their altered conformation as a consequence of an amino acid change, these proteins are able to function and interact at higher temperature. In fact, due to the increased kinetic activity, the complex can adjust to accommodate the mutant protein. However, at lower temperatures the
altered shape and slower movement of the mutant protein prevents it from interacting efficiently (O’Donovan and Ingraham, 1965).

The remarkable property of self-assembly of ribosomal proteins is highly dependent upon temperature, an observation that led Nomura to the brilliant decision to screen cold-sensitive lethals of *Escherichia coli* for defects in ribosomal proteins (Guthrie et al., 1969). It was found that a significant proportion of cold-sensitive lethal mutations in this organism do indeed encode defective ribosomal proteins.

Although cold-sensitive mutations are isolated to a lesser extent, they are included in several other specific studies such as sporulation (Esposito and Esposito, 1968) and mitochondrial petite (Weisblum and Davies, 1968) in yeast and developmental mutants in *Drosophila* (Foster and Suzuki, 1970).

### 1.2.6.3 Temperature- and Cold-Sensitive Mutations Differ Strongly in Genetic Distribution and Target Genes

Edgar and Lielausis (1964) found that ts mutations were widely distributed over the genome of bacteriophage T4, although they were not distributed completely randomly. They observed long segments in which they didn’t find any ts mutations, while other regions were densely populated with such mutations. However, they showed that modifications of the isolation procedures did yield ts mutations located in otherwise sparsely populated segments of the genome. Thus, it would appear probable that many of the blank segments were regions of low detectable mutability rather than regions in which ts mutations cannot occur. In *Drosophila* it was observed that, given a proper screening protocol, temperature-sensitivity is a property that can be readily detected for a wide spectrum of loci throughout the genome (Suzuki, 1970). As a consequence, ts lethals are defective with respect to a wide range of functions.

In contrast, cold-sensitive mutants are remarkably clustered in a small number of cistrons (Scotti, 1968; Cox and Strack, 1970) and involve changes primarily in regions of proteins concerned in regulatory functions. Striking is the case of ribosome assembly in *Escherichia coli*, where extensive searches for ts mutants were fruitless while cold-sensitive mutants were readily obtained (Tai et al., 1969).

### 1.2.7 Different Strategies can be Followed to Generate Temperature-Sensitive Mutants

#### 1.2.7.1 Temperature-Sensitive Mutants are Typically Generated by Random Mutagenesis Followed by Laborious Screening Procedures

Ts mutants are usually point mutations, which are often generated by random mutagenesis followed by screening of large numbers of progeny. In most of the organisms studied, the highest frequency (25-50%) is observed when chemical mutagenesis is applied using the alkylating agent EMS (ethyl methane sulfonate). γ-ray-induced ts mutations show a frequency of 3%-3.5%. In the case of prokaryotes and yeast, such procedures work well because simple screens or selection methods exist and a large number of progeny can be simultaneously screened using simple plate assays. In more complicated organisms, however, such an approach suffers from several drawbacks. Since mutations will be generated throughout the genome, a large number of progeny need to be screened before a ts mutant is obtained. In the case of *Drosophila melanogaster* and *Arabidopsis thaliana*, this number is typically of the order of several hundred thousand. Screening such a large number of progeny can be extremely laborious in situations where a simple screen does not
exist. Furthermore, in organisms with long generation times or in cases where it is not possible to obtain large numbers of progeny, random mutagenesis of the entire genome cannot be used to isolate ts mutants.

Identifying and recovering ts mutations is also hindered by their rarity, a particular problem in higher organisms. It is also clear that ts alleles cannot be generated for all proteins and some remain refractory even following extensive mutagenesis (Harris et al., 1992).

1.2.7.2 Heat-Inducible Degron: An Alternative Method for Constructing Temperature-Sensitive Mutants

Features of proteins that confer metabolic instability are called degradation signals, or degrons (Varshavsky, 1991). The essential component of one degradation signal, the first to be discovered, is a destabilizing N-terminal residue of a protein (Bachmair et al., 1986; Varshavsky et al., 1992, 1996a). This signal is called the N-degron. The collection of N-degrons containing different destabilizing residues yields a rule, termed the N-end rule of protein degradation, which relates the in vivo half-life of a protein to the identity of its N-terminal residue. The N-end rule pathway is present in all organisms examined, including the bacterium Escherichia coli (Tobias et al., 1991; Shrader et al., 1993), the yeast Saccharomyces cerevisiae (figure 1.7; Bachmair and Varhavsky, 1989) and mammalian cells (Gonda et al., 1989; Lévy et al., 1996). In eukaryotes, the N-degron consists of at least two determinants: a destabilizing N-terminal residue and a specific internal lysine residue (or residues) of a substrate. The lysine residue is the site of attachment of a multiubiquitin chain. Ubiquitin is a 76-residue protein whose covalent conjugation to other proteins plays a role in a number of cellular processes, primarily through routes that involve protein degradation.

Dohmen et al. (1994) devised a method for constructing ts mutants, which involves making fusions of ubiquitin-ts-arginin dihydrofolate reductase to the protein of interest (figure 1.8). Arginin-dihydrofolate reductase (Arg-DHFR) bearing an amino-terminal arginin is long-lived in the yeast Saccharomyces cerevisiae, even though arginin is a destabilizing residue in the N-end rule of protein degradation. A temperature-sensitive derivative of Arg-DHFR (Arg-DHFR-ts) was identified that is long-lived at 23 °C but rapidly degraded by the amino-end rule pathway at 37 °C. Fusions of ts Arg-DHFR to either Ura3 (orotidine 5'-phosphate decarboxylase) or Cdc28 (the main cyclin-dependent kinase involved in cell cycle control in budding yeast) of Saccharomyces cerevisiae confer ts phenotypes specific for these gene products. Thus, Arg-DHFR-ts is a heat-inducible degradation signal that can be used to produce ts mutants without a mutational screening.

Although this is a powerful and elegant approach, it remains to be seen how well it works in other organisms. In some cases, it is possible that the fusion partners may also affect the normal functioning of the protein at temperatures below the restrictive temperature.

1.2.7.3 For Globular Proteins Temperature-Sensitive Mutants Can Be Predicted and Designed Solely Based on the Amino Acid Sequence

A method to predict those residues in a protein sequence that, when appropriately mutated, are most likely to give rise to a ts phenotype was described by Varadarajan et al. (1996). Their approach was based on the fact that mutations at buried residue positions cause much larger changes in protein stability than mutations at surface positions (Pakula et al., 1986; Kellis et al., 1989; Shortle et al., 1990). Decreased
protein stability is correlated with reduction of the protein levels in vivo and with the generation of a ts phenotype (Parsell et al., 1989; Pakula et al., 1989). The strategy has therefore been to predict those residues in the sequences that are likely to be buried in the protein structure. Values of two sequence-based parameters, the average hydrophobicity (Rose and Roy, 1980; Hopp and Woods, 1981) and the hydrophobic moment (Eisenberg, 1984), were used for the prediction. The average hydrophobicity is calculated over a seven-residue window centered around the residue of interest. The hydrophobic moment is a vectorial sum of hydrophobicities calculated over a nine-residue window, with a phase angle optimized for detection of amphiphilic helical structures, which typically would not have a high hydrophobicity. Stringent cut-offs for these two parameters ensured that it was possible to predict buried hydrophobic residues with an accuracy >80%. The method successfully predicted several known ts and partially active mutants of T4 lysozyme, λ repressor, gene V protein and staphylococcal nuclease. Furthermore, it correctly predicts residues that form part of the hydrophobic cores of
FIGURE 1.8 THE HEAT-INDUCIBLE DEGRON CAUSES A PROTEIN TO BE DEGRADED AT 37°C (SANchez-DIAz ET AL., 2004).

λ repressor, myoglobin and cytochrome b562.

It should be noted that the authors did not attempt to predict all possible ts mutants. Ts mutations can result from substitutions at non-core residues and at surface positions that may be important for folding, proteolytic sensitivity or interaction with other proteins. Rather, the method is designed to yield ts mutants at sufficiently high frequencies to be experimentally useful in systems in which it is not feasible to screen large numbers of progeny generated by random mutagenesis.

In a more recent report Chakshusmathi et al. (2004) tested this algorithm to further refine and improve the method. Initially, they used Escherichia coli cytotoxin CcdB as a model system and developed simple procedures for generating ts mutants at high frequency through site-directed mutagenesis. Putative buried, hydrophobic residues are selected through analysis of the protein sequence, as described above. Residue burial is then confirmed by ensuring that substitution of the residue by asparagin leads to protein inactivation. At such sites, a ts phenotype can typically be generated either by substitution of two predicted, buried residues with the remaining 18 amino acids or by introduction of lysin, serin, alanin and tryptophan at three to four predicted buried sites. By using these design strategies, 17 tight ts mutants of CcdB were isolated at four predicted buried sites. The rules were further validated by making several ts mutants of the yeast transcriptional activator Gal4 at residues 68, 69 and 70. Gal4 binds to a sequence called UAS (upstream activating sequence consisting of tandem 17 base pairs, imperfect repeats) and activates transcription of genes downstream of this sequence (Bram and Kornberg, 1985; Fedor et al., 1988; Giniger et al., 1985). Gal4 is absent in higher eukaryotes. However, if expressed under control of a tissue- or developmental stage-specific promoter, the gene placed downstream of UAS will be expressed in a Gal4-dependent pattern. The UAS-Gal4 system is widely used in cell culture and in a variety of organisms, including yeast, fruitflies, zebrafish, mice and frogs (Andrulis et al., 1998; Bernard et al., 1993; Brand and Perrimon, 1993; Hartley et al., 2002; Ornitz et al., 1991; Rahner et al., 1996; Scheer and Carnos-Orthega, 1999). No ts mutants of Gal4 have currently been reported. Such mutants would be very useful, because they would permit reversible, conditional expression of the very large number of UAS constructs that have already been made.
1.2.7.4 Temperature-Sensitive Mutant Inteins Inserted Within Two Transcription Factors Provide a New Tool for Controlling Gene Expression in Yeast and Flies

Zeidler et al. (2004) combined ts inteins - protein elements capable of self-excision from a precursor protein and ligation of the flanking protein regions - with the yeast transcription factors Gal4 and Gal80 to create a system that overcomes the problems associated with the isolation of ts mutants using random mutagenesis (figure 1.9).

Intein-mediated protein splicing is a post-translational, autocatalytic reaction requiring no known host cell cofactors, making inteins very portable. However, each intein is more or less sensitive to proximal host protein sequences, which are the substrate of the intein enzymatic activity (Chong et al., 1998; Noren et al., 2000; Southworth et al., 1999). Over 170 inteins have been catalogued in InBase, the intein database (Perler, 2002). In naturally occurring precursors, protein splicing is usually required for host protein activity. However, when inserting inteins into new sites, host protein activity of the precursor and the spliced product should be tested using inactive and active inteins, respectively. The positional effects of protein splicing in foreign contexts can often be overcome by trying several intein insertion sites. Zeidler et al. have eliminated this problem by inserting temperature-dependent splicing mutants of the Saccharomyces cerevisiae vacuolar ATPase subunit (VMA) intein into transcriptional regulators that control expression of essentially any target protein.

The transcriptional regulators they chose to modify were Gal4 and Gal80. Gal80 functions by binding to Gal4 and blocking the latter's ability to activate transcription. Zeidler et al. demonstrated intein-mediated conditional target gene expression in yeast, Drosophila melanogaster tissue culture cells and in transgenic Drosophila melanogaster imaginal discs (cell cluster progenitors of adult organs). This method is readily transferable to the multitude of genes currently regulated by Gal4.

They initially genetically selected temperature-dependent splicing mutants of the S. cerevisiae VMA intein in Gal4 (Gal4INT<sub>ts</sub>) in yeast. These mutant inteins were then transferred to two sites in Gal80 (Gal80INT<sub>ts</sub>). Efficient, temperature-dependent splicing of both Gal80INT<sub>ts</sub> constructs was observed in yeast, but only one Gal80INT<sub>ts</sub> fusion spliced efficiently in D. melanogaster cells, reflecting the complexity of generating a universal ts intein. This is possibly the first example of transfer of ts splicing to a second host protein, indicating that the ts phenotype is intrinsic to the intein.

Transcription of a target gene is stimulated when Gal4 binds to an upstream activator sequence (UAS) inserted 5' to the gene (see also §1.2.7.3). A second layer of control is added by Gal80, which represses transcription of the target gene by binding to Gal4. When the ts intein is present in Gal4, target gene expression is activated only at the permissive-splicing temperature. When the ts intein is present in Gal80, Gal4 activity is only repressed at the permissive-splicing temperature.

Using this system, any Gal4 line can be directly converted into a conditional expression system by merely crossing with a Gal80INT<sub>ts</sub> or Gal4INT<sub>ts</sub> line. The combination of available temporal-, tissue- and stage-specific expression pattern drivers provide a multitude of combinations of Gal4 and Gal80INT<sub>ts</sub> interactions, resulting in an enormous diversity of potential target protein expression patterns. The ability to activate Gal80INT<sub>ts</sub> by shifting to the permissive-splicing temperature provides another layer of control.

The elegance of this system was demonstrated in the transgenic Drosophila melanogaster experiments. This is the first example of intein-mediated protein splicing in a multicellular animal (as opposed to splicing in plants or pathogenic bacteria, viruses and yeast that invade complex organisms).
Figure 1.9 Conditional target gene expression controlled by temperature-sensitive splicing of intein fusions (Zeidler et al., 2004). Gal4 is a two-domain protein that activates transcription of genes preceded by an upstream activator sequence (UAS). The Gal4 N-terminal binding domain (BD) binds to UAS, whereas the carboxy-terminal activation domain (AD) stimulates transcription of the target gene. Gal80 represses Gal4 by binding to the Gal4 activation domain. (a) A temperature-sensitive splicing mutant of the S. cerevisiae VMA intein (I) was cloned into Gal4. Spliced Gal4 is fully functional, whereas the Gal4INTts precursor fails to activate transcription of the target gene. (b) A temperature-sensitive splicing mutant of the S. cerevisiae VMA intein (I) was cloned into Gal80. The Gal80INTts precursor has no effect on Gal4, whereas spliced Gal80 binds to the transcription activation domain (AD) of Gal4 repressing transcription of the target gene.

Expression of UAS-green fluorescent protein (GFP) reflected the spatial location and temperature-dependent splicing of aperus enhancer driven Gal4INTts. In a second experiment, Gal80INTts was under the control of the decapentaplegic enhancer and Gal4 was under control of the patched enhancer, yielding different spatial patterns of expression of the repressor and activator in imaginal discs. Gal4 activation of UAS-GFP and UAS-Notch was dependent on the spatial location of both Gal4 and Gal80, and on temperature-dependent splicing of Gal80INTts.

The ts intein system is directly applicable to any organism that uses Gal4 activation. It is limited to cells or organisms that can undergo the requisite temperature shift, making it impractical in many warm-blooded animals. However, other ways of controlling protein splicing exist or are in development. These include splitting the precursor and splicing in trans (Noren et al., 2000) or using small molecules to activate or inhibit splicing (Mootz and Muir, 2002). Several genetic selection systems are now available to screen for ts inteins and protein splicing inhibitors and activators (Paulus, 2003).

One of the most significant contributions of the authors is the concept of controlling a universal regulator with ts inteins. The plethora of Gal4 and UAS lines currently available and easily adapted to this system makes the conditional splicing of Gal4INTts and Gal80INTts extremely important new genetic tools. They have demonstrated conditional protein splicing in a complex eukaryote and have identified ts mutants that could possibly be transferred to other host proteins, providing a framework for making other ts regulator proteins.

Although the above-described approaches offer elegant alternatives to overcome the problems associated with random mutagenesis, the latter is still the most widely used method to induce (possibly ts) point mutations. Despite the mentioned drawbacks, it is a technically and experimentally simple and effective procedure that has been used for more than 60 years now in all kinds of organisms.

1.3 Establishing Gene Function Through Mutagenesis

Gene function can be revealed through analysis of mutants by either forward or reverse genetics. The methodologies that are currently available to obtain mutants with desired phenotypes or mutations in a gene of interest and to clone a mutant gene will be discussed in detail for the model plant Arabidopsis thaliana.
1.3.1 Arabidopsis thaliana as an Experimental System for Molecular Genetic Studies

The small mustard weed Arabidopsis thaliana (family Brassicaceae) was known to botanists for at least four centuries and had been used in experimental research for about half a century, before Friedrich Laibach outlined the advantages of using it in genetic experiments (figure 1.10; Laibach, 1943; Rédei, 1992). In 1943 he also suggested that it could be used as a plant model system. Several features make Arabidopsis amenable to classical experimental genetics: a small size, a rapid generation time (5-6 weeks under optimum conditions), the ability to grow well under controlled conditions (either in soil or in defined media), high fecundity (up to 10,000 seeds per plant) and the ease with which a mutant line can be maintained (by self-fertilization) and out-crossed.

As well as having all the key features of a model system for classical genetics, Arabidopsis has the smallest known genome among flowering plants (125Mb), with fewer repetitive sequences than any known higher plant, greatly facilitating molecular studies and map-based cloning (Leutwiler et al., 1984; Pruitt and Meyerowitz, 1986; The Arabidopsis Genome Initiative, 2000). In addition, Arabidopsis can easily be transformed by Agrobacterium-mediated gene transfer, which is a prerequisite for many molecular genetic experiments. The advantages of using Arabidopsis in such experiments have been reviewed extensively (Meinke et al., 1998; Meyerowitz, 1989; Meyerowitz and Pruitt, 1985). Although Arabidopsis had been used for classical genetic studies for a long time, its role was not comparable with that of Drosophila in the systematic analysis of developmental and metabolic processes. Only once it was recognized that Arabidopsis had a small genome and, therefore, the potential to combine classical genetics with molecular studies, did it advance to being the most studied higher plant. The success of Arabidopsis as a model organism is mainly due to its amenability to forward genetic screens, by which genetic variation is artificially induced and mutagenized plants are screened for phenotypes of interest. Mutagenesis in Arabidopsis mainly relies on ethylmethane sulphonate (EMS) and, to a much lesser extent, fast neutron radiation (FNR). The study culminates in the molecular isolation of the mutated gene, which causes the observed phenotype. So, a function can be assigned to a molecular change in the inherited material of an individual. As more scientists joined the Arabidopsis community, experience was gained and tools were developed to carry out these screens with the aim of isolating genes that are involved in all aspects of plant life.

1.3.2 Forward Genetics: From Phenotype to Genotype

Forward genetics aims to identify the sequence change that underlies a specific mutant phenotype. The mutant phenotype can be the result of deliberate mutagenesis or be defined on the basis of existing variation. When the mutation
causing the phenotype is the result of a T-DNA or transposon insertion, rapid identification of the gene of interest is at least theoretically possible by locating the sequence tag and analyzing its neighboring sequences. However, the identification of a gene affected by a chemically or radiation-induced mutation requires a more laborious map-based cloning approach in which markers linked to the mutated gene are used to delimit the region containing the gene of interest (Østergaard and Yanofsky, 2004).

1.3.2.1 Mutagenesis in Arabidopsis

In order to conduct genetic analysis, at least two alleles for a given locus must exist and the facile production of these alleles through mutagenesis has been the focus of considerable research since the early part of the 20th century (Koornneef, 2002).

Classical Mutagenesis

Most of the mutations available in Arabidopsis thaliana have been induced by some form of ionizing radiation or by chemical agents. Since the frequency of insertional mutagenesis is generally lower, it may be advantageous to first define a locus by these classical methods. Once one knows that a particular kind of mutation is possible, it is then practical to plan a strategy for the isolation of the corresponding gene.

The generation of plants, grown from the mutagen-treated parental genotype, is called the M₁ generation and is heterozygous for mutations. The embryos within the seeds formed on such plants and the plants that grow out of these seeds represent the M₂ generation. The M₂ is the first generation following mutagenesis in which homozygous recessive mutations can be detected. For this reason, it is the generation most frequently used in mutant screens (Feldmann et al., 1994).

Plants from Treated Seeds Are Chimeras

Seeds are the most commonly mutagenized Arabidopsis tissue, because of their ready availability, robustness in mutagenesis and ease of propagation (figure 1.11). Each cell within the original treated seed is mutagenized in its own unique way. Thus, different cells of the same seed will contain different mutations. The M₁ plant that sprouts from the seed is therefore a chimera of genetically different cell lineages, each of which we would expect to be heterozygous for whatever new mutations have been induced. The population of M₂ seeds from a given M₁ plant contains both heterozygotes and homozygotes for any mutation that was induced in a cell whose descendants gave rise to gametes.

Fortunately, a small number of cells in the mature seed give rise to the body of the plant and subsequently contribute to the next generation. So, only those mutations that occurred in cells that form the germ line (precursor cells of the gametes) can be detected in the M₂ generation. The number of cells that contribute to the germ line was called the genetically effective cell number (GECN) by Li and Redei (1969) and was estimated to be, on average, two in Arabidopsis. The consequence is that half of the progeny of an individual M₁ plant derives from a sector in which a specific gene (A) is not mutated (AA) and the other half from the sector that was heterozygous for this gene (Aa). The segregation ratio observed for aa in the progeny of this chimeric plant is then 7A (AA + Aa) : 1aa. This 7 : 1 ratio is obtained from a 4 : 0 ratio from the non-mutated sector and a 3 : 1 ratio from the sector carrying the mutated allele a. The sectors that originate from the germ line cells present in seeds are not identical in size and their contribution to the main inflorescence on which seeds are harvested, changes during development. Parts of the plant that are formed later (the top) are mostly composed of only one sector from only one germ line cell.
The most practical way to induce heritable mutations in Arabidopsis is by seed mutagenesis (1). The targets of seed mutagenesis are the diploid cells of the fully developed embryo (2) covered by the seed coat. To assess the effectiveness of mutagenesis, it is crucial to know how many of the targeted cells will eventually contribute to the next generation. Although plants do not have a determined germ line, in practice, the functional germ line can be defined as the number of cells in the shoot meristem of the embryo (3) that will contribute to the seed output (the genetically effective cell number [GECN]). The GECN is determined genetically and might vary between different species and for different developmental stages. If the segregation in the M2, the generation derived from mutagenized M1 plants, is 3:1, the functional germ line consists of a single cell at the time of mutagenesis (GECN=1). In Arabidopsis, recessive mutants segregate in a ratio of 7:1 in an M2 population (4), therefore, GECN = 2. Because the functional germ line consists of two cells at the time of mutagenesis, the developing M1 (which develops from the mutagenized seed) is chimeric and consists of two sectors, which might vary in size (5). Mutations segregate 4:0 (6) or 3:1 (7), depending on the sector from which the M2 seeds derive. To recover a potential mutation, it is therefore necessary to analyse M2 seeds from different branches of the M1 plant.

The consequence is that variation occurs between M1 plants and that the harvesting policy also determines how many sectors contribute to the progeny. Chimerism disappears when the mutation is transmitted through the progeny because every individual of this progeny is derived from a single zygote cell.

**To Pool or Not to Pool?**

To reduce labor, M2 seed is often collected from pools of M1 plants, which is a good strategy for mutations that do not cause death or sterility when homozygous. If the mutations of interest are expected to cause lethal or sterile phenotypes, then M2 seeds should be collected from individual M1 plants. This way, when an interesting mutant is found among the progeny of a particular M1, plants heterozygous for the mutation can be recovered from the siblings of the dead or sterile plant. Another advantage of collecting seeds from individual plants is that it prevents the same allele from being recovered more than once. If two plants with the same phenotype are found in the same pool, there is a good chance that they will carry the same mutation. Plants from different pools will carry independent mutations. So, the higher number of M1 pools, the more independent alleles will be recovered.

**Physical Versus Chemical Mutagens**

The most commonly used physical mutagens are ionizing radiation (γ- and X-rays) and fast neutron bombardment. Each type of radiation produces deletions at a high frequency. The deletions are supposed to be larger after fast neutron radiation. The latter type of radiation yields densely ionized tracks compared to the mainly sparse ionizations caused by γ- and X-rays. For this reason fast neutrons are sometimes preferred. However, facilities where one can apply this type of radiation are rare in comparison to sources of γ- and X-rays. Deletions are attractive for many experiments because the mutants are often true nulls. In addition, a deletion is often detectable with Southern blot and PCR analysis, which facilitates the recognition of the gene to be cloned using map-based cloning. Although large deletions (parts of chromosome arms) can occur, they are rarely observed in M2 populations because a large deletion in most cases affects the viability of germinating pollen grains that then will not contribute to fertilization. This effect of reduced transmission of mutated pollen is called certation.
Chemical mutagenesis was developed by Charlotte Auerbach and her collaborators in England prior to and during the Second World War as part of a chemical warfare program. Chemical mutagens can be divided into two classes: the alkylating agents and the cross-linking agents.

The alkylating mutagens cause high point mutational densities with only low levels of chromosome breaks that would cause aneuploidy, reduced fertility and dominant lethality. Ethyl methane sulphonate (EMS) has been the most commonly used mutagen in plants and induces large numbers of recessive mutations per genome. Other alkylating agents such as ethynitrosourea (ENU) have also been used effectively (Gilchrist and Haughn, 2005). They typically introduce small alkyl groups onto the bases themselves, thus modifying their capacity for base pairing. An alkylated guanine (G) pairs with thymine (T), rather than with cytosine (C) (Ashburner, 1990). Thus, if this lesion is not repaired before the next replication, a T will be incorporated instead of a C. Following the second replication event, the original GC base pair will be replaced by an AT, resulting in GC → AT transition. Modification of T by some alkylating agents can also produce TA → CG transitions by a similar mechanism. Accordingly, a substantial fraction of the mutants produced by alkylating mutagens are either missense or nonsense mutations (Greene et al., 2003; Henikoff et al., 2003).

The second class of chemical mutagens is the cross-linking agents. Cross-linking mutagens such as diepoxybutane (DEB) and cis-platinum (II) diaminodichloride (cis-platinum) can chemically interlock two sites on the DNA molecules. Cross-links that connect two sites on the same strand are known as intra-strand cross-links and those that connect opposite strands are referred to as inter-strand cross-links. Both types of cross-links prevent replication and are difficult for the cell to repair. When incorrectly repaired, such lesions often produce small deletions.

Physical and chemical alkylating mutagens have a different mutant spectrum, which is most obvious when one compares the ratio $M_1$ sterility versus embryolethal and chlorophyll mutants. For instance, in an experiment where EMS and X-rays were compared, an EMS treatment resulting in 44% sterility (percentage reduction of fertilized ovules compared to control) gave 18.2% embryonic lethals and 9.4% chlorophyll mutants, whereas an X-ray treatment that gave almost the same percentage of sterility produced only 5.9% embryo-lethals and 1.0% chlorophyll mutants (Van der Veen, 1966). This relative high sterility is due to the fact that irradiation leads to chromosome breaks and chromosomal aberrations, giving rise to meiotic disturbances and, therefore, sterility. Base pair substitutions caused by alkylating agents may lead to specific amino acid changes, which may alter the function of proteins but do not abolish their functions as deletions and frame-shift mutations mostly do. Plant species differ in their sensitivity for specific chemicals. In barley, for example, sodium azide is a very potent mutagen, but it is hardly effective in Arabidopsis, where EMS is preferred (Gichner and Veleminsky, 1977; Kleinhofs et al., 1978). However, several other chemicals have shown to be effective in Arabidopsis.

Mutagen Dose

The higher the mutagen dose, the more efficient the experiment, and the higher the number of possible mutants. However, there are good reasons not to use the highest mutagen dose possible, since this also will lead to an increase in unwanted mutations at other loci, many of which lead to sterility or even lethality. This makes a relatively low dose attractive in those cases where the amount of work and costs involved in growing more $M_1$
plants and screening more M$_2$ plants is not excessive. In those cases where it is expensive to grow many M$_1$ plants and when laborious mutant screens are applied, the costs involved justify a higher dose. This approach then leads to more ‘dirty’ mutants that need ‘cleaning up’ by subsequent backcrosses to the WT.

A simple theoretical model (Malmberg, 1993) predicts that the optimum yield of mutations for a given mutagen should be the dose when 37% of the M$_1$ seeds survive to reproduction and contribute to the M$_2$. This may be a useful starting point but the ultimate measure of success is the production of mutants. Two traits for calibrating the effectiveness of mutagen and dose are the frequencies of embryonic lethals and the production of albino seedlings. Both traits are readily scored and should be abundant, as they represent multilocus targets. Embryonic lethals can be scored directly in the siliques of the M$_1$ plants, appearing as a 3:1 segregation of pale white embryos instead of the normal green. After EMS mutagenesis, the percentage of M$_1$ plants segregating embryonic lethals can be as high as 5-10%.

**Mutation Frequencies**

Mutation frequencies (m) should be based on the frequency at which mutations occur per treated cell. Usually, these are expressed per locus or per group of mutants. In practice, these estimates are based on the estimate of the number of mutants found in the M$_2$ in relation to the number of M$_2$ plants that are screened. It is also possible to base the estimate on the number of M$_1$ plants whose progeny were screened for mutants.

The method of Gaul (Koornneef et al., 1982; Yonezawa and Yamagata, 1975) is the most simple and expresses the number of mutants found per number of M$_2$ plants divided by f, the average mutant frequency in the progeny of a heterozygote (f = 0.25 for recessive mutations). Therefore, m = m’ / n x f in which m’ is the number of mutants found per n M$_2$ plants screened. This estimate is independent of the number of M$_1$ plants from which these are derived and also from the degree of chimerism. By dividing m by two the frequency of mutation per haploid genome can be obtained.

Li and Rédei (1969) derived the same frequency from the number of M$_1$ plants whose progeny were tested. This procedure takes into account the GECN and also requires that all mutations will be detected and that, therefore, the size of individual M$_1$ progenies should be large enough not to miss the mutant by chance. When GECN = 2, the mutant occurs only at a frequency of 1/8 which requires 23 plants not to miss the aa plant by chance in a given progeny at P<0.05. According to the method of Li and Rédei the mutation frequency per cell is calculated with the formula: m = M / (S x GECN), where M is the number of M$_1$ progeny segregating for a mutation (type) among S progeny that were screened.

Haughn and Sommerville (1987) have estimated that an M$_1$ population size of roughly 125,000 plants, after EMS mutagenesis, will effectively saturate the genome for all possible EMS-inducible base changes. Rédei (1975) describes similar calculations to optimize mutant isolation strategies.

Using various EMS-based protocols, mutation frequencies ranging from 1/300 to 1/30,000 per locus have been reported. On the basis of these mutation rates, a GECN value of 2 and the Poisson distribution (Rédei and Koncz, 1992), the number of M$_1$ plants required is ~700 and ~70,000 respectively.

In practice, many screens are based on 2,000-3,000 M$_1$ plants after exposure to 15-20 mM EMS for 12 hours, which is at sub-saturation level but typically yields three or more alleles per locus for easily scorable floral mutants (Bowman, non published results).
Although a detailed explanation of insertional mutagenesis goes beyond the scope of this thesis, a brief mention, referring to some excellent reviews on this subject, is warranted.

Currently, T-DNA and transposons are the two main insertional mutagens widely used for gene disruption in Arabidopsis (Fedoroff, 1989; Errampalli et al., 1991; Bancroft and Dean, 1993a; Fedoroff and Smith, 1993; Aprioz-Leehan and Feldmann, 1997; Martienssen, 1998; Krysan et al., 1999; Parinov and Sundaresan, 2000).

An advantage of using T-DNA as an insertional mutagen is that it directly generates stable insertions into genomic DNA and does not require additional steps to stabilize the insert. On the other hand, complex patterns of T-DNA integration, including transfer of vector sequences adjacent to T-DNA borders and the large frequency of concatemeric T-DNA insertions, can complicate further PCR analysis (De Buck et al., 2000; Kononov et al., 1997). Small and major chromosomal rearrangements induced by T-DNA integration have been frequently observed, leading to difficulties in the genetic analysis of the insertion, such as mutant phenotypes that are not correlated with the T-DNA insertion (Nacry et al., 1998; Laufs et al., 1999). Nevertheless, the development of highly efficient methods of T-DNA transformation for Arabidopsis has made it feasible for laboratories to generate thousands of transformants relatively rapidly (Krysan et al., 1999; Clough et al., 1998).

While T-DNA insertions are easily generated in Arabidopsis, it is difficult to generate a large collection of independent T-DNA lines in plant species for which transformation methods are more laborious. In contrast, transposon mutagenesis can be accomplished using a limited number of starter lines generated by transformation (Ramachandran and Sundaresan, 2001). Transposons integrate in the genome as single intact elements, as compared to the complex integration patterns that are frequently generated by T-DNA, and can be used for reversion analysis by remobilizing the transposon insertion. It is well established that transposons of the Ac-Ds family preferentially jump to linked sites (Bancroft and Dean, 1993b; Machida et al., 1997; Ramachandran and Sundaresan, 2001). This property of transposable elements can be exploited for targeted (i.e. localized) transposon mutagenesis of a chromosomal region, which has been used in endogenous systems (Sundaresan, 1996) as well as in heterologous systems (Smith et al., 1996; Mussett et al., 2003; Zhang et al., 2003). Localized transposon mutagenesis can also be used to generate double knockouts of tandemly duplicated genes through reinsertion of transposed elements (Tantikanjana et al., 2004), which would otherwise require a tedious screen for recombinants of two closely linked insertions within the duplicated genes.

1.3.2.2 The Design of Forward Genetic Screens

A genetic screen is a procedure to identify and select individuals which possess a phenotype of interest. A screen for new genes based on the observation of a particular mutant phenotype is often referred to as a forward genetic screen. On the other hand, identifying new mutant phenotypes caused by disruption of genes that have already known sequences is called a reverse genetic screen. Although the complete sequence of the Arabidopsis genome is known, its functional characterization depends on well-designed forward genetic screens, which remain a powerful strategy to identify genes that are involved in many aspects of the plant life cycle.

The pioneers of Arabidopsis genetics carried out screens in strains that were collected from the wild and isolated mutants that showed altered leaf shape, altered flowering time and flower morphology (McKelvie, 1962) or that were
characterized by embryo (Meinke and Sussex, 1979) or seedling lethality (Mayer et al., 1991). In addition, screening on defined media allowed the recovery of mutants that were defective in phytohormone synthesis and perception (Bleecker et al., 1988; Guzman and Ecker, 1990; Koornneef et al., 1982; Koornneef et al., 1984; Maher and Martindale, 1980) and other biosynthetic pathways (Li and Rédei, 1969). Screens that use brute force biochemical assays and screening under specific biotic and abiotic environmental conditions led to the analysis of the lipid biosynthesis pathway (Browse et al., 1985; James and Dooner, 1990; Lemieux et al., 1990) and to the genetic analysis of photomorphogenesis (Quail, 1998) and plant-pathogen interactions (Glazebrook et al., 1997), respectively. These screens, however, only allowed identifying obvious phenotypes. Therefore, genetic screens became more sophisticated through the design of particular genetic backgrounds in which more precise questions can be addressed.

The success of a forward genetic screen depends mainly on the definition of a suitable genetic background and an easy and tight procedure to identify the mutants of interest.

Whereas a WT background is suitable for recovering mutants with striking phenotypes, it is not suitable for detecting additional loci that are redundant to a previously identified gene. Neither does a WT background allow the identification of mutants that are involved in processes that do not immediately lead to a morphological or easily scorable biochemical phenotype. At present, the establishment of transgenic approaches that use drug or herbicide resistance genes for selection, and reporter genes to construct visible phenotypes, allows Arabidopsis researchers to engineer a scorable phenotype for virtually any process in the life cycle of a plant.

The simpler the screening procedure, the better suited it will be for screening large numbers of plants, thereby increasing the chances of identifying mutants of interest. Ideally, if a selection method is available, this is preferable to screening, as it is less laborious. If screening cannot be avoided, it should be easy and tight (the fraction of false positives should be low). If neither selection nor an easy screening procedure is possible, the assay should at least be very informative, as is the case for enhancer detection and gene trapping, which provide detailed information about the expression of a gene. Two-step screens offer a successful way to isolate mutants that are not easy to identify (Bartee et al., 2001; Moore et al., 1997). The first step consists of screening both for an easily scorable trait that can be expected to alter in the mutant candidates and for a secondary phenotype that is characteristic for the mutants targeted. As the secondary phenotype could also result from phenomena that are unrelated to the biological process of interest, the mutant candidates must be further characterized in a second step using the more laborious assay to select the informative mutants.

1.3.2.3 THE ISOLATION OF TAGGED MUTATED GENES

Using T-DNA tagging, several Arabidopsis genes have been identified and cloned. Once co-segregation of T-DNA along with the phenotype has been established for a given mutant, isolation of the disrupted gene can be achieved by a number of strategies.

Sequences flanking the insertion can be easily identified from single or low copy number lines using inverse PCR (IPCR) and thermal asymmetric interlaced (TAIL) PCR (Liu et al., 1995). Recently, Antal et al. (2004) have reported a simplified version of TAIL PCR, called single oligonucleotide nested (SON)-PCR, which involves only two rounds of PCR with two or three nested sequence primers.

Another strategy employed, called plasmid rescue (Yanofsky et al. 1990; Feldmann, 1992), makes use of insertion within T-DNA of an ‘ori’ sequence, a
replication of origin required for replication of plasmids in *E. coli*, along with a bacterial selection marker.

Finally, T-linker specific ligation PCR (T-linker PCR), described by Yuanxin et al. (2003), provides another simple and efficient way to amplify the T-DNA flanking region.

1.3.2.4 The Isolation of Untagged Mutated Genes using Map-Based Cloning Approaches

If a mutation is caused by the insertion of a T-DNA or transposon, the inserted sequence provides a tag pointing directly to the gene. However, when radiation or chemical agents are used a map-based cloning (or positional cloning) approach is required for the isolation of the mutated genes. Although this procedure is more laborious and time-consuming, there are several good reasons to keep screening for radiation or chemically induced mutants.

Besides from being easy to apply and inducing a high number of mutations per plant, the use of radiation or chemical mutagens has additional advantages. For instance, mutations with special properties such as weak, dominant or conditional alleles are more likely to be created using radiation or chemical agents as mutagens. This is in contrast to insertion mutagenesis, which tends to result in complete knockouts of a gene, making it difficult to associate a phenotype other than death with essential genes. It was reported that leaky mutations in *VTC1*/*CYT1* can result in ozone sensitivity and reduced vitamin C levels in Arabidopsis, while knockout mutations cause embryolethality (Conklin et al., 1999; Lukowitz et al., 2001). Moreover, key regulatory steps in biochemical pathways are often found through dominant point mutations that prevent feedback inhibition of an enzyme, such as anthranilate synthase (Kreps et al., 1996; Li and Last, 1996) or aspartic acid kinase (Heremans and Jacobs, 1997). Such dominant mutations would not be found by using insertional mutagenesis. Also, many chemically induced second-site mutations have been identified.

The big advantage of map-based cloning is that it is an approach without prior assumptions. It is a process of discovery that makes it possible to find mutations anywhere in the genome, including intergenic regions and genes that do not resemble any gene with known or inferred function. A prerequisite for map-based cloning is the availability of a genome map. Such a map shows the positions of genes and other sequence features, which serve as reference points and guide one to the region containing the gene of interest. Conventionally, the construction of genome maps, or genome mapping, has been divided into two categories: genetic mapping and physical mapping.

**Genetic Mapping**

Genetic mapping uses genetic techniques, such as cross-breeding and pedigree analysis, to construct maps showing the positions of genes and other sequence features on a genome. The principles of inheritance as first described by Mendel in 1865 were the first step towards genetic mapping.

**Mendel’s Laws of Inheritance**

From the results of his breeding experiments with peas, Mendel concluded that each pea plant possesses two alleles for each gene, but displays only one phenotype. This is easy to understand if the plant is homozygous for a particular trait, as then it possesses two identical alleles and displays the appropriate phenotype. However, Mendel showed that if two pure-breeding plants with different alleles are crossed, all the progeny (F1 generation) display the same phenotype.
These F1 plants were heterozygous for the phenotypes being studied, since each inherited one allele from each parent. The phenotype expressed in the F1 plants is dominant over the second, recessive phenotype.

Mendel carried out additional crosses that resulted in the postulation of two laws of genetics. The First Mendelian Law states that alleles segregate randomly (figure 1.12). This means that each member of an F1 progeny has the same chance of inheriting one parent’s allele as it has of inheriting the other parent’s allele. The Second Law described by Mendel is that pairs of alleles segregate independently, so that the inheritance of alleles of one gene is independent of the inheritance of alleles of another gene. Mendel’s theory about the existence of ‘unit factors’ (genes) and the way they are inherited was absolutely correct.

We know now that the two alleles of a gene are carried by the two homologous chromosomes in a diploid cell and that the way in which alleles are transmitted from parent to offspring corresponds to the events that occur when haploid gametes are produced during meiosis. The simple dominant-recessive rule proposed by Mendel, however, can be complicated by situations he did not encounter with the pea plants. One of these is incomplete dominance, where the heterozygous phenotype is intermediate between the two homozygous forms. Another complication is codominance, when both alleles are detectable in the heterozygote. Mendel made only one major error: his Second Law does not allow for linkage, the possibility that alleles of two genes will be inherited together because they are on the same chromosome. The discovery of linkage by Mendel’s followers led to the methods used for genetic mapping.

The Discovery of Linkage

As early as 1903, Walter S. Sutton, who along with Theodor Boveri united the fields of cytology and genetics, pointed out the likelihood that in organisms there are many more ‘unit factors’ (later called alleles) than chromosomes. Soon thereafter, genetic investigations with several organisms revealed that certain genes were not transmitted according to Mendel’s second law of independent assortment. When studied together in matings, these genes seemed to segregate as if they were somehow joined or linked together. Further research showed that such genes were part of the same chromosome and they were indeed transmitted as a single unit. We know now that most chromosomes consist of very large numbers of genes and, in fact, contain sufficient DNA to encode thousands of these units. Genes that are part of the same chromosome were said to be linked and to demonstrate linkage in genetic crosses. However, when the early geneticists carried out genetic crosses similar to those published 40 years earlier by Mendel, they discovered that very few pairs of genes displayed complete linkage. Pairs of genes were either inherited independently, as expected for genes in different chromosomes, or if they showed linkage then it was only partial linkage: sometimes they were inherited together and sometimes they were not (figure 1.13). The resolution of this contradiction between theory and observation was the critical step in the development of genetic mapping techniques.

The Behavior of Chromosomes during Meiosis Explains Why Genes Display Partial and Not Complete Linkage

The Drosophila geneticist, Thomas H. Morgan, first discovered the phenomenon of X-linkage. In his studies, he investigated numerous Drosophila mutations located on the X chromosome. When he analyzed crosses involving only one trait, he was able to deduce the
mode of X-linked inheritance. However, when he made crosses that simultaneously involved two X-linked genes, his results were at first puzzling.

In 1911, Morgan proposed that the gene separation he observed could be explained based on the cytological studies of F. Janssens. Janssens had observed that synapsed homologous chromosomes in meiosis wrapped around each other, creating chiasmata where points of overlap are evident. Morgan stated that these chiasmata could represent points of genetic exchange. After performing numerous crossing experiments, he concluded that linked genes exist in a linear order along the chromosome and that a variable amount of exchange occurs between any two genes. Furthermore, he explained that the closer two genes are on a chromosome, the less likely it is that a chiasma is formed between them and that a genetic exchange will occur. The term crossing over was first proposed by Morgan to describe the physical exchange leading to recombination (Griffiths et al., 1996).

Nowadays, crossing over is viewed as an actual physical breaking and rejoining process that occurs during the meiotic prophase. The double-strand-break-repair model (Szostak et al., 1983) is acknowledged to best explain meiotic reciprocal recombination. In this model two sister chromatids break at the same point and their 5' ends are degraded. In the next step one of the 3' tails invades
the intact homolog, displacing one of the strands. This 3' end serves as a primer for polymerase, which extends the tail until it can eventually be joined to a 5' end. Meanwhile, the displaced strand serves as a template to fill the gap left in the other broken strand. Two Holliday junctions form and may produce recombinant flanking DNA, depending on how they are resolved.

Crossing-over provides an enormous potential for genetic variation in the gametes formed by any individual. This type of variation, together with that resulting from independent assortment of the chromosomes, ensures that all offspring will contain a diverse mixture of maternal and paternal alleles.

From Partial Linkage to Genetic Mapping

Morgan’s student, Alfred H. Sturtevant, was the first to realize that his mentor’s proposal could also be used to map the sequence of linked genes. Sturtevant knew from Morgan’s work that the frequency of exchange could be...
taken as an estimate of the relative distance between two genes. He argued that the recombination frequencies between linked genes are additive and constructed a map of three X-linked Drosophila genes. He defined one genetic map unit as the distance between genes for which one product of meiosis out of hundred is recombinant. So, one map unit corresponds to a recombination frequency of 0.01 (or one percent) and is now referred to as one centimorgan (cM) in honor of Morgan’s work. In addition to these three genes, Sturtevant considered two other genes on the X-chromosome and produced a more extensive map including all five genes. He and a colleague, Calvin Bridges, soon began a search for autosomal linkage in Drosophila. By 1923, they had clearly shown that linkage and crossing over were not restricted to X-linked genes. Rather, they had discovered linked genes on autosomes, between which crossing over occurred (Griffiths et al., 1996).

Although Sturtevant’s basic principles are accepted as correct, it turns out that his assumption about the randomness of crossovers was not entirely justified. Comparisons between genetic maps and the actual positions of genes on DNA molecules, as revealed by physical mapping and DNA sequencing, have shown that some regions of chromosomes, called recombination hot spots, are more likely to be involved in crossovers than others. This means that a genetic map distance does not necessarily indicate the physical distance between two genes or DNA features. Also, we now realize that a single chromatin can participate in more than one crossover at the same time, but that there are limitations on how close together these crossovers can be, leading to more inaccuracies in the mapping procedure. Despite these qualifications, linkage analysis usually makes correct deductions about gene order, and distance estimates are sufficiently accurate to generate genetic maps that are of value as frameworks for genome sequencing projects.

Markers For Genetic Maps

The First Genetic Maps Were Constructed Using Genes as Markers

The first genetic maps, constructed in the early part of the 20th century for organisms such as Drosophila, used genes as markers. However, geneticists soon realized that only a limited number of genes were responsible for visual phenotypes whose inheritance can be studied, and in many cases the analysis is made less easy because more than one gene affects a single feature. To make genetic maps more comprehensive, it became essential to find characteristics that were more numerous, more distinctive and less complex than visual ones. Especially in bacteria and yeast, where few visual characteristics are present, using biochemistry to distinguish phenotypes was a solution to this problem. Also in humans biochemical phenotypes scorable by blood typing have been studied in the 1920’s (Yamamoto et al., 1990). Although genes are very useful markers, they are by no means ideal. Especially with larger genomes such as those of vertebrates and flowering plants, a map based entirely on genes is not very detailed.

Molecular Markers for Genetic Mapping

Restriction fragment length polymorphism (RFLP) markers have been used extensively for the construction of genetic maps (table 1.1). RFLPs are variations in the length of the DNA fragments produced after cleavage with a restriction enzyme ( Tanksley et al., 1989). Among the various molecular markers developed, they were the first to be used in human genome mapping (Botstein et al., 1980) and later they were adopted for plant genome mapping (Weber and Helentjaris, 1989). Rice is the most
extensively analyzed species by RFLPs (Causse et al., 1994; Kurata et al., 1994).

The first technique used to type RFLPs was Southern hybridization. The DNA fragments resulting from restriction are separated from one another by agarose gel electrophoresis and the ones being studied detected by hybridization probing. The polymorphism detected by RFLP is very reliable as it involves the recognition by specific restriction enzymes and hybridization with a specific probe. RFLPs are codominant since both alleles can be directly detected. However, such RFLP analysis is labour intensive and time consuming.

PCR is now more commonly used to detect RFLPs (Konieczny and Ausubel, 1993). Here, the region of DNA containing the RFLP is amplified before treatment with the restriction enzyme. The amplified fragments, called cleaved amplified polymorphic sequences (CAPS) can then be visualized on agarose gel using ethidium bromide. This is a less sensitive detection method than hybridization probing so that the background of unamplified fragments does not show up.

Another well known probe-based marker is the variable number tandem repeat (VNTR) or minisatellite, first discovered in humans by Alec Jeffreys in the 1980’s, and extensively used in forensic DNA profiling (Jeffreys et al., 1985). VNTRs consist of tandem arrays of short repeated sequences (a few tens of nucleotides in length) highly dispersed throughout the genome at numerous loci. They are embedded in unique flanking sequences and the loci are hypervariable in terms of their number of repeat units. Not only are there many different loci, but multi-allelic forms of single loci exist as well at the population level due to unequal crossing over.

VNTRs tend to be found more frequently near the ends of chromosomes, which is less convenient. Most of the VNTRs are longer than 300 base pairs since the repeat units are relatively long and there tend to be many of them in a single array. Therefore they are less amenable to PCR and are detected by hybridization probing.

Random amplified polymorphic DNA (RAPD) analysis involves amplification with PCR using a single or a combination of typically 10-mer oligonucleotides (Welsh and McClelland, 1990; Williams et al., 1990). Variation is based on the position and orientation of primer-annealing sites and the interval they span. The amplification products are separated on agarose gel and stained with ethidium bromide. RAPDs are dominant in the sense that the presence of a RAPD band does not allow distinction between hetero- and homozygous states. Their main disadvantages are poor reliability and reproducibility and their sensitivity to experimental conditions (Karp et al., 1996).

Simple sequence length polymorphisms (SSLPs) (also termed simple sequence repeats (SSRs), short tandem repeats (STRs), sequence tagged microsatellite sites (STMS) or microsatellites) are a type of insertion/deletion polymorphisms (InDels) and exploit the variability of short repetitive sequences for mapping purposes (Bell and Ecker, 1994). A primer pair is used to amplify a fragment containing a short repetitive element. The length of this repeat differs between different alleles. Consequently, PCR products will also differ in length. Typically, SSLPs may be dinucleotides \([\text{AC}]_n, (\text{AG})_n, (\text{AT})_n\), trinucleotides \([(\text{TCT})_n, (\text{TTG})_n]\), tetranucleotides \([(\text{TATG})_n]\) and so on, where \(n\) is the number of repeating units of the microsatellite. In addition to occurring at many different loci, they can also be polyallelic. SSLPs have been applied extensively towards genetic and physical mapping in plant and animal species. \((\text{AT})_n\) dinucleotides are the most abundant type of SSLP in plants. They are codominant and their detection on agarose gel is inexpensive and straightforward. However, specific primers need to be designed and tested, which is a time-consuming and costly business in species
of which the genome sequence is not known. The use of random primers, as in RAPD analysis, avoids this up-front input. Also non-repetitive InDels are used as molecular markers for genetic mapping. For instance for Arabidopsis, 18579 Indels have been identified (http://www.arabidopsis.org/Cereon/index.html), of which only 10% were associated with simple sequence repeats but most were found in non-repetitive sequences.

Sequenced characterized amplified regions (SCARs) are DNA fragments amplified by PCR using specific 15-30 base pair primers, designed from nucleotide sequences established in cloned RAPD fragments. By using longer PCR primers, SCARs do not face the problem of low reproducibility generally encountered with RAPDs. Obtaining a codominant marker may be an additional advantage of converting RAPDs into SCARs. However, SCARs may exhibit dominance when one or both primers partially overlap the site of sequence variation. SCARs are locus specific and have been widely applied in gene mapping studies (Nair et al., 1995; Nair et al., 1996; Paran and Michelmore, 1993).

The amplified fragment length polymorphism method (AFLP) combines the use of restriction enzymes with PCR amplification of fragments and detects either a single nucleotide polymorphism (SNP) in or immediately adjacent to one or both restriction sites, or an insertion/deletion (InDel) event in the sequence of one of the alleles (table 1.1; Vos et al., 1995). The AFLP technique is highly reproducible and requires no prior sequence information. It can thus be applied directly to any organism. Genomic restriction fragments are adaptor-ligated and PCR primers are designed based on adaptor and restriction site sequences. Subsets of fragments can subsequently be specifically amplified by adding random selective nucleotides at the 3’ end of the primers. The number of selective nucleotides needed depends on the size of the genome studied. Thus, although it is based on restriction and random selection, the targeted subsets of PCR templates are highly specific and the procedure is robust. A great advantage is that one basic set of primers can be used for organisms with comparable genome sizes.

Single nucleotide polymorphisms (SNPs) are individual point mutations in the genome (table 1.1). There are vast numbers of these, some of which give rise to RFLPs or CAPS markers, but many of which do not because the sequence in which they lie is not recognized by any restriction enzyme. Derived cleaved amplified polymorphic sequences (dCAPS) (Michaels and Amasino, 1998; Neff et al., 1998) can exploit almost all single nucleotide changes for mapping purposes. To achieve this, a mismatched PCR primer is designed next to the polymorphic position such that an artificial restriction site is created. Consequently, at this position the sequence of all PCR products is changed with respect to the genomic sequence. Restriction sites with interrupted palindromes are frequently used to design dCAPS markers because they allow to position the mismatched nucleotide at a distance from the 3’ end of the oligonucleotide where it is less likely to interfere with the priming of Taq polymerase. The size difference after digestion is rather small and essentially determined by the length of the mismatched primer. Disadvantages of using CAPS and dCAPS for genotyping include the extra time and cost involved in the restriction enzyme digestion and the possibility of a false result attributable to incomplete digestion by the restriction enzyme.

It is also possible to detect SNPs using allele-specific PCR primers, where the 3’ end of a primer has a perfect match with one allele and a mismatch with the other allele (Ugozzoli and Wallace, 1991).
**Table 1.1 The main categories of molecular markers.**

<table>
<thead>
<tr>
<th>Marker Name</th>
<th>Variation Type</th>
<th>Information Content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNP$^1$</td>
<td>INDEL$^2$</td>
</tr>
<tr>
<td>RFLP</td>
<td>+</td>
<td>(+)</td>
</tr>
<tr>
<td>PCR-RFLP</td>
<td>+</td>
<td>(+)</td>
</tr>
<tr>
<td>RAPD</td>
<td>+</td>
<td>(+)$^4$</td>
</tr>
<tr>
<td>AFLP</td>
<td>+</td>
<td>(+)$^4$</td>
</tr>
<tr>
<td>SSCP$^7$</td>
<td>+</td>
<td>(+)$^4$</td>
</tr>
<tr>
<td>Microsatellite</td>
<td>-</td>
<td>(+)$^7$</td>
</tr>
<tr>
<td>SNP</td>
<td>+</td>
<td>(+)$^8$</td>
</tr>
</tbody>
</table>

1 Single Nucleotide Polymorphism: any kind of base substitution. The fact that SNPs appear both as a variation type and a marker name is due to the fact that in reality, many techniques used for genotyping SNPs are grouped under this generic marker name.

2 Insertions and deletions.

3 Variable number of tandem repeats.

4 Although the RAPD, AFLP, RFLP, PCR-RFLP and SSCP techniques will detect base substitutions in the vast majority of cases, the two other types of DNA variation can also be analyzed.

5 In some instances, more than two alleles can be analyzed.

6 With an automatic sequencer, some markers can be scored as co-dominant.

7 Variations in PCR product length can be due to a deletion in the sequence flanking the microsatellite.

8 Many SNP detection techniques can also be used for scoring small insertions or deletions (INDELs).

9 Single strand conformation polymorphism (SSCP) is based on differences in folding conformation of single stranded DNA, when placed in non-denaturing conditions.

In theory, such primers can be used to preferentially amplify one allele of a SNP, but in practice a single-basepair change is often not enough to allow reliable differentiation between the two alleles of a SNP (Kwok et al., 1990; Cha et al., 1992). Drenkard et al. described a modification of the allele-specific amplification procedure (single nucleotide amplified polymorphism [SNAP]) (Drenkard et al., 2000). In this method, additional mismatches are introduced in the amplifying primers to maximize the difference in the amplification efficiencies of the two alleles of the SNP. Primer base pair changes that allow differential amplification of SNP sites can be predicted using the SNAPER program (http://patho.mgh.harvard.edu/ausubelweb). As with CAPS, SNAP markers are codominant and can be detected on agarose gels. However, it is necessary to run two PCR reactions (one for each allele of the SNP) to get complete SNAP genotyping data.

SNPs can also be typed by methods that do not involve gel electrophoresis. This is important because gel electrophoresis has proved to be difficult to automate so any detection method that uses it will be relatively slow and labor intensive. Oligonucleotide arrays (Gene Chips) contain thousands of oligonucleotides annealed to a glass slide. Such arrays allow the detection of SNP polymorphisms by differential hybridization in a highly parallel and automated manner (Lipshultz et al., 1999).

The Taq-Man PCR assay is designed to detect SNPs in a high-throughput manner through the release of fluorescent reporter dye from a quencher on the same oligonucleotide by 5' nuclease activity (Livak, 1999). By using more than one reporter dye, it is possible...
to detect different alleles of a SNP in a single reaction. The relatively high price of oligonucleotides tagged with reporter and quencher dyes makes this method cost-effective only if a large number of reactions need to be run with each SNP marker.

In pyrosequencing, an enzymatic cascade and luminometric detection system is used to measure the pyrophosphate that is released as a result of nucleotide incorporation (Ahmadian et al., 2000; Alderborn et al., 2000). Because 20 or more nucleotides are determined by this method, it is possible to detect several closely linked SNPs at once. The pyrosequencing method can be automated but has the disadvantage that it does not work well on stretches of repeated nucleotides.

DHPLC allows the detection of SNPs through different retention time of heteroduplex and homoduplex DNA in reversed-phase HPLC under partially denaturing conditions (Spiegelman et al., 2000). DHPLC allows detection of SNP polymorphisms in PCR-amplified DNA up to about 1,000 base pairs in size.

The method of fluorescence resonance energy transfer combines PCR and oligonucleotide ligation to detect SNPs (Chen et al., 1998). Dye-labeled oligonucleotide probes are used in this assay, and allele-specific ligation is detected by fluorescence resonance energy transfer, which only occurs when two dye-labeled oligos are joined by ligation.

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry can be used to rapidly detect SNPs in short DNA pieces by differences in molecular mass (Wada and Yamamoto, 1997).

A disadvantage of most high-throughput methods for detecting DNA polymorphisms is the high initial equipment cost, which results in a high per-assay cost for a lab that does not need to perform large numbers of genotyping reactions on a routine basis.

**Physical Mapping**

Although genetic mapping has been pursued in plants and animals for decades, it is only relatively recently that advances in cloning and clone fingerprinting have allowed the construction of physical maps (figure 1.14). A physical map is an ordered set of DNA fragments, among which the distances are expressed in physical distance units (base pairs). These days, a physical map usually comprises a set of ordered large insert clones such as bacterial artificial chromosomes (BACs), which have largely replaced yeast artificial chromosomes (YACs) as the preferred building blocks of a physical map. Physical maps can be independent of genetic information but are more valuable if linked to genetically mapped markers, and are even more powerful if integrated with genomic sequence data. A plethora of physical mapping techniques have been developed, including restriction mapping (Meyers et al., 2004; Strachan and Read, 1999), fluorescent in situ hybridization (FISH) (Dear, 1997; Heiskanen et al., 1996) and sequence tagged site (STS) mapping (Dear, 1997).

**The Arabidopsis Genome Sequence**

The sequence of the first plant genome was completed in December 2000, and it was the third complete genome of a higher eukaryote (The Arabidopsis Genome Initiative, 2000), after Drosophila melanogaster (Adams et al., 2000; Myers et al., 2000) and Caenorhabditis elegans (C_elegans_Sequencing_Consortium,1998). The Arabidopsis sequence represented 115 million base pairs (Mb) of euchromatin out of the estimated 125 Mb total. It was completed by the traditional bacterial artificial chromosome (BAC)-based approach from a minimal tiling path of overlapping large insert clones.
Figure 1.14 Comparison between the genetic and physical maps of *Saccharomyces cerevisiae* chromosome III (Oliver et al., 1992). The comparison shows the discrepancies between the genetic and physical maps, the latter determined by DNA sequencing. Note that the order of the upper two markers (glk1 and cha1) is incorrect on the genetic map, and that there are also differences in the relative positioning of other pairs of markers.

The sequence covered all ten chromosome arms, including parts of the centromeres (figure 3.15). A major finding revealed by the five chromosome sequences was evidence for genome-wide duplication followed by gene loss and major translocations. Tandem duplications are also extensive. Overall, only one of three genes does not have a close family member. The Arabidopsis gene repertoire of 11,000–15,000 gene families is comparable in number to other sequenced organisms, highlighting the similarity of life’s instructions that stem from our common single-celled ancestors. The total number of Arabidopsis genes was initially estimated at 25,490 and later revised to 30,700. What seemed remarkable when the human genome was published (Lander et al., 2001; Venter et al., 2001), was that Arabidopsis, the simplest plant genome, had a similar number of genes to humans. Certainly the number of genes does not reflect the complexity of the organism. Plants often contain many more genes than animals (mainly due to polyploidy or large-scale duplication). More likely, organism complexity is related to the levels of molecular interactions and regulatory circuitry using a similar genetic parts list.

The Arabidopsis genome contains an amazing array of genes encoding enzymes involved in primary and secondary metabolism, which is a unique feature of plants. These enzymes are the equipment of the plant chemical factories to build all required metabolic molecules and to generate an arsenal of specialized compounds. Because plants are sessile, they cannot move to avoid biotic attack or abiotic stress, or to find mating partners. Thus, they depend heavily on chemical signals. One example is the large gene family of cytochrome p450s, with more than 300 members involved in small molecule biosynthesis and detoxification (Paquette et al., 2000). Arabidopsis also has a large number of transcription factors (approximately 1500), many of which are in families unique to plants, such as the AP2/EREBP, RAV, NAC, ARF, and AUX/IAA families. Plants seem to lack many of the transcription factors families found in animals, such as nuclear steroid receptors. Arabidopsis has nearly 1000 serine/threonine kinases and more than 600 are receptor-like ser/thr kinases (Morris and Walker, 2003), but no receptor-like tyrosine kinases were found. Bacterial-like histidine kinases are also present. Some have been recruited as ethylene receptors or as red light photoreceptors, the phytochromes. Although showing sequence homology with histidine kinases, the plant phytochromes and several of the ethylene receptors have evolved ser/thr kinase activity (Xie et al., 2003). Plants have the extraordinary ability to photosynthesize, and hundreds of genes have been termed identified that are likely involved in light harvesting, chlorophyll biosynthesis, CO₂ fixation, or are a part of the two core energy - generating
photosystems. Light also regulates development, in a process photomorphogenesis (Neff et al., 2000). The genome contains, in addition to the red light-absorbing phytochromes, blue light-absorbing cryptochrome, and phototropin photoreceptors, as well as hundreds of putative downstream light-signaling proteins (Cashmore, 2003). We know the function of only a few dozen light-signaling proteins. Other unique gene functions of plants and yeast are creating electrochemical gradients using mainly proton-type ATPases, whereas C. elegans and Drosophila use mainly sodium-type ATPases. Thus, transport is usually coupled with protons rather than sodium ions. In addition, compared to animals, water channels (aquaporins) are highly over-represented in the Arabidopsis genome. There are also cellular differences between plants and animals at the genome level. No genes encode intermediate filaments in Arabidopsis, whereas actin and α- and β-tubulin are present. The plant cell wall surely is unique, and homologues of animal cytoskeleton anchorage proteins (which link with the extracellular matrix) have not been seen in the Arabidopsis genome. Plants also seem to have a different repertoire of small G-protein-signalling molecules. Different mechanisms are used to protect plants and animals against their biotic environment. No major histocompatibility complex genes or antibody-like genes were identified in plants; however, plant-specific nucleotide binding site leucine rich repeat (NBS-LRR)-type disease resistance (R) genes are abundant and can be grouped into subfamilies (Meyers et al., 2004). The 149 R genes are present at several loci throughout the Arabidopsis genome and occur as singletons or in highly polymorphic clusters. The sequence of the first plant genome revealed extensive horizontal transfer of genes from an ancestral cyanobacterium-like endosymbiont to the plant cell nucleus. New studies show that up to 18% of all plant genes originated from this engulfment, which resulted in the chloroplast (Marlin et al., 2002).
Mitochondrial genes have also been transferred to the nucleus. Strikingly, a large translocation of 620 kb brought several duplicated and rearranged copies of the mitochondrial genome near the centromere on chromosome 2 (Stupar et al., 2001). Compared to Drosophila, C. elegans, and yeast, there are about 150 protein families that are unique to Arabidopsis.

**MAP-BASED CLONING: GENERAL CONSIDERATIONS**

Map-based cloning is the process of identifying the genetic basis of a mutant phenotype by looking for linkage of the mutant phenotype to markers whose physical location in the genome is known. In the past, the paradigm used for map-based cloning has been one in which 'chromosome walking' played a central, and very time-consuming, role. First, either random molecular markers or established molecular linkage maps were used to conduct a search to localize the gene adjacent to one or more markers. A 'chromosome walk' was then initiated from the closest linked marker and a series of overlapping clones (cosmids or YACs) isolated. This continued either until one arrived at another molecular marker known to be situated on the opposite side of the target gene or until there was another indication that the walk had actually passed over the target gene. This was followed by cloning, complementation by transformation and de novo determination of the sequence of the entire region of interest to high quality without a previously determined WT DNA sequence as a guide (Arondel et al., 1992; Giraudat et al., 1992; Leung et al., 1994; Meyer et al., 1994; Mindrinos et al., 1994). In large genomes 'chromosome walking' was hampered not only by the large amounts of DNA being traversed, but also by the high frequency of repetitive DNA. Furthermore, libraries of large inserts (YAC libraries) used for 'chromosome walking' in large genomes have often contained chimeric clones or clones that have become rearranged after cloning.

The strategy of 'chromosome walking' was based on the assumption that it is difficult and time-consuming to find DNA markers that are physically close to a gene of interest. Technological developments invalidate this assumption for many species. As a result, the mapping paradigm has now changed such that one first isolates one or more DNA markers at a physical distance from the targeted gene that is less than the average insert size of the genomic library being used for clone isolation. The markers are then used to screen the library and isolate (or 'land on') the clone containing the gene.

Today, the advances in sequencing projects, the wealth of available marker systems and the progress made in methods to detect DNA polymorphisms make relatively rapid map-based cloning of a gene in model species feasible (Jander et al., 2002).

**MAP-BASED CLONING IN ARABIDOPSIS**

Map-based cloning of genes in Arabidopsis is greatly facilitated by the recent sequencing of the Col-0 and partial sequencing of the Ler genome. These two ecotypes were sequenced because they are among the most commonly used ecotypes in Arabidopsis research. Since the work of George Rédei in the 1950's, they have been the subjects of literally thousands of papers that have been published on the genetics, molecular biology, and biochemistry of Arabidopsis.

The Col-0 ecotype was the subject of a large international sequencing project, which has produced a nearly complete sequence using a clone-by-clone approach (The Arabidopsis Genome Initiative, 2000). Partial genomic sequence data generated from other ecotypes can be positioned on the framework of Col-0 genome sequence and sequencing of individual genes from
mutants or from other ecotypes has become routine. It is simply a matter of designing PCR primers based on the Col-0 sequence, amplifying the desired gene, and sequencing the product.

The sequence of the Ler genome was determined by low coverage shotgun sequencing at Cereon Genomics (Jander et al., 2002). This project generated approximately 700,000 500 base pair sequence traces. More than 200,000 of these traces were chloroplast, mitochondrial, or ribosomal DNA and were not used for the assembly. This left 498,037 traces (or 263 Mbp) of good quality raw sequence. Assembly of the sequences produced 50,262 contigs of an average size of 1.5 kb and 31,044 single-read sequences. The size of the assembled dataset was 92.1 Mbp, suggesting that approximately 70% of the genome is covered at the nucleotide level. It was found that at least some sequence from over 95% of all genes is present in the data assembled from the low coverage shotgun approach.

For Arabidopsis researchers who are interested in map-based cloning, the value of two genome sequences greatly exceeds that of only one such sequence. Whereas the availability of the genome sequence of a single ecotype mainly facilitates DNA sequencing in the final stages of a mapping project, data from two genomes make it possible to develop a database of DNA polymorphisms that can be used as genetic markers.

Besides their known genome sequences other available tools and resources make Col and Ler the combination of choice for map-based cloning in Arabidopsis (Lukowitz et al., 2000). For the Col accession, for instance, various libraries were made from WT plants and were deposited at the Arabidopsis stock centers, including genomic and cDNA phage-libraries, and libraries of yeast- and bacterial artificial chromosomes and of transformation-competent artificial chromosomes, which may be used for molecular complementation (Liu et al., 1999). Most existing mutations, including mutations causing visible phenotypes that can be employed as genetic markers, have been induced either in a Col or a Ler background. Col and Ler are also the parents of a widely used collection of recombinant inbred lines (Lister and Dean, 1993). Hundreds of markers have been analyzed in these lines, and the genetic map produced from this work has become the standard against which other Arabidopsis genetic maps are aligned.

Map-Based Cloning Strategies: Choosing the Appropriate Marker Systems

Map-based cloning uses the fact that, as distances between the gene of interest and the analyzed markers decrease so does the frequency of recombination. To position any locus in the Arabidopsis genome using a large mapping population (1500-2000 individuals) in a 10 kb interval one would need approximately 12,000 well positioned markers. This is feasible with presently available technologies (Peters et al., 2003). However, to perform efficient map-based gene cloning it is important to follow a good strategy by using the appropriate marker systems in the different successive steps of the procedure.

Regional Targeting with High-Volume Marker Technologies

The probability of identifying one or more markers within a specified physical distance of a target gene decreases the closer one gets to the target. Several strategies have been developed that allow one to screen a large number of random molecular markers in a relatively short time and to select just those few markers that reside in the vicinity of the target gene. These methods rely on two principles: (i) the development of high-volume marker technology, which allows
hundreds or even thousands of potentially polymorphic DNA segments to be generated and visualized rapidly from single preparations of DNA and (ii) the use of segregating populations to identify, among these thousands of DNA fragments, those few that are lying in a region adjacent to the gene of interest.

High-volume marker systems such as amplified fragment length polymorphisms (AFLPs) (Vos et al., 1995; Zabeau, 1993) and high throughput single nucleotide polymorphism (SNP) detection (Cho et al., 1999; Gresham et al., 2006) have been used with success in several plant species. These methods allow a global comparison of genomes for genetic differences.

The AFLP method relies on PCR used in conjunction with primers that amplify random sites throughout the genome. Amplification products are analyzed on acrylamide gels to allow visualization of variation among individuals. The fact that multiple AFLP markers can be obtained per PC was exploited by Peters et al. (2004) to develop an AFLP-based genome-wide mapping strategy.

Borevitz and colleagues developed a high-throughput genotyping platform for the identification of SNPs by hybridizing genomic DNA from Arabidopsis thaliana accessions to an RNA expression GeneChip (AtGenome1). Using this approach several high heritability development and circadian clock traits were mapped via direct hybridization of mutant DNA to ATH1 expression arrays (Hazen et al., 2005).

By using these high-volume marker technologies, thousands of loci scattered throughout the genome can be assayed in a matter of weeks or months. The next issue, however, is to determine which of the amplified loci lie near the target gene.

In the past few years two approaches have proven to be effective for identifying from a large number of markers the few that reside near the locus of interest. Both involve the use of genetic stocks that are almost identical, except in the regions flanking the targeted gene.

A first and generally applicable strategy has been referred to as the bulked segregant analysis (BSA) method and relies on the use of segregating populations (Giovanonni et al., 1991; Michelmore et al., 1991). BSA requires the generation of populations of bulked segregants (bulks). When P1 and P2 are crossed, the F2 generation derived from this cross will segregate for alleles from both parents at all loci throughout the genome. If the F2 population is divided into two pools of individuals with contrasting phenotypes on the basis of screening at a single target locus, these two pools (bulk 1 and bulk 2) will differ in their allelic content only at loci contained in the chromosomal region close to the target gene. Bulk 1 individuals, selected for the recessive phenotype, will contain only P2 alleles near the target, while bulk 2 plants, selected for the dominant phenotype, will contain alleles from both P1 and P2. Bulk 1 and 2 will both contain alleles from both P1 and P2 at loci unlinked to the gene of interest.

The use of high volume marker technologies is particularly helpful during the first steps of the map-based cloning procedure, when no information about the location of the gene of interest is available.

The second strategy uses nearly isogenic lines (NILs) but this method can only be applied if inbred lines are generated that differ at the targeted locus (figure 1.16; Alonso-Blanco and Koornneef, 2000). NILs are created when a donor line (P1) is crossed to a recipient line (P2). The resulting F1 hybrid is then crossed back to the P2 recipient to produce the backcross generation (BC1). From BC1, a single individual that contains the dominant allele of the target gene from P1 is selected. Selection for the target gene is performed using markers. The BC1 individual is subsequently back-crossed again to P2, and the cycle of back-cross selection is repeated for a number of generations. In the BC7 generation most of
the genome will be derived from \( P_2 \), except for a small heterozygous chromosomal segment containing the selected dominant allele, which is derived from \( P_1 \). Lines homozygous for the target gene can be selected from the BC\(_1\)F\(_2\) generation. The homozygous BC\(_1\)F\(_2\) line is said to be nearly isogenic with the recipient parent \( P_2 \). The genetic profiles of pairs of NILs can be compared using high-volume marker technologies. If the same polymorphisms are observed between the NIL and \( P_2 \) and between \( P_1 \) and \( P_2 \), they should be derived from a locus tightly linked to the gene of interest.

**Fine-Mapping with Individual, Simple and Robust PCR Markers**

Once a chromosomal segment has been defined in which the gene of interest is localized, it is practical to start using individual markers to further narrow down the interval. The Arabidopsis InDel polymorphisms (http://www.arabidopsis.org/Cereon/index.html) are a good source for such PCR markers but any other easy to use marker (for instance, SSLP) will do. The flanking PCR markers are tested on 1,000-2,000 individuals from a segregating population (e.g. F\(_2\) population). The markers can be analyzed simultaneously, which makes identifying recombinants in the region of interest straightforward. The selected recombinants can subsequently be used to further delimit the area containing the target gene to a region of approximately 5 to 10 kb. Markers necessary for this process can be developed from the available Cereon InDel and SNP collection (Jander et al., 2002).
Outline of a Typical Map-Based Cloning Experiment

Creating a Mapping Population

The first step in the mapping process is to create a mapping population by crossing the mutant to a WT plant of another ecotype (in practice Col and Ler is the most common combination). Usually, it is preferred to clean up the genetic background of the mutant by at least one back-cross cycle. F₁ seeds are sown and, as the plants are growing, it is possible to perform phenotype analysis. Presence or absence of the mutant phenotype in the F₁ generation will suggest whether the mutation of interest is likely to be dominant or recessive. F₂ seeds are collected from self-pollination of the F₁ plants and the phenotype of the F₂ plants is determined, unless the trait can only be scored in the progeny (F₃) seed. A 3:1 segregation of WT versus mutants indicates a monogenic trait. Mapping resolution can be referred to as the average distance between two recombination breakpoints in a given mapping population. In Arabidopsis, a genetic distance of 1% recombination (or 1 centiMorgan (cM)) corresponds, on average, to a physical distance of about 250 kb. However, the ratio between genetic and physical distance is not constant and varies with respect to chromosome position and different mapping populations. The main determinant of mapping resolution is the size of a mapping population. Resolutions in the range of 10 to 40 kb can be obtained in mapping populations of about 1000 plants. Typically, a DNA fragment of this length contains between two and ten genes.

Detecting Linkage

Once the F₂ generation of a mapping cross is available, the mutation is assigned a rough position on the genetic map by identifying linked genetic markers. Next, the two closest linked markers are defined (two markers that are about 5 cM apart and define a genetic interval containing the mutation).

As already described above high-volume marker technologies, such as AFLP, allow to detect linkage and non-linkage very quickly since a large number of markers dispersed throughout the genome can be visualized simultaneously using just a few PCs. Peters et al. (2004) exploited the fact that multiple AFLP markers can be obtained per PC to develop an AFLP-based genome-wide mapping strategy (figure 1.17). Their strategy establishes linkage to a <6 Mb region by analyzing 20-30 mutant individuals from a segregating population with eight PCs that provide a well-dispersed grid of 85 AFLP markers covering the genome. Further analysis of approximately 120 mutant individuals will typically exhaust the available Arabidopsis AFLP map (Peters et al., 2001) and identifies a 200-800 kb region. To reduce the number of PCR reactions (especially when individual PCR markers are used to detect linkage) DNA samples can be pooled for bulked segregant analysis (Michelmore et al., 1991).

Cho et al. (1999) described another genome-wide mapping strategy using SNPs to provide a first rough localization of the mutation on the Arabidopsis genome. To obtain a higher resolution (<3.5 cM) only a very limited number of F₂ individuals (28) have to be analyzed. However, it is array-based (Affymetrix, Santa Clara, California, USA) and thus relatively expensive.

Before performing a large-scale screen to select recombinants for further fine-structure mapping, it is essential to confirm that there are no comparable and cloned mutants located in the identified region. This can be checked using the sequence-based map of Arabidopsis genes with mutant phenotypes (Meinke et al., 2003). Potential candidates can be checked by
performing allelism test crosses. If it is certain that the gene to be cloned is still unidentified fine-mapping can be initiated.

Collecting Recombinants in the Vicinity of the Mutation and Fine-Structure Mapping

Searching a large mapping population for chromosomes with a recombination in the vicinity of the mutation can be done effectively if two markers are known that are closely linked and that flank the mutation on both sides (figure 1.18). In a large F2 population (1500-2000 individuals) the majority of plants will have the same genotype for both flanking markers (Lukowicz et al., 2000; Peters et al., 2004). These plants are either not recombinant or, in very rare cases, carry two recombination events. The majority is not informative for mapping the mutation and can be discarded without further analysis. Thus, the bulk of the mapping population is only analyzed with two

---

**Steps**

- Generating F2 mapping population
- Establish linkage and non-linkage e.g. using AFLP-based genome-wide mapping strategy (Peters et al., 2004) → <6Mb
- Identify a 200-800 kb region containing the gene of interest with AFLP markers
- Check known mutants in the identified region
- Identify and test individual flanking PCR markers for fine-mapping purposes
- Select recombinants in a large mapping population with the flanking markers
- Fine-mapping: use recombinants to narrow down the region with InDel polymorphisms and/or SNPs
- Interval <40 kb → track down mutation by -sequencing region containing target gene -genetic complementation with wild type gene

**Required**

- Mutant and wild type in different genetic backgrounds
- 20-30 F1 individuals (mutants), AFLP map
- ~120 F2 individuals (mutants), AFLP map
- TAIR, sequence-based map of Arabidopsis genes with mutant phenotypes
- e.g. InDels, SSLPs
- 1000-2000 F2 individuals (mutants and WTs)
- e.g. InDels, SNPs
- sequencing facilities
- Agrobacterium-mediated transformation

---

**Figure 1.17** A map-based cloning procedure [after Peters et al., 2004]. The material needed to map a mutation in Arabidopsis thaliana is given for each step in the procedure. See Peters et al. [2001] for the AFLP map referred to in steps 2 and 3. For the sequence-based map of Arabidopsis genes with mutant phenotypes referred to in step 4, see Meinke et al. (2003). See http://www.arabidopsis.org/ for the TAIR database; see http://www.arabidopsis.org/Cereon/index.html for further details of InDel polymorphisms and SNPs.
flanking markers and not considered further. A minority of the F2 plants will show different genotypes for the two flanking markers. Obviously, these plants carry a chromosome with a recombination close to the mutation and are informative for further mapping.

If the mutation is recessive, viable, and fertile, it is possible to utilize only plants that show the mutant phenotype for fine mapping. Since all these plants should be homozygous mutant for the gene of interest, no further analysis of the F3 generation is required. Alternatively, all F2 plants can be included in the mapping population. In this case the genotype with respect to the mutation must be determined for all plants that have been found to carry a recombination event in the vicinity of the gene. Obviously, this often requires analysis of the F3 generation.

Although a bulked segregant analysis is a very effective way to detect linkage, it usually does not allow determination of the order of closely linked loci on the chromosome. This can only be done by analyzing three-point mapping data. To find flanking markers that can be used to select recombinant chromosomes as described above, it is therefore necessary to examine individual F2 plants with markers from the region. Therefore, all F2 individuals first have to be subjected to a screen with the two flanking markers. Since the advantage of high-volume marker technologies is lost at this point, it is advised to use other PCR-based markers. For Arabidopsis, as already mentioned before, the Cereon collection of InDels and SNPs can be used for this purpose. InDel polymorphisms flanking the region of interest are identified and primers are
developed to score these flanking markers in a co-dominant way. The identified recombinants can then be used to further narrow down the region of interest.

Identification of the mutated gene

Once an interval of less than 40kb containing the mutation of interest has been determined, this entire region can be sequenced to find the mutation (Jander et al., 2002; Tanksley et al., 1995). Otherwise the number of F2 plants needed to find recombinants in such a small interval would be very large. Because the sequence of the Col-0 genome is known, one efficient way to sequence the mutant region is to design PCR primers to amplify overlapping segments of about 500 bp spanning the entire 40 kb interval. These segments are then sequenced and assembled, the sequence is compared with that of the WT plant (Col-0 or Ler), and the mutation is identified. In the case of a mutation in the Ler background, it is necessary to also sequence the Ler WT for comparison at every location where a difference to the WT Col-0 is found. In the case of a mutation in Col-0, a published sequence is available. However, it is necessary to confirm that any nucleotide that diverges from the published Col-0 sequence was induced by the mutagenesis treatment and is not present in the WT progenitor strain. This is because strain to strain differences exist in Col-0 WT, and even at the high quality standard of the Col-0 sequence, sequencing errors are expected and found (Jander et al., 2002).

The ultimate way to prove that a particular gene is mutated is to complement the mutant phenotype by transformation with the WT gene. This latter approach provides the most direct and definitive evidence. Genetic complementation by Agrobacterium-mediated transformation is routine in Arabidopsis and has been used in several instances to confirm that a candidate gene is responsible for a particular phenotypic trait.

1.3.3 Reverse Genetics: From Genotype to Phenotype

Rapid progress being made in genome sequencing projects provides raw material for the potential understanding of gene function, so effective reverse genetic strategies are increasingly in demand. Sequence information alone may be sufficient to consider a gene to be of interest, because sequence comparison tools that detect protein sequence similarity to previously studied genes often allow a related function to be inferred. Hypotheses concerning gene function that are generated in this way must be confirmed empirically. Experimental determination of gene function is desirable in other situations as well, for example, when a genetic interval has been associated with a phenotype of interest. In such cases, the functions of genes in an interval can be deduced from the phenotypes of induced mutations. Furthermore, the dissection of gene interactions often requires the availability of a range of allele types. However, most available methods for inferring function rely on techniques that produce a limited range of mutations, are labor-intensive or unreliable, or are limited to species in which special genetic tools have been developed. Just as the discovery of induced mutations led to forward genetics, the introduction of rapid reverse genetic methods can have great impact.

Several general strategies have been used to obtain reduction-of-function or knockout mutations in model organisms, including insertional mutagenesis (Alonso et al., 2003; Bouche and Bouchez, 2001; Long and Coupland, 1998; Parinov and
Sundaresan, 2000; Sussman et al., 2000) and RNA suppression (Chuang and Meyerowitz, 2000; Waterhouse et al., 1998), which have been widely used in plants. Insertional mutagenesis is now largely an in silico procedure for Arabidopsis researchers, as searchable databases of flanking sequences from T-DNA and transposon insertions are available on-line (Krysan et al., 2002). RNA suppression currently requires considerable manual effort, but it has the potential of reducing expression of repeated genes, which are especially common in plants, including Arabidopsis. However, because these techniques rely either on Agrobacterium T-DNA vectors for transmission or on endogenous tagging systems, their usefulness as general reverse genetics methods is limited to very few plant species. Recently, an additional, extremely powerful tool has become available: high-throughput screening for induced point mutations. This allows the identification of induced lesions in a particular gene of interest (targeted, induced lesions in genomes, TILLING) (Colbert et al., 2001). TILLING is a large-scale reverse genetic strategy that uses a high-throughput method based on an endonuclease that preferentially cleaves mismatches in heteroduplexes between WT and mutant DNAs. Subsequent analysis of cleavage products on a sequencing gel allows for the rapid identification of induced point mutations. Unlike insertional mutagenic approaches, which often provide knockout mutations, EMS-mutagenesis-based TILLING yields point mutations that provide an allelic mutant series. For instance, it might identify sublethal alleles, which are especially valuable for analyzing the phenotype of essential genes. TILLING has the potential to identify conditional (temperature-sensitive) alleles in a relatively simple way. Several TILLING reverse genetics projects are now underway in diverse plant species, including Arabidopsis (Till et al., 2003), Lotus (Perry et al., 2003), and Brassica oleracea (Gilchrist et al., unpublished results), and in some animal species (Hurlstone et al., 2003; Smits et al., 2004; Stemple, 2004; Wienholds et al., 2003). Reverse genetics approaches provide a powerful tool to study genes of unknown function and therefore complement the study of genes identified by forward genetic approaches.

1.4 Plant Functional Genomics Through Transcriptome Analysis

Reverse genetics represents one category of experimental tools that can be used to decipher the biological significance of the genome sequences from various organisms that are currently accumulating. However, information on both the physical and functional annotation of the genome can also be gained through transcript profiling (Hughes et al., 2001; Shoemaker et al., 2001). In recent years, transcript profiling has become synonymous with gene expression analysis, largely because of the technical difficulties and greater molecular complexity of proteomics and metabolomics (Smith, 2000). Such correlations are acceptable in many cases (Celis et al., 2000), even though the term ‘gene expression’ is often used to refer more directly to the compendium of gene products that ultimately cause cellular responses, and these are more often than not proteins. In some instances, protein levels are not reflected by alterations in mRNA level, and their activities are often controlled by post-translational modifications (Gygi et al., 1999). This means one has to take care with both experimental design and interpretation if one wishes to extrapolate transcript levels to those of protein (proteomics) or protein activity (metabolomics). Even so, it is generally accepted and there is much experimental evidence to support the statement that...
while post-transcriptional events play a role in modulating gene expression the primary level of control is at transcription. Therefore, the accessibility of transcript profiling in recent years has allowed the establishment of various high-throughput methodologies of gene expression analysis. These methodologies differ in their convenience, expense, number of transcripts assayed and sensitivity (Kuhn, 2001).

1.4.1 Overview of Global Transcript Profiling Methods

1.4.1.1 Sequence-Based Analysis of Gene Expression

In the early nineties, the automatization of the sequencing protocol allowed scientists for the first time to build up a large data collection of so-called expressed sequence tags (ESTs) in model organisms. ESTs are created by sequencing the 5’ and/or 3’ ends of randomly isolated gene transcripts that have been converted into cDNA (Adams et al., 1991). Generating sequences from cDNA fragments serves two purposes, the discovery of new genes and the assessment of their expression levels in the representative tissue (Ewing et al., 1999; Mekhedov et al., 2000; White et al., 2000). The basis of the approach is that the level of an mRNA species in a specific tissue is reflected by the frequency of occurrence of its corresponding EST in a cDNA library. In this respect the technology is distinct from ratio-based methodologies, such as microarraying, in being immediately quantitative (Bohnert et al., 2001). EST technologies are attractive because they do not rely on established sequence data from the organism under study, and they also fit well with labs already equipped to carry out high-throughput DNA sequencing (Adams et al., 1991). However, even at a few dollars per sequence the process can be expensive if one desires to progress beyond cursory screening of abundant mRNAs to in-depth analysis (Ohlrogge and Benning, 2000). In addition to the statistical problems in sampling small numbers from a large population (Audic and Claverie, 1997), there are also problems of bias in cloning and cDNA synthesis, though this last problem and others associated with data normalization are not specific to this technology. Auxiliary techniques are available that reduce the amount of sequencing. These include subtraction hybridization (Sargent, 1987) and related methods, representational difference analysis (RDA) (Hubank and Schatz, 1994) and suppression subtractive hybridization (SSH) (Diatchenko et al., 1996). Numerous variants on these technology themes are also available (Bonaldo et al., 1996; Sagerstrom et al., 1997).

Currently, there are nearly 20 million ESTs in the NCBI public collection, more than 4 million of which derive from plants (http://www.ncbi.nlm.nih.gov/dbEST/). With the rapid expansion of available EST data (e.g., http://www.arabidopsis.org; http://www.gramene.org; http://www.medicago.org; http://www.sgn.cornell.edu; http://www.tigr.org), opportunities for digital analysis of gene expression will continue to expand.

By developing Serial Analysis of Gene Expression (SAGE), Velculescu et al. (1995) addressed the problem of reducing the costs of traditional EST sequencing, and further modifications have allowed this procedure to handle small amounts of tissue (Datson et al., 1999; Peters et al., 1999). Though a similar sequence-based method to EST analysis, SAGE achieves a cost saving by the concatenation of multiple sequence tags of 10–14 base pairs, prior to cloning. By the size selection of inserts containing 25–50 tags, a comparable reduction of cost or increase in depth of analysis can be achieved over the sequencing of single ESTs. However, this increased efficiency comes at the
price of more extensive sequence reads. Consequently, this technology is best applied to organisms whose genomic sequences are known or that have a substantial cDNA sequence database. Because SAGE tags tend to originate from the 3’ portion of transcripts, they are less effectively screened against most cDNA libraries, which are generally sequenced only from their 5’ ends. Even with a reference database, because the tags are so short, there can be a redundancy of matches. SAGE is not very convenient for the comparison of many different samples and for the study of less abundant transcripts. SAGE has been used sparingly in plant research, in contrast to the numerous human and yeast studies (http://www.sagenet.org/).

As the name suggests, Massively Parallel Signature Sequencing (MPSS) has effectively tackled the problems of EST sequencing with regard to speed and depth of analysis, both through refined molecular biology and, equally important, good quality automation and data management (Brenner et al., 2000a, b). The method, developed at Lynx Therapeutics, Inc., California, is based on the in vitro cloning on microbeads of cDNA fragments from an mRNA population using the Megacclone technology (figure 1.19). To achieve this, a collection of oligonucleotide tags is synthesized from a defined set of 4-mers. This collection is so large that when attached to 3’ cDNA restriction fragments, tag-cDNA conjugates are obtained with virtually every polynucleotide having a unique tag. These conjugates are then amplified and hybridized to anti-tag sequences specific to separate microbeads. As a result, each microbead hybridizes about 100,000 copies of a single species from the interrogating cDNA. The cDNA fragments can then be sequenced in a flow cell by a method involving type II restriction endonucleases and the tandem ligation of differential adapters, one set of which are fluorescently labeled. A camera with image processing software tracks the fluorescence of individual beads through the multiple hybridization, ligation, and cleavage steps.
1.4.1.2 Fragment-Based Analysis of Gene Expression

Differential display (Liang and Pardee, 1992; Welsh et al., 1992) uses low-stringency PCR, a combinatorial primer set, and gel electrophoresis to amplify and visualize larger populations of cDNAs representing mRNA populations of interest. Since 1992, there has been a veritable cornucopia of published variations on this theme of differential display (Matz and Lukyanov, 1998). At the time, these methods provided a far more cost effective approach to identifying differentially expressed genes compared to existing subtractive methods. Consequently, they have been used extensively in plant research (Visioli et al., 1997; Sablowski and Meyerowitz, 1998). However, a survey of such studies shows that while differential display has been an effective means of gene discovery it has had less value as a means of gene expression profiling (Sambrook and Russell, 2001). This stems from the inherent nature of the PCR processes used in this and associated methods (McClelland et al., 1995). Because arbitrary primers are used at low annealing temperatures to allow for priming at multiple sites, the amplified products are not solely dependent on the initial concentration of a particular cDNA, but are also a function of the quality of match of the primers to the template. Predictably, under such conditions profiles are highly influenced by the PCR conditions as well as the sample quality (Matz and Lukyanov, 1998). This problem manifests itself in a high percentage of false-positive bands (Sun et al., 1994; Sompayrac et al., 1995), lack of sensitivity (Bertioli et al., 1995), and difficulties with reproducibility (Haag and Raman, 1994; Zhang et al., 1998). The use of longer primers has reduced but not fully alleviated problems of false-positive bands (Zhao et al., 1995; Martin and Pardee, 1999). Similar modifications have been employed to address sensitivity (Ikononmov and Jacob, 1996) and reproducibility problems (Linskens et al., 1995). However, these changes represent modifications that do not change the inherent problem of arbitrarily primed PCR reactions. In conclusion, differential display methods are relatively cheap and simple means of screening for differentially expressed genes, and are particularly good where the availability of RNA is limited (Renner et al., 1998; Bosch et al., 2000). However, they are not very accurate in quantitatively profiling global levels of gene expression, as illustrated by the number of false-positives generated.

More recently the principles of AFLP (Vos et al., 1995) have been applied to cDNA templates (cDNA-AFLP) (Bachem et al., 1996; Bachem et al., 1998; Goossens et al., 2003; Vandenabeele et al., 2003) and this approach has been used to identify differentially expressed genes involved in a variety of plant processes (Bachem et al., 2001; Bryne et al., 2002; Dellaqi et al., 2000; Durrant et al., 2000; Qin et al., 2000). This technique offers several advantages over more traditional approaches. Of particular importance is the fact that poorly characterized genomes can be investigated in a high throughput manner. Because the stringency of cDNA-AFLP PCR reactions is quite high (which is not the case with differential display) the fidelity of the cDNA-AFLP system allows much greater confidence in acquired data and differences in the intensities of amplified products can be informative (Bachem et al., 1996). In addition, this technique allows a wide variety of tissue types, developmental stages, or time points to be compared concurrently. As with the other profiling methods described here, the sensitivity of cDNA-AFLP is only limited by the ability of cDNA libraries to capture low-abundance transcripts. Sequencing of cDNA libraries is a more direct and comprehensive approach to gene expression profiling (Adams et al., 1991; Okubo et al., 1992), but this method requires substantial resources for cloning and sequencing, and is less sensitive to
low-abundance transcripts as mentioned above.

Two other fragment-based methods for whole-genome expression analysis were introduced by which cDNAs can be displayed combined with a high sample throughput. A first technology is called GeneCalling, which assays transcript abundance by processing poly-A+ RNA-derived cDNA through a restriction digestion utilizing optimized pairs of 6 base pair recognition restriction enzymes. Resultant gene fragments are ligated and end-labeled with 5'-fluorescamine and 3' biotin adapters and subjected to PCR amplification. Subsequently, the fragments are resolved by gel electrophoresis (Bruce et al., 2000; Shimkets et al., 1999). Differentially expressed fragments are identified and confirmed against a virtual digest of the appropriate databases. By combining the sequence knowledge derived from the two restriction enzymes with the length of a given fragment, one is able to 'call' putative gene assignments for each fragment. This information is sufficient for database cross-referencing to determine the identity or novelty (in the case no 'Gene Call' can be made in the database) of a particular gene. This technique proves to be very sensitive, but on the other hand is very laborious and time-consuming.

Another method is named TOtal Gene expression Analysis or TOGA (Sutcliffe et al., 2000). This approach utilizes sequences near the 3' ends of mRNA molecules to give each mRNA species present in the organism a single identity. This is achieved by exact matching by an equimolar mixture of 48 anchor primers to initiate reverse transcription. One member of this mixture initiates synthesis at a fixed position of each mRNA species, hereby defining a 3' endpoint. By restriction digestion with two enzymes, uniform fragment lengths for each cDNA species are obtained. After PCR amplification, the fragments are visualized and the status of each mRNA can be compared among samples in an automated fashion. Its identity can be electronically matched with sequences of known mRNAs in databases. This method is sensitive and its single-product-per-mRNA feature allows a near complete coverage of all mRNAs present in a given organism. However, as like ‘GeneCalling’, this method requires a high input of labor and time.

### 1.4.1.3 Hybridization-Based Analysis of Gene Expression

The principles underlying the hybridization of complementary nucleotide sequences are embodied in the structure of duplexed nucleic acids, and have been exploited experimentally for decades (Gillespie and Spiegelman, 1965). In this time, nucleic acid hybridization has been used in a variety of guises in the quantification of plant RNA levels. In the 1970s, it was used to study sequence complexity (Goldberg et al., 1978). Also in that decade, Southern developed a method using a solid support in hybridization studies of DNA fragments separated by gel electrophoresis (Southern, 1975). This led to a major advance in the analysis of gene expression with the development of northern-transfer hybridization (Alwine et al., 1977). This technique has been immensely useful over the years, though ironically the approach is less global and the method less genomic in nature than preceding solution-based systems. With the availability of nucleotide sequences and clones as physical reagents, hybridization-based approaches now allow for the simultaneous analysis of tens of thousands of genes. Quite literally, one can globally survey the transcription of a plant by hybridization. The interest in this form of transcript profiling has been stimulated by the development of two parallel microarray-based technologies, one based on spotting cDNA fragments (cDNA microarrays) (Schena et al., 1995), the other, the arrayed synthesis of
oligonucleotides (GeneChips) (Lockhart et al., 1996). These two methods have been extensively reviewed in recent years (Duggan et al., 1999; Lipshutz et al., 1999) and both have notable and distinct advantages.

cDNA arrays can be prepared directly from existing cDNA libraries, a large number of which are in the public domain. Thus, fabrication of cDNA arrays is only dependent upon availability of ordered clone collections and appropriate arraying and scanning instrumentation (Clark et al., 1999; Drmanac and Drmanac, 1999; Eisen and Brown, 1999). Once a set of corresponding PCR products has been generated, arrays can be created in multiple versions containing the entire set of available sequences or subsets of sequences resulting in more focussed arrays suitable for specific research applications (regulatory-, pathway-, stage- or response-specific arrays) (Jiao et al., 2003). Such arrays are also useful for reducing a statistical problem of scale (large numbers of features and low number of replications common in microarray experiments). Here, a large array might be used to identify differentially expressed genes of interest, which could then be re-arrayed as a smaller array and used in subsequent experiments. One benefit of this approach is that it can free up resources that can be used to increase experimental replication and thereby increase precision. Another advantage of cDNA arrays is that they can be used in “two-color” co-hybridization experiments that allow direct comparisons of transcript abundance in two mRNA populations of interest. Although this strategy generates comparative expression ratios instead of measuring absolute expression levels, it is effective for comparative expression profiling and reduces experimental variation that arises in microarray data collected from different chips (Aharoni and Vorst, 2001).

Like cDNA arrays, oligonucleotides can be printed using robotic instrumentation and (once appropriate oligonucleotides have been synthesized) sub-arrays for specific research applications can be fabricated easily. The main limitations in development of oligonucleotide arrays are the costs associated with sequence selection and oligonucleotide synthesis. As these costs continue to decline oligo-based arrays are likely to become more predominant in the near future because they offer a number of important advantages over cDNA arrays. One such advantage is the fact that oligo-based arrays can be fabricated using microfluidic technology, which utilizes light to direct the synthesis of short oligonucleotides onto a suitable matrix (photolithography) (Fodor et al., 1991; Fodor et al., 1993; Pease et al., 1994). Photolithography is particularly useful because it allows for the fabrication of extremely high-density arrays (>300 000 elements/ 1.28 cm²) (Lipshutz et al., 1999). Another important advantage is that the probes in an oligonucleotide array are designed to represent unique gene sequences such that cross-hybridization between related gene sequences (genes belonging to a gene family or genes with common functional domains) is minimized to a degree dependent upon the completeness of available sequence information. Cross-hybridization between homologous sequences continues to be problematic when using cDNA arrays. Furthermore, the array elements in an oligonucleotide array are typically designed to have uniform length, uniform melting temperatures, and to be of uniform concentrations, which can significantly reduce experimental variation and thereby increase statistical power and precision. The primary disadvantage of oligo-based arrays is that oligonucleotide sets can be very expensive because of the extensive sequence data and computational input required for
designing gene-specific oligonucleotide probes. A common limitation of all array approaches is the requirement of significant amounts of RNA for the preparation of fluorescently labelled targets.

Several microarray platforms have been produced based on the Arabidopsis sequence, including first generation arrays produced by a public consortium (The Arabidopsis Functional Genomics Consortium or AFGC; Wisman and Ohlrogge, 2000) and a commercial array that included more than 8,000 Arabidopsis genes produced by Affymetrix Inc. (Santa Clara, CA) (Zhu and Wang, 2000). The
Affymetrix GeneChip arrays are comprised of sets of 25-base oligonucleotides synthesized in situ via a photolithographic process (Lockhart et al., 1996). The most recent generation of the Arabidopsis Affymetrix GeneChip can monitor up to 24,000 gene sequences. Affymetrix also offers arrays for a variety of other model organisms including rat, mouse, Escherichia coli, Pseudomonas aeruginosa, yeast, Caenorhabditis elegans and Drosophila melanogaster. A second commercial array, produced by Agilent Technologies Inc. (Palo Alto, CA), represents 28,500 Arabidopsis genes and is based on the process of ink-jet printing of 60-base probes (Hughes et al., 2001).

Recently, Allemeersch et al. (2005) have assessed the performance of the CATMA microarray designed for Arabidopsis thaliana transcriptome analysis and compared it with the Agilent and Affymetrix commercial platforms. The CATMA array consists of gene-specific sequence tags of 150 to 500 bp, the Agilent array of 60-mer oligonucleotides, and the Affymetrix GeneChip of 25-mer oligonucleotide sets. They have matched each probe repertoire with the Arabidopsis genome annotation (The Institute for Genomic Research release 5.0) and determined the correspondence between them. Array performance was analyzed by hybridization with labeled targets derived from eight RNA samples made of shoot total RNA spiked with a calibrated series of 14 control transcripts. CATMA arrays showed the largest dynamic range extending over three to four logs. Agilent and Affymetrix arrays displayed a narrower range, presumably because signal saturation occurred for transcripts at concentrations beyond 1,000 copies per cell. Sensitivity was comparable for all three platforms. For Affymetrix GeneChip data, the RMA software package outperformed Microarray Suite 5.0 for all investigated criteria, confirming that the information provided by the mismatch oligonucleotides has no added value. In addition, taking advantage of replicates, they conducted a robust statistical analysis of the platform propensity to yield false positive and false negative differentially expressed genes, and all gave satisfactory results. The results establish the CATMA array as a mature alternative to the Affymetrix and Agilent platforms (Crowe et al., 2003; Hilson et al., 2004).

### 1.4.2 Molecular Phenotyping Through Comprehensive Transcriptome Analysis

Whole-genome expression profiling is facilitated by the development of DNA microarrays and represents a major advance in genome-wide functional analysis. In a single assay, the transcriptional response of each gene to a change in cellular state can be measured. This change can be a process such as cell division or a response to a chemical or a genetic perturbation (such as a mutation). Because the relative abundance of transcripts is often linked to specific cellular needs, most expression profiling studies to date have focused on the identification of genes that are involved in specific processes of interest, such as systemic acquired resistance, key pathways involved in the circadian clock in Arabidopsis thaliana, the analysis of developmental and sex-regulated gene expression in Caenorhabditis elegans or the global transcriptional response of Saccharomyces cerevisiae to ionic radiation (De Sanctis et al., 2001; Harmer et al., 2000; Jiang et al., 2001; Schenk et al., 2002).

However, the idea that the collection of genome-wide expression profiles in a cell, which consists of thousands of individual observations, can serve as a transcriptional fingerprint and constitutes in this manner a complete and detailed molecular phenotype has only recently begun to receive attention. Provided that the cellular transcriptional response to disruption of different steps in the same pathway is similar, and that there are sufficiently unique transcriptional
responses to the perturbation of most cellular pathways, the systematic characterization of novel mutants could be carried out with a single genome-wide expression measurement.

Using a comprehensive database of reference profiles, the pathway(s) perturbed by an uncharacterized mutation would be identified by simply asking which expression patterns in the database its molecular phenotype most strongly resembles, in a manner analogous to fingerprinting. An important aspect of this approach is that it should be equally effective at determining the consequences of pharmaceutical treatments and disease states. A sufficiently large and diverse set of profiles obtained from different mutants, treatments, and conditions would also result in a relatively comprehensive identification of coregulated transcript groups, allowing additional hypotheses to be drawn regarding the functions of genes based on the regulatory characteristics of their own transcripts (Eisen et al., 1998).

This strategy was elegantly illustrated by Hughes et al. (2000) who analyzed global gene expression in yeast (Saccharomyces cerevisiae) by monitoring mRNA expression changes of approximately 6,300 genes under 300 experimental conditions. These data were used to build a reference database or a compendium of expression profiles. The 300 conditions were highly diverse, involving deletion of 279 individual genes from all major functional classes as well as treatment with 13 drugs. Among the deletion mutants, 69 functionally uncharacterized open reading frames (ORFs) were analyzed. These widely divergent experiments led to an alteration in expression of 4,553 genes, or approximately three-fourths of the entire genome in at least one experiment. Moreover, in 96% of the experimental conditions, the expression of at least one gene was significantly altered by more than two-fold and in 41% of the experiments the expression levels of more than 20 genes were significantly changed. Hughes et al. used their compendium database to identify matching patterns of expression among genes of known function and uncharacterized ORFs. By identifying similar 'fingerprint' patterns, the authors were able to hypothesize, and subsequently verify, the diverse functions of eight orphan ORFs that encode proteins involved in sterol metabolism, cell wall function, mitochondrial respiration or protein synthesis.

To extend this paradigm, they demonstrated that it is possible to identify the target or pathway affected by a therapeutic drug whose mechanism of action is unknown. The expression profile observed in yeast treated with the topical anaesthetic dyclonine closely matched that observed in yeast with a perturbed ergosterol pathway, particularly erg2Δ yeast, deficient in Erg2p, a sterol C-8 isomerase. Validation experiments implicated Erg2p as the target of dyclonine. Interestingly, a sequence homology search revealed the likely human homologue to be a α-receptor, which is not a sterol isomerase, but rather a neurosteroid-interacting protein that regulates potassium channels and presumably accounts for the anaesthetic actions of dyclonine in humans.

The power of the microarray technology, which allows the global analysis of patterns of simultaneous gene expression and moreover the ability to link the database of expression patterns obtained using one microarray to other databases, was also exploited extensively for the classification of human tumours (Ahr et al., 2001; Huang et al., 2003; Sorlie et al., 2001; Sorlie et al., 2003; Van ’t Veer et al., 2002). The potential of the approach was first demonstrated by Golub et al. (1999) who determined the RNA expression profiles of a series of acute myeloid leukaemias and acute lymphoblastic leukaemias. A class discovery program automatically identified the distinction between these
leukaemia groups without prior knowledge of the diagnosis. Using this data set, the program was then able to classify new leukaemia cases.

Perou et al. (2001) obtained microarray portraits of a set of 65 surgical specimens of human breast tumours from 42 individuals. The authors selected a subset of 496 genes that showed considerable variation in expression levels between different tumours. When variations in level of expression within the genes were used to order the tissue samples, a striking division of the tumours into oestrogen receptor (ER)-positive and ER-negative categories occurred. Interestingly, the ER-positive group was characterised by the expression of many genes specific to breast luminal cells. The ER-negative tumours could be further divided into several subgroups. A subgroup of six tumours had an expression pattern characteristic of breast basal (myoepithelial) cells, whereas another subgroup was partially characterised by high expression levels of genes associated with overexpression of the ERBB2 oncogene. On the basis of these results the authors suggested that we are far from a complete understanding of the diversity of human breast tumours.

In yet another study, Pérez-Amador et al. (2001) used DNA microarray analysis for the further molecular characterization of the dst1 mutant of Arabidopsis. The dst (downstream) mutants were isolated originally as specifically increasing the steady state level and the half-life of DST-containing transcripts. As such, they offer a unique opportunity to study rapid sequence-specific mRNA decay pathways in eukaryotes. These mutants show a three-fold to four-fold increase in mRNA abundance for two transgenes and an endogenous gene, all containing DST elements, when examined by RNA gel blot analysis. However, they show no visible aberrant phenotype. The authors used DNA microarrays to identify genes with altered expression levels in dst1 compared with the parental plants. In addition to verifying the increase in the transgene mRNA levels, which were used to isolate these mutants, they were able to identify new genes with altered mRNA abundance in dst1. RNA gel blot analysis confirmed the microarray data for all genes tested and also was used to catalog the first molecular differences in gene expression between the dst1 and dst2 mutants. These differences revealed previously unknown molecular phenotypes for the dst mutants. Cluster analysis of genes altered in dst1 revealed new coexpression patterns that prompt new hypotheses regarding the nature of the dst1 mutation and a possible role of the DST-mediated mRNA decay pathway in plants.

For Arabidopsis a special initiative, called the CAGE project, has been set up in 2002 to build a gene expression reference database. A Consortium of European Arabidopsis functional genomics centers has teamed up with bioinformatics partners that contribute expertise in microarray data processing, analysis and storage/distribution. A total of 2000 Arabidopsis samples are being produced and analyzed under largely standardized conditions. These samples are profiled on CATMA microarrays containing gene specific probes for most Arabidopsis genes, to build a compendium of expression profiles. The data will be assessed for statistical significance and submitted to the ArrayExpress database at the European Bioinformatics Institute (EBI). EBI will deliver specific CAGE ontology, and data submission pipelines. The compendium data will be annotated and analyzed for content and confirmation of gene function.

Another example of a compendium of gene expression data is the Genevestigator Arabidopsis microarray database and analysis toolbox (Zimmermann et al., 2004), which is a database and Web-browser data mining interface for Affymetrix GeneChip data. Users can query the database to retrieve the expression patterns of individual genes.
throughout chosen environmental conditions, growth stages, or organs. Reversely, mining tools allow users to identify genes specifically expressed during selected stresses, growth stages, or in particular organs. Using GENEVESTIGATOR, the gene expression profiles of more than 22,000 Arabidopsis genes can be obtained, including those of 10,600 currently uncharacterized genes.

The fundamental advantage of a compendium approach over conventional assays is that it substitutes a single genome-wide expression profile in the place of many conventional (often tedious) assays that measure only a single cellular parameter. Because the compendium approach to determining gene function does not rely at all on the regulatory characteristics of the gene of interest, it has a significant advantage over the widely accepted idea of assessing gene function based on co-regulation. Furthermore, the same compendium used to characterize mutants can also be used to characterize other perturbations, including treatments with pharmaceutical compounds, and potentially disease states as well.
We designed a screening procedure to isolate temperature-sensitive mutations that affect plant growth and development. By applying our screening on an EMS mutagenized population of Arabidopsis, we obtained a collection of twelve growth mutants, each of which potentially corresponds to an unknown (essential) plant gene.

A general phenotypic analysis allowed us to classify the twelve mutants into two distinct categories. A first category consists of eight mutants (dgr1 to dgr8; defective in growth), which are inhibited in overall plant growth. A second category comprises four mutants (dlr1 to dlr4; defective in lateral root formation) that show defects specifically in the process of lateral root formation.

Since we aimed to eventually identify novel essential genes involved in plant growth and development, dgr1 and dgr2 appeared, based on their specific phenotypes, to be the most representative candidates to focus our further research on. Whereas both mutants are inhibited in overall plant growth and development, dgr1 was more specifically selected for its irreversible phenotype and dgr2 as a representative for a subclass of mutants with a reversible phenotype.

In first instance, our study demonstrates that temperature-sensitive mutants provide a powerful tool for studying gene function in vivo. Secondly, we were able to isolate twelve interesting mutants, for which further phenotypic and molecular genetic characterization is expected to provide a diversity of information, potentially useful for further understanding of the biological mechanisms underlying plant growth and development.
2.1 INTRODUCTION

Since the publication of the Arabidopsis genome sequence (The Arabidopsis Genome Initiative, 2000), which was the first to be completed among flowering plants, the fundamental question in the field of plant growth and development has been how many and which of these thousands of genes are essential to produce a viable and fertile plant with a normal phenotype. To determine the biological significance of a gene product and its relevance to plant growth and development both forward genetics and saturation mutagenesis on a genomics level are required. When gene functions are not redundant, this objective can be met by investigating which gene disruptions result in a mutant phenotype. Using T-DNA insertion mutagenesis, Tzafrir and co-workers (2004) recently constructed a first large-scale dataset of essential Arabidopsis genes, required for normal embryo development. Although the identification of essential genes is straightforward using regular knock-out mutants, the study of gene function in vivo is often problematic due to the resulting lethal phenotypes.

Screening for conditional lethal mutants provides a means to genetically study mutations in essential genes. The potential to control the activity of individual genes reversibly and quantitatively has opened up new perspectives for the in-depth analysis of gene function. Temperature-sensitive (ts) mutations are the most widely used type of conditional mutations in which there is a marked drop in activity of the gene product when the gene is expressed above (or below) a certain (restrictive) temperature. Below (or above) this temperature, at so-called permissive temperatures, the phenotype of the mutant is similar to that of the WT (WT). This way, gene function can be studied at the restrictive temperature, while mutant lines can be maintained and propagated at the permissive temperature.

Although there are other excellent inducible systems for gene expression (Gossen and Bujard, 1992; Gossen et al., 1995), ts mutants have several unique advantages over such systems, including fast temporal response, high reversibility and the applicability to any tissue type or developmental stage of an organism upon which temperature changes can be imposed. Such organisms include viruses (Shenk et al., 1975), prokaryotes (Fangman and Novick, 1968; Carl, 1970; Wechsler and Gross, 1971; Kirschbaum et al., 1975), yeasts (Hartwell, 1967; Hartwell et al., 1974), nematodes (Hirsh and Vanderslice, 1976; Miwa et al., 1980; Cassada et al., 1981; Kemphues et al., 1988a, Kemphues et al., 1988b), insects (Suzuki, 1970; Suzuki, Kaufman and Falk, 1976; Hall, 1978a; Hall and Greenspan, 1979), amphibia, fish (Nechiporuk et al., 2003; Parichy et al., 2003), and plants.

In the past, several ts mutants have been isolated in plants and most of them show normal phenotypes at low temperature and mutant phenotypes at high temperature. The deficiens mutant of Antirrhinum majus, for instance, was used in the functional analysis of homeotic genes (Schwarz-Sommer et al., 1992; Zachgo et al., 1995). Other examples include carrot variants unable to complete embryo development (Breton and Sung 1982; Schnall et al. 1988), auxin-auxotrophic mutants of Nicotiana plumbaginifolia (Fracheboud and King, 1991) and the argentia mutant of Zea mays, being delayed in normal development of photosynthetic cell types (Langdale et al., 1987).

In Arabidopsis, the ts mutant rsw1, which harbours a point mutation in the catalytic subunit of cellulose synthase, has contributed greatly to the molecular analysis of cellulose biosynthesis (Arioli et al., 1998; Peng et al., 2000, Lane et al., 2001). Other examples of Arabidopsis ts mutants include the add mutants (Pickett et al., 1996), defective in early shoot development, and the srd and rrd mutants, disturbed in the redifferentiation of shoots from root explants or roots from
hypocotyl explants, respectively (Ozawa et al., 1998; Sugiyama, 1999; Yasutani et al. 1994). Only in rare cases, has high temperature been permissive and low temperature restrictive. Two examples are sweetclover mutants defective in chlorophyll production (Bevins et al., 1993) and the Arabidopsis mutant, fab2, which overproduces the fatty acid stearate (Lightner et al., 1994).

In this report we describe a screening procedure to isolate ts mutations that affect plant growth and development. By applying our screening on an EMS mutagenized population of Arabidopsis, we obtained a collection of twelve growth mutants, each of which potentially corresponds to a novel (essential) plant gene.

Initially, a general phenotypic analysis allowed to classify the twelve mutants into two distinct categories. A first category consisted of eight mutants (dgr1 to dgr8; defective in growth) inhibited in overall plant growth. The observed phenotype was either reversible (for dgr2 to dgr8) or irreversible (for dgr1), which, at the higher temperature, was reflected in growth arrest or lethality, respectively. A second category comprised four mutants (dlr1 to dlr4; defective in lateral root formation) that showed defects specifically in the process of lateral root formation. Of these, two mutants (dlr1 and dlr2) were not able to form lateral roots at all, whereas the remaining two mutants (dlr3 and dlr4) produced smaller lateral roots compared to WT plants. At a permissive temperature of 18°C the phenotypes of the ts mutants were comparable to those of WT plants.

In a next stage, a more detailed phenotypic characterization was focused on two specific mutants, dgr1 and dgr2. Both mutants were selected from the first class, since we were particularly interested in isolating genes essential for the growth and development of the whole plant. Dgr1 was furthermore chosen for its irreversible phenotype, whereas dgr2 was representative for the subclass of mutants with a reversible phenotype. A Mendelian segregation analysis was performed on both dgr1 and dgr2 as a first step towards the isolation of the corresponding genes by positional cloning.

### 2.2 Experimental Procedures

#### 2.2.1 Plant Material

The plant material used in the present study was derived from the Col strain (Col-0) of Arabidopsis thaliana (L.) Heynh that is referred to as the WT in this paper. Chemical mutagenesis was performed with ethyl-methane sulfonate (EMS) (15µM) for 24 hours at 24°C on a population of 200,000 seeds (M1 generation). The M2 progeny was screened for ts mutations. Subsequently, heritability of the observed phenotype was tested with M3 plants produced by self-fertilization of M2 plants of putative mutant lines. The M3 seeds are used to characterize the mutants identified.

#### 2.2.2 Growth Conditions

Seeds were sterilized by a ten-minute incubation with a 1.5% sodiumhypochlorite solution (Merck, Germany). Sterile seeds were sown on MS germination medium (0.44% MS, incl. micro- and macro-elements, no vitamins; 0.05%MES; 1% sucrose; 0.8% agar; pH5.7) (Murashige and Skoog, 1962) in square Petri dishes (120mm x 120mm, Greiner bio-one, Frickenhausen, Germany). After sowing, Petri dishes were placed vertically in a growth chamber under continuous light provided by V.H.O. Sylvania fluorescent white tubes at an irradiance of 50 µmol photons m⁻² s⁻¹ (Osram, Milan, Italy) or by Lumilux Cool White light tubes (Osram, Milan, Italy) at an irradiance of 50 µmol photons m⁻² s⁻¹.
2.2.3 Screening for Arabidopsis Temperature-Sensitive Growth Mutants

EMS mutagenized seeds were subdivided in M1 batches consisting of 800 individual seeds. Seedlings were grown at the permissive temperature of 18°C and were allowed to self-pollinate, which resulted in the corresponding M2 seed batches.

For 150 M2 batches, 1000 seeds of each batch were arbitrarily chosen and grown in vitro for seven days at the permissive temperature. This was done by taking aliquots of 25 milligrams of seeds (±1000 seeds) from each M2 seed batch and washing the seeds for 30 minutes with water and subsequently for five minutes with 70% ethanol. The sterile seeds were sown on sterile strips of filter paper, which were subsequently placed onto MS germination medium. Each collection of 1000 seeds, originating from a single M2 batch, was divided over four Petri dishes, which were placed at the permissive temperature for seven days.

Seedlings were subsequently screened for deficiencies in lateral root formation at the restrictive temperature of 28°C in order to isolate mutants with ts defects in potentially essential genes involved in growth and developmental processes. This was done by transferring the strips of filter paper supporting the one-week old seedlings to new sterile Petri dishes containing the same MS medium supplemented with 10^{-6} M NAA (1-naphtalenecetic acid, NAA, Sigma, N-0640, Bornem, Belgium) after autoclaving. Those Petri dishes were placed at the restrictive temperature. After four days, the seedlings were examined for the presence or the absence of lateral roots. The candidate mutants failing to form normal lateral roots at the restrictive temperature were rescued by transferring them back to the permissive temperature on fresh sterile medium. The seedlings that recovered were transferred to soil to allow self-fertilization and seed production. In order to confirm the ts phenotype, the resulting M3 seeds of these rescued plants were re-screened, using the same procedure.

2.2.4 Plant Phenotypic Analysis

On soil, plants were sown and grown on a mixture of leaf mould, clay and sand in eight-centimeter pots and were watered daily with tap water.

In hydroponic culture, seedlings were grown on Hoagland nutrient solution diluted four times with demineralized water (six plants per liter of solution). No aeration of the solution was needed. The solution was replaced after the first month of culture and subsequently every two weeks.

Under both conditions, mutant and WT plants were grown at the permissive temperature in a growth chamber under continuous light at an irradiance of 50µmol photons m^{-2}.s^{-1} provided by V.H.O. Sylvania fluorescent white tubes (Osram, Milan, Italy) until the first flower primordia were visible at the center of the rosette. They were then transferred to the restrictive temperature and examined two weeks later.

2.2.5 In vitro Growth Studies

2.2.5.1 Auxin Studies

Ten mutant seeds together with five WT seeds were sown on one Petri dish. For each mutant line three Petri dishes were sown. After sowing, the Petri dishes were placed vertically at the permissive temperature for seven days. Subsequently, plants on the first Petri dish were transferred to fresh MS germination medium without NAA and placed at the restrictive temperature. Plants on the second Petri dish were transferred to fresh MS germination medium supplemented with 10^{-4} M NAA and were subsequently placed at the restrictive temperature.
Plants from the third Petri dish were transferred to fresh MS germination medium supplemented with $10^{-4}$M NAA and stayed at the permissive temperature. After four days, pictures were taken using a CAMEDIA C-3040 zoom digital camera (Olympus, Tokyo, Japan).

2.2.5.2 Callus Initiation Experiments (Work partially done by Antoine Adamantidis, Pierre Lejeune and Sandra Ormenese, University of Liege, Sart-Tilman)

Using the following procedure, calli were initiated from root explants (Valvekens et al., 1988). Sterile seeds were sown on MS medium and placed for two weeks at the permissive temperature. Subsequently, the roots of the seedlings were excised and transferred to fresh MS medium supplemented with 0.05 mg/L 6-furfurylaminopurine (kinetin, Fluka Chemika, Germany) and 0.5 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D; Sigma, St. Louis, MO 63178, USA). Half of the explants were then incubated at the permissive temperature and the other half at the restrictive temperature under the light conditions specified above. After six to eight weeks, the explants were examined for the presence of calli and pictures were taken using a CAMEDIA C-3040 zoom digital camera (Olympus, Tokyo, Japan).

2.2.5.3 Growth Analysis of the Primary Root

On each Petri dish fifteen WT and fifteen mutant seeds were sown. For each mutant line, two Petri dishes were sown and placed vertically at the permissive temperature for seven days. Subsequently, the two plates were placed vertically for four days at the restrictive temperature. Eventually only one plate was transferred back to the permissive temperature for five days. After sixteen days, whole plates were photographed using a CAMEDIA C-3040 zoom digital camera (Olympus, Tokyo, Japan) and scanned on a color copier CLC-iR C3200 (Canon, Tokyo, Japan). During these sixteen days, the length of the main root was marked daily by scratching a horizontal line in the Petri dish at the height of the main root tip with a razor blade. Plate scans were analyzed using the image analysis software program Scion Image Beta 4.03 (Scion Corporation, Maryland, USA), the difference in length between each mark was determined. The growth rate for mutant or WT plants was then calculated by dividing the mean length difference of fifteen mutant roots or fifteen WT roots, respectively, by the time elapsed between the two marks.

2.2.5.4 Analysis of Cortical Cell Lengths in the Primary Root

Mutant test plants were grown on MS medium for seven days at the permissive temperature and then transferred to the restrictive temperature for various periods of time (between two and five days), depending on the mutant line. Simultaneously, mutant control plants and WT plants were grown continuously at the permissive temperature and under the same conditions as the mutant test plants, respectively. Subsequently, the seedlings were harvested and fixed for two hours using 4% paraformaldehyde (Merck, Germany) in 20mM Pipes buffer (Merck, St. Louis, MO 63178, USA). After fixation seedlings were washed three times for ten minutes with 20mM Pipes buffer + 1mM CaCl$_2$ + 0.5% Triton X-100. Plants were subsequently mounted on microscope slides in glycerol. Roots were viewed under DIC (differential interference contrast) optics with a 20x objective (Leica DMLB, Germany). A focal plane was chosen by visualizing the quiescent center. For each root a series of overlapping images of the two cortical cell files in this focal plane were captured using a charge-coupled device camera (Zeiss AxioCam, Germany), fitted to the microscope. The series of images
continued until cell length exceeded the width of the video image (when average cell size was approximately 100 µm). Composite images were created (PanaVue ImageAssembler, Canada) and used to measure the length of every cell in each cortical cell file that could be followed over most of its length (ImageJ software). The position of each cell was defined by its midpoint relative to the quiescent center.

For both mutant plants grown at the restrictive temperature, mutant plants grown at the permissive temperature and WT plants grown at both temperatures, cell lengths from the cortical cell files of five roots were combined and then smoothed and interpolated using the procedure described by Beemster and Baskin (1998). The smoothing procedure fitted a series of overlapping, independent polynomials to the data. First, a given number of data points was selected symmetrically around the first desired x, x being the position of each cortical cell relative to the tip of the root. The number of data points used for each fit was determined as the number of cells between the inflection point where cell length started to increase and the most basal data point and ranged from 15 to 25. The value of x was then increased by 25 µm and the process was repeated. For positions at the beginning and end of the series, the data were not symmetric around x but contained the same number of points. Therefore, a second step was introduced, in which the equally spaced data obtained from step 1 were smoothed further using a similar procedure. The length of the interval used for step 2 was defined as the distance between the same inflection point and the quiescent center and ranged between 450 and 550 µm. The iteration in step 2 was repeated until the change in cell length between successive iterations became smaller than 0.5 µm for all positions. The smoothing algorithm was implemented as a macro for the software program Excel (version 2002; Microsoft Corporation).

2.2.6 Back-Crossing and Mendelian Segregation Analysis

Mutant M₃-plants (grown on soil) were back-crossed with WT Col plants (grown on soil) at the permissive temperature. This was done under a binocular (Leica MZ 16F, Germany) using tweezers to transfer pollen from the WT Col plant to the stamen of the mutant M₃ plant. Through selfing of the resulting BC₁-plants BC₁:F₂-seed batches were obtained for each mutant line. The segregation of the BC₁:F₂-families was tested for a 3:1 ratio of WT to mutant phenotypes. Therefore, for each family at least 50 plants were grown for seven days in vitro on MS medium at the permissive temperature and subsequently transferred to the restrictive temperature. For each mutant line a number of plants showing a mutant phenotype at the restrictive temperature were rescued by transferring them to soil and back to the permissive temperature. Self-pollination of these mutant plants led to the production of mutant seed batches for each mutant line of which the genetic background was purified for fifty percent. The homogeneity of the resulting mutant seed batches was tested by growing for each batch at least 50 plants for seven days at the permissive temperature, subsequently transferring these plants to the restrictive temperature, and checking the consistency of the mutant phenotype.

2.3 Results

2.3.1 Identification of Temperature-Sensitive Growth Mutants of Arabidopsis thaliana (results obtained by Antoine Adamantidis, Pierre Lejeune and Sandra Ormenese, University of Liège, Sari-Tilman)

We screened an EMS mutagenized population of Arabidopsis (Col accession) for deficiencies in lateral root formation, when grown at a restrictive temperature of 28°C in the presence...
of auxin (figure 2.1). By selecting for defects in the generation of lateral roots, we aimed to isolate growth mutants in a reasonably fast and straightforward manner. Furthermore, we used exogenous auxin during the entire screening procedure, since it is expected to be easier to distinguish mutant plants that are not able to form any lateral roots from WT plants producing many lateral roots in the presence of auxin.

Initially, 380 candidate mutants were isolated and transferred to the permissive temperature (18°C) to allow self-pollination. The M₃ progenies were re-screened, using the same screening procedure, in order to confirm the heritability of the ts phenotype. Most of the M₃ plants either showed a WT phenotype or did not germinate well. In the end, we obtained twelve ts lines, disrupted in the production of lateral roots at the restrictive temperature (table 2.1).

Based on the mutant phenotype in vitro at the restrictive temperature, mutants were classified into two distinct categories. A first category consisted of
eight mutants (dgr1 to dgr8) exhibiting an overall growth arrest and thus complete inhibition of the development of both shoot and root organs. At the restrictive temperature, dgr2, 3, 4, 5, 6, 7 and 8 showed a reversible phenotype if plants were transferred back to the permissive temperature within four days. Dgr1, on the other hand, showed an irreversible growth arrest followed by lethality within four days at the restrictive temperature (figure 2.2; B and C).

A second class comprised four mutants (dlr1 to dlr4) with deficiencies in the production of lateral roots. Two mutants, dlr1 and dlr2, were not able to form any lateral roots at the restrictive temperature, although growth of all the other plant organs proceeded normally (figure 2.2; D and E). The other two mutants, dlr3 and dlr4, produced smaller lateral roots compared to the WT (figure 2.2; F).

At the permissive temperature of 18°C the phenotype of most mutants was comparable to the WT phenotype. However, three mutants (dgr3, dgr4, and dgr8) were significantly smaller than WT plants at the permissive temperature.

2.3.2 General Phenotypic Analysis of the Twelve Isolated Temperature-Sensitive Mutants

To get a better insight in the different phenotypes of the dgr and dlr mutants, we initially carried out a general phenotypic characterization. To this end, we first examined the effect of exogenous auxin on the mutant phenotypes. Next, the phenotypes of the mutants were studied during growth on soil and in a hydroponic system. Furthermore, to check whether cell proliferation was affected, we tried to initiate callus from mutant root explants. Finally, an analysis of the growth rate of the main root was performed.

2.3.2.1 Auxin Dependency of the Temperature-Sensitive Phenotypes

We examined the phenotype of the twelve ts mutants both in the presence and in the absence of 10^{-6}M exogenous auxin to investigate whether this phytohormone, initially used in the standard screening procedure, has an effect on the mutant phenotype in vitro.

As expected, in the presence and in the absence of exogenous auxin, generation of lateral roots was inhibited in all mutants (figure 2.3; A and B). Our results also show that exogenous auxin
does not induce a changed phenotype in comparison to the WT at the permissive temperature (figure 2.3; C).

2.3.2.2 Mutant Phenotypes at the Vegetative and Reproductive Developmental Stage

To further characterize the phenotype of the isolated mutants, they were grown on soil and in a hydroponic system. The plants were initially cultivated at the permissive temperature and subsequently transferred to the restrictive temperature when the first flower buds appeared. This developmental stage was chosen to examine the ts defects during the reproductive stage. Two weeks after the transfer to the restrictive temperature, plants were examined and photographed.

Grown on soil, dgr mutants showed an inhibition of growth and development at the restrictive temperature but not at the permissive temperature (figure 2.4; B-E). For some mutants, this inhibition was quite severe as in dgr1 (figure 2.4; B, G, and H), where rapid necrosis was observed, and in dgr6 (figure 2.4 C, and M), where growth at the higher temperature induced flower abortion. As also observed in vitro, dgr phenotypes were reversible when the period at the restrictive temperature did not last longer than four days. However, dgr1 was an exception by showing overall growth arrest followed by lethality. In contrast, dlr mutants grown on soil grew better at the restrictive temperature than at the permissive temperature (figure 2.4; A and F).

Cultivation of the mutants in a hydroponic system allowed us to observe the root system and to compare it with WT roots. At the restrictive temperature dgr mutants like for example dgr6, showed disruptions in the formation of secondary roots but also the growth of the root and shoot was affected (figure 2.4 K, and L; figure 2.5). The dgr6 mutant turned necrotic at the apex and did not bolt at 28°C, whereas flowers of dgr3, were aborted and sterile (figure 2.4 I, and J).

2.3.2.3 Effect of the Mutations on Cell Proliferation

To determine whether the ts mutations affect fundamental activities required for cell proliferation, the initiation of callus from root explants was examined in each mutant both at the permissive and at the restrictive temperature. Dgr mutants were unable to initiate calli at the restrictive temperature (figure 2.6). At the permissive temperature, normal callus formation was observed compared to the WT. On the other hand, all dlr mutants initiated callus both at the permissive and the restrictive temperature (data not shown).

2.3.2.4 In Vitro Growth Analysis of the Primary Root Meristem

For the dgr mutants, we performed an in vitro growth analysis of the root apical meristem in order to further characterize their growth behavior at the restrictive temperature. Since the main root growth is not affected in the dlr mutants, these were not included in the experiment. For the eight dgr mutants a growth decrease of the main root and eventually a complete arrest were observed after transfer to the restrictive temperature. The root growth defects were reversible for all mutants except for dgr1 (figure 2.7). In this mutant, lethality occurred after four days at the restrictive temperature. In addition, dgr3, dgr4, and dgr8 showed already at the permissive temperature significant growth rate differences compared to WT roots (data not shown).
Chapter Two – Isolation of Temperature-Sensitive Growth Mutants of Arabidopsis thaliana

Figure 2.3 Effect of exogenous auxin on the phenotype of dgr and dlr mutants. Plants were grown for seven days at 18°C (permissive temperature) in the absence of exogenous auxin. Subsequently, they were transferred to the restrictive temperature (28°C) (A), or they were first transferred to fresh medium supplemented with 10^{-6} NAA and then transferred to 28°C (B) or they were first transferred to fresh medium supplemented with 10^{-6} NAA and further grown at 18°C (C). Pictures were taken four days later. Dotted lines separate wild type (left) from mutant (right) plants. Bar = 1 cm
Chapter Two – Isolation of Temperature-Sensitive Growth Mutants of Arabidopsis thaliana

Figure 2.4 Effect of Temperature on Growth and Development of Different Temperature-Sensitive Mutants.

Results obtained by Antoine Adamantidis, Pierre Lejeune and Sandra Ormenese, University of Liege, Sart-Tilman. Plants were germinated and grown at the permissive temperature of 18°C until the first flower buds appeared. Then they were transferred to the restrictive temperature of 28°C. Pictures were taken two weeks after transfer to the restrictive temperature.

Pictures (A) through (H) (left: restrictive temperature; right: permissive temperature), and (M) illustrate culture on soil; pictures (I) through (L) illustrate hydroponic culture.

(A) WT; (B) dgr1; (C) dgr6; (D) dgr2; (E) dgr7; (F) dkr2; (G) and (H) necrotic dgr1 plants; (I) flowers of dgr3 at the permissive temperature, and (J) at the restrictive temperature; (K) hydroponic cultivation of dgr6 at the restrictive temperature, and (L) at the permissive temperature; (M) flowers of dgr6 at the restrictive temperature.
**Figure 2.5** Root systems of WT (left) and dgr6 (right)

Plants (results obtained by Antoine Adamantios, Pierre Lejeune and Sandra Ormenese, University of Liege, Sart-Tilman). Plants were cultivated in a hydroponic system for three weeks at the permissive temperature followed by two weeks at the restrictive temperature.

---

### 2.3.5 Further Phenotypic Analysis of the Mutants dgr1 and dgr2

In the next experiments as well as in the following chapters three and four we have focussed our efforts on the phenotypic and molecular characterization of the mutants dgr1 and dgr2. We selected these two mutants, defective in overall growth, since the main goal of my research was to isolate genes essential for the growth and development of the whole plant. Furthermore, dgr1 was selected for its irreversible phenotype, whereas dgr2 was chosen as a representative for the subclass of mutants with a reversible phenotype.

---

#### Table 2.1 Summary of Phenotypic Analysis of the 12 Is Arabidopsis Mutants

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th>Plant phenotype</th>
<th>Effect of exogenous auxin</th>
<th>Callus initiation</th>
<th>In vitro growth analysis of the primary root</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>in vitro</td>
<td>On soil</td>
<td>In hydroponic system</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18°C</td>
<td>28°C</td>
<td>18°C</td>
<td>28°C</td>
</tr>
<tr>
<td>Class 1</td>
<td>dgr1²</td>
<td>No lateral roots</td>
<td>Overall growth is inhibited</td>
<td>Formation of normal lateral roots</td>
<td>Normal growth</td>
</tr>
<tr>
<td></td>
<td>dgr2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dgr3¹</td>
<td>Normal growth</td>
<td>Overall growth is inhibited</td>
<td>Formation of normal lateral roots</td>
<td>Normal growth</td>
</tr>
<tr>
<td></td>
<td>dgr4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 2</td>
<td>dgr5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dgr6</td>
<td>Normal growth</td>
<td>No lateral roots</td>
<td>Temperature-sensitive phenotypes are rescued by exogenous auxin</td>
<td>No difference between the mutant phenotypes in the presence or in the absence of exogenous auxin</td>
</tr>
<tr>
<td></td>
<td>dgr7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dgr8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dlr1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dlr2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dlr3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dlr4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In contrast to all the other mutants, dgr3, dgr4, and dgr8 show a slower growth than the WT at the permissive temperature.

³The phenotype of the mutant dgr1 is the most strongly affected by high temperature since lethality already occurs after four days at the restrictive temperature.

*Experiment is not applicable to the mutant.

*No data available.
2.3.5.1 Microscopic Study of the Primary Root Meristem

By means of microscopic visualization of the primary root meristem and cortical cell length measurements going from the root tip to the beginning of the mature root zone, we tried to obtain a better insight in the biological processes that are affected in both the mutants dgr1 and dgr2 (figure 2.8).

No clear differences neither in root structure nor shape were observed between WT and dgr1 roots. Only in a few cases we noticed cells that were slightly misshaped. However, this could be due to plasmolysis, caused by the method of slide preparation.

In contrast, the meristematic zone in dgr2 roots was dramatically reduced at the restrictive temperature. The more distant from the root tip, the more cells resembled WT cortical cells again. Striking is the severely swollen and misshaped nature of the cortical cells in this mutant (figure 2.8).

Cell lengths in cortical cell files were measured going from the quiescent center to the mature root zone. When cell length is plotted in function of root length, it is clear that throughout the root growth zone dgr1 cells are significantly shorter than WT cells. Furthermore, an increased growth zone (meristem and elongation zone) is observed, which has almost twice the length of the growth zone in the WT or in the dgr1 mutant root grown under permissive conditions (figure 2.9).

Dgr2 cortical cells are significantly larger in the growth zone all the way to the quiescent center. It is unlikely that such large cells in the vicinity of the quiescent center are still dividing. Therefore, a reduction of the meristem in the roots could be the cause of the root growth arrest in this mutant. Furthermore, the growth zone in the root is only half of the length of the growth zone in WT or dgr2 mutant roots at the permissive temperature.

2.3.5.2 Mendelian Segregation Analysis

Mutant dgr1 and dgr2 M2-plants were back-crossed with WT Col plants at the permissive temperature. Through selfing of the resulting BC1-plants BC1F2 seed batches were obtained for each mutant line. For both mutants the segregation was scored by screening the individuals of the BC1F2 generation for defects in lateral root formation at the restrictive temperature. None of the tested dgr1 and dgr2 populations showed a segregation ratio different from the expected 3:1 ratio at a 5% significance level (table 2.2).
2.4 Discussion

We isolated twelve novel ts mutants of Arabidopsis thaliana that are defective in growth and development. By selecting for deficiencies in lateral root formation, we aimed to isolate these growth mutants in a reasonably fast and straightforward manner. Furthermore, we used exogenous auxin during the entire screening process, since it is expected to be easier to distinguish mutant plants that are not able to form lateral roots from WT plants producing many lateral roots in the presence of auxin.

A general phenotypic analysis allowed to classify the twelve mutants into two categories: (i) a first category consisted of eight mutants (dgr1 to dgr8) that were inhibited in overall plant growth; the observed phenotype was either reversible (dgr2 to dgr8) or irreversible (dgr1), which, at the higher temperature, resulted in growth arrest or lethality, respectively; (ii) a second category comprised four mutants (dlr1 to dlr4) that showed specific defects in lateral root formation. Of these, two mutants (dlr1 and dlr2) were not able to form lateral roots at all, whereas the remaining two mutants (dlr3 and dlr4) produced smaller lateral roots compared to WT plants. At a permissive temperature of 18°C the phenotypes of the ts mutants were comparable to those of WT plants.

2.4.1 In Δgr Mutants Temperature-Sensitivity is Conferring Throughout Vegetative and Reproductive Development

When grown on soil, Δgr mutants showed an inhibition of growth and development at the restrictive temperature. For some mutants, this inhibition was quite severe as in dgr1, where rapid necrosis and lethality was observed, and in dgr6, where growth at
Chapter Two – Isolation of Temperature-Sensitive Growth Mutants of Arabidopsis thaliana

**Figure 2.8** Cellular Phenotype in the Primary Root of dgr1 and dgr2. Microscopic (DIC) pictures of the main root of WT (A) and the temperature-sensitive mutants dgr1 (B) and dgr2 (C) are shown. After seven days at the permissive temperature, WT, dgr1 and dgr2 plants were transferred to the restrictive temperature for five, two and five days, respectively.

**Figure 2.9** Cortical Cell Lengths in the Primary Roots of dgr1 and dgr2. Dgr1 plants were grown for seven days at 18°C and then for two days at 28°C (A: ▲). Dgr2 plants were grown for seven days at 18°C and then for five days at 28°C (B: ▲). WT plants were grown simultaneously under the same conditions (A and B: •). Control plants (dgr1 (A) and dgr2 (B): ■) were continuously grown at 18°C. Cell length measurements of the cortical cell file were performed starting from the quiescent center to the mature root zone. (symbols represent means +/- SE (n=5), when larger than the data point)
### Table 2.2 Mendelian Segregation Analysis for ts Mutants dgr1 and dgr2

<table>
<thead>
<tr>
<th>Cross</th>
<th>Number of Plants</th>
<th>Experimental Segregation</th>
<th>Theoretical Segregation</th>
<th>$\chi^2$-Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>WT</td>
<td>Mutants</td>
<td>WT</td>
</tr>
<tr>
<td>$\text{ColWT}(1) \times dgr1(1)$</td>
<td>64</td>
<td>48</td>
<td>16</td>
<td>48</td>
</tr>
<tr>
<td>$\text{ColWT}(2) \times dgr1(2)$</td>
<td>70</td>
<td>55</td>
<td>15</td>
<td>52.5</td>
</tr>
<tr>
<td>$\text{ColWT}(3) \times dgr1(3)$</td>
<td>70</td>
<td>57</td>
<td>13</td>
<td>52.5</td>
</tr>
<tr>
<td>$\text{ColWT}(4) \times dgr1(4)$</td>
<td>40</td>
<td>30</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>$\text{ColWT}(5) \times dgr1(5)$</td>
<td>40</td>
<td>29</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>$dgr1(1) \times \text{ColWT}(1)$</td>
<td>70</td>
<td>54</td>
<td>16</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr1(2) \times \text{ColWT}(2)$</td>
<td>70</td>
<td>54</td>
<td>16</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr1(3) \times \text{ColWT}(3)$</td>
<td>70</td>
<td>57</td>
<td>13</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr1(4) \times \text{ColWT}(4)$</td>
<td>70</td>
<td>58</td>
<td>12</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr1(5) \times \text{ColWT}(5)$</td>
<td>70</td>
<td>55</td>
<td>15</td>
<td>52.5</td>
</tr>
<tr>
<td>$\text{ColWT}(1) \times dgr2(1)$</td>
<td>70</td>
<td>55</td>
<td>15</td>
<td>52.5</td>
</tr>
<tr>
<td>$\text{ColWT}(2) \times dgr2(2)$</td>
<td>70</td>
<td>53</td>
<td>17</td>
<td>52.5</td>
</tr>
<tr>
<td>$\text{ColWT}(3) \times dgr2(3)$</td>
<td>70</td>
<td>54</td>
<td>16</td>
<td>52.5</td>
</tr>
<tr>
<td>$\text{ColWT}(4) \times dgr2(4)$</td>
<td>65</td>
<td>52</td>
<td>13</td>
<td>48.75</td>
</tr>
<tr>
<td>$\text{ColWT}(5) \times dgr2(5)$</td>
<td>70</td>
<td>56</td>
<td>14</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr2(1) \times \text{ColWT}(1)$</td>
<td>70</td>
<td>53</td>
<td>17</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr2(2) \times \text{ColWT}(2)$</td>
<td>67</td>
<td>56</td>
<td>11</td>
<td>50.25</td>
</tr>
<tr>
<td>$dgr2(3) \times \text{ColWT}(3)$</td>
<td>70</td>
<td>58</td>
<td>12</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr2(4) \times \text{ColWT}(4)$</td>
<td>70</td>
<td>53</td>
<td>17</td>
<td>52.5</td>
</tr>
<tr>
<td>$dgr2(5) \times \text{ColWT}(5)$</td>
<td>60</td>
<td>41</td>
<td>19</td>
<td>45</td>
</tr>
</tbody>
</table>

* $\chi(0.05;1)^2=3.84$ (for 3:1 ratio)

the higher temperature induced flower abortion and sterility.

Cultivation of the mutants in a hydroponic system furthermore revealed that dgr mutants showed disruptions in the formation of secondary roots but also the growth of the root and shoot was affected. The dgr6 mutant turned necrotic at the apex and did not bolt at 28°C, whereas flowers of dgr3, were aborted and sterile.

These results indicate that the DGR genes are required for normal root and shoot meristem functioning throughout vegetative and flower development.

Although ts mutations provide an excellent opportunity to define the specific developmental stage(s) at which the protein encoded by the mutated gene acts, in Arabidopsis this was only tested in a few studies. The add mutants, for instance, that were isolated as nine-day old seedlings with ts disruptions in apical meristem function, proved to confer temperature-sensitivity throughout the vegetative stage (Pickett et al., 1996). Older add1 plants were arrested in...
development when shifted from low to high temperature, which indicated that ADD1 was required not only for the organization or initiation of the shoot apical meristem, but also for normal meristem function throughout the Arabidopsis life cycle. A similar result was seen with add2 plants.

Our dgr mutants showed temperature-sensitivity throughout the vegetative and the reproductive developmental stage, suggesting that the mutated genes play a vital role during the entire plant life cycle. Such genes may be involved in processes that are required to be permanently active, such as housekeeping genes or genes involved in cell proliferation processes.

2.4.2 Dgr Mutations Affect Processes Essential for Overall Plant Growth

At the restrictive temperature, the overall growth of the plant is inhibited in all dgr mutants, implying that the mutations affect genes involved in fundamental processes of growth and development, possibly the same in shoot and root. The temperature effects are reversible for seven mutants when the shift to the restrictive temperature is less than four days. Only dgr1 shows an irreversible mutant phenotype and becomes necrotic and dies after four days at the restrictive temperature.

Several mutants with an altered morphology of the whole plant have been described in literature. In the glabra2 (gl2) mutant, epidermal patterning is altered both in the root and in the shoot suggesting that GL2 controls epidermal cell fate in the root and the shoot (Hung et al., 1998). Another set of genes appears to perform the same function in root and shoot in ground tissue patterning: SHORT-ROOT (SHR) and SCARECROW (SCR) are required for a normal endodermis (also known as ‘starch-sheath’) in the stem, hypocotyls and the root (Fukaki et al., 1998). These studies indicate that many genes involved in fundamental cellular processes are simultaneously active in several organs of the plant or even throughout the entire plant. However, using classical mutants, only those genes of which the function is not essential to plant survival, will be picked up.

Therefore, the dgr ts mutants, showing an overall growth arrest at the restrictive temperature, and in particular the conditional lethal mutant dgr1, are expected to harbour mutations in genes essential for plant life that are simultaneously active in all organs of the plant. Identifying such genes that function in the fundamental set of processes that are required to produce a viable plant with a normal phenotype, will contribute a lot to the understanding of the basic mechanisms of plant growth and development.

2.4.3 Dlr Mutations Specifically Disrupt Lateral Root Formation

Lateral root formation is specifically affected in the dlr mutants. The mutated genes of dlr1 and dlr2 seem to regulate the initiation steps of the lateral root formation while the mutated genes of dlr3 and dlr4 probably control the subsequent stages during which meristem organization takes place within the primordium. Three aberrant lateral root formation (alf) mutants with altered position, development or number of lateral roots were isolated by Celenza et al. (1995). The alf1-1 mutant forms an increased number of lateral roots. The alf3-1 mutant is defective in the development of lateral root primordia into mature lateral roots such as the phenotype of mutants dlr3 and dlr4. On the other hand, the phenotype of the alf4-1 mutant is comparable to that of mutants dlr1 and dlr2 in which no lateral roots can be formed. The alf3-1 mutant can be rescued by auxin (IAA), whereas this is not possible for the alf4-1 mutant. Since the alf4 mutant shows only defects in lateral root formation...
and not in primary root formation, it is suggested that certain genes exist, which specifically function in the process of lateral root formation.

De Smet et al. (PhD thesis) used a genome-wide transcript profiling approach to identify genes, specifically expressed during lateral root initiation. Based on their transcript profiling, they recovered and analyzed mutants for eighteen genes expressed during lateral root initiation. Mutations in three genes (at2g33830; dormancy/auxin associated family; at4g28640; IAA11 and at2g28620; kinesin motor protein-related) strongly reduced the number of lateral roots per centimeter with more than 40%. However, these mutants also showed a slower growth of the main root.

Since only very few mutants are described that are specifically affected in lateral root formation and show normal growth of all other plant organs, the further characterization of the dlr mutants would be useful to identify the genes with such specific functions.

2.5 Conclusion

In a first aspect, our study demonstrated that ts mutants provide an extremely powerful tool for studying gene function in vivo. They represent a reversible mechanism to lower the level of a specific gene product at any stage in the growth of the organism simply by changing the temperature of growth.

In a second aspect, we have isolated twelve interesting mutants, each potentially corresponding to a novel (essential) gene. Further phenotypic and molecular genetic characterization of the ts growth mutants, described in this chapter, is expected to provide various information, which will be useful for the further understanding of the biological mechanisms underlying plant growth and development.

For the research described in the next chapters, we have focussed our efforts on the molecular characterization of the mutants dgr1 and dgr2. We selected these two mutants, defective in overall growth, since our main goal was to isolate genes essential for the growth and development of the whole plant. Furthermore, dgr1 was selected for its irreversible phenotype, whereas dgr2 was chosen as a representative for the subclass of mutants with a reversible phenotype.
ACKNOWLEDGEMENTS

I would like to thank Dr. Sandra Ormenèse, Prof. Dr. Annie Jacqmard and Dr. Pierre Lejeune for the large amount of work they invested in the mutant screening procedure and the initial phenotypic characterization of the ts mutants.
CHAPTER THREE

MAP-BASED CLONING OF THE TEMPERATURE-SENSITIVE MUTATIONS dgr1 AND dgr2 IN ARABIDOPSIS

In this study, we describe the map-based cloning of two temperature-sensitive mutations, dgr1 (defect in growth 1) and dgr2 (defect in growth 2) that we previously identified by screening an EMS mutagenized population of Arabidopsis. Plants harbouring the dgr1 or dgr2 mutations are rapidly arrested in growth when transferred to the restrictive temperature. Furthermore, dgr1 exhibits irreversible growth defects, eventually resulting in lethality after four days at the restrictive temperature.

Using an AFLP-based genome-wide mapping strategy, we mapped both mutations to a chromosomal segment of the Arabidopsis genome. This approach proved to be very useful in the initial stage of the map-based cloning process, when no knowledge about the location of the gene of interest was yet available.

For the subsequent fine-mapping of dgr1, we used individual PCR-based InDel markers from the Arabidopsis Cereon collection. As a result, we succeeded in cloning the DGR1 gene or at2g45790, and revealed that it encodes an eukaryotic phosphomannomutase. This enzyme was never functionally characterized before in Arabidopsis but has been known to play a crucial role in several fundamental cellular processes, including cell wall biosynthesis, glycoprotein synthesis and plant ascorbic acid biosynthesis in other species. Through our work on the dgr1 mutant, we have cloned and characterized a novel essential plant gene and also provided a new and useful tool for further dissection of the several (potentially unknown) pathways involving phosphomannomutase in Arabidopsis.

The DGR2 gene was mapped to a chromosomal region of 3.5Mb on the top of chromosome II, consisting of NOR2 and the upper telomere. This 3.5Mb region contains only a very small sequenced 38.5 kb fragment preceding the NOR region, hampering further fine-mapping of the DGR2 gene. Despite the fact that NOR2 and the telomeric region largely consist of non-protein coding genes, they fulfil important functions in plant growth and development, which will be discussed in detail in this chapter.
Chapter Three – Map-Based Cloning of the Temperature-Sensitive Mutations dgr1 and dgr2 in Arabidopsis

3.1 INTRODUCTION

Now that the genome of Arabidopsis has been sequenced, this cruciferous plant has emerged as the preferred reference organism for understanding the function of unknown genes in higher plants. The most common method for elucidating the function of unknown genes is to use various mutagenesis procedures, such as insertional mutagenesis (Feldmann, 1991; Koncz et al., 1992; Bancroft and Dean, 1993; Aarts et al., 1995), gene silencing (Baulcombe, 1996; Kooter and Mol, 1993), and physical or chemical mutagenesis (Redei and Koncz, 1992). Gene disruption by insertional mutagenesis can be accomplished by either transposon mutagenesis (Bancroft and Dean, 1993; Aarts et al., 1995) or Agrobacterium-mediated T-DNA transformation (Feldmann, 1991; Koncz et al., 1992). The advantage of insertional mutagenesis is that the gene affected can be identified easily. However, with few exceptions, such as the reported overexpression of PIF3 (for phytochrome-interacting factor) after T-DNA insertion in its promoter region, insertional mutagenesis lacks the versatility to create partial loss-of-function or gain-of-function alleles, which are often very informative in elucidating the biological function of genes. In contrast, chemical or physical mutagenesis can generate simple sequence mutations, such as point mutations or small deletions and insertions, and thus can be used to dissect functional domains of proteins by producing a variety of different phenotypes associated with a specific genetic locus. The major drawback in generating mutations by chemical or physical means is the ability to identify the locus responsible for the mutant phenotype because locus identification, unless limited to a few candidate genes (McCallum et al., 2000), entails map-based cloning, a process that until recently has been considered time-consuming and expensive. However, with the public accessibility of the complete and annotated sequence of the Arabidopsis genome, this situation has radically changed. Chromosome walks, formerly the most tedious and technically problematic steps in the process, are a thing of the past. Nowadays, molecular mapping in Arabidopsis is vastly enhanced by systematically exploiting the available sequence information.

Sequence variation between individuals includes single nucleotide polymorphisms (SNPs) and deletions/insertions (InDels). These make up the backbone for the mapping and positional cloning of genes responsible for phenotypic traits. By tracking DNA sequence variants that cosegregate with heritable traits, the genes giving rise to these traits can be localized to specific chromosomal locations. The majority of known DNA sequence variants in use over the past 20 years have been polymorphisms affecting the recognition sequence of a restriction endonuclease, commonly referred to as restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980). With the development of the polymerase chain reaction (PCR), novel DNA fingerprinting methods were developed, such as CAPS (cleaved amplified polymorphic sequences; Weining and Langridge, 1991; Konieczny and Ausubel, 1993) and AFLP (Vos et al., 1995). The CAPS method uses amplified DNA fragments that are digested with a restriction endonuclease to display any polymorphic restriction sites, and the products are analyzed by agarose gel electrophoresis. Using the AFLP method, sets of restriction fragments are visualized by PCR without knowledge of nucleotide sequence.

In this study, we used an AFLP-based genome-wide mapping strategy (Peters et al., 2001; Peters et al., 2004) to map the two temperature-sensitive (ts) mutations in Arabidopsis, called dgr1 and dgr2. This genome-wide mapping strategy proved to be very helpful at the initial steps of the map-based cloning process, when no knowledge about the location of the gene of interest is available. Briefly, for
each mutation we analyzed approximately 120 F2 segregants with eight primer combinations (PCs), providing a well-dispersed grid of 85 AFLP markers covering the genome. Segregation analysis established linkage to a chromosomal region of less than 5 Mb. For dgr1 and dgr2, segregation analysis of additional AFLP markers in the target region resulted in a chromosomal segment of less than 700 kb and 3.5 MB, respectively.

For the subsequent fine-mapping of the mutated gene the advantage of AFLP (i.e. many physically dispersed markers per PC) is lost and therefore further segregation analyses were based on individual, robust, PCR-based markers (InDels) from the Cereon collection (http://www.arabidopsis.org/Cereon/index.html).

Using the above-described approach, we were able to identify the DGR1 gene as a phosphomannomutase, an essential enzyme functioning in several biological processes, such as glycosylation and vitamin C (ascorbic acid) production in plants. Moreover, we mapped the DGR2 gene to a chromosomal region of 3.5Mb on the top of chromosome II. Unfortunately, sequence information is unavailable for this chromosomal segment, which mainly consists of the highly repetitive nucleolar organizing region (NOR) and the upper telomere of chromosome II.

3.2 Experimental Procedures

3.2.1 Mutagenesis

The plant material used in the present study was derived from the Columbia-0 strain (Col-0) of Arabidopsis thaliana (L.) Heynh that is referred to as the wild type (WT) in this paper. Chemical mutagenesis was performed with ethyl-methane sulfonate (EMS) (15µM) for 24 hours at 24°C on a population of 200,000 seeds (M1 generation). The M2 progeny was screened for ts mutations. Subsequently, heritability of the observed phenotype was tested with M3 plants produced by self-fertilization of M2 plants of putative mutant lines.

3.2.2 Growth Conditions

Seeds were sterilized by a ten-minute incubation with a 1.5% sodiumhypochlorite solution (Merck, Germany). Sterile seeds were sown on MS germination medium (Murashige and Skoog, 1962) in square Petri dishes (120mm x 120mm, Greiner bio-one, Frickenhausen, Germany). After sowing, Petri dishes were placed vertically in a growth chamber (temperature specified below) under continuous light provided by V.H.O. Sylvania fluorescent white tubes at an irradiance of 50 µmol PAR photons m\(^{-2}\)s\(^{-1}\) (Osram, Milan, Italy) or by Lumilux Cool White light tubes (Osram, Milan, Italy) at an irradiance of 50 µmol PAR photons m\(^{-2}\)s\(^{-1}\).

Ts phenotypes were identified in vitro by growing plants for seven days at a permissive temperature of 18°C and transferring them subsequently to a restrictive temperature of 28°C (further experimental details are described in chapter 2).

3.2.3 Plant Material

EMS is a mutagen that typically induces hundreds of point mutations per genome (for more details see chapter one: § 1.3.2.1 and figure 1.11). To avoid phenotypic interference of side mutations (different from the mutation in the gene of interest), it is usually recommended to perform at least one back-cross between the mutant line of interest and the wild type (of the same accession), before initiating the different successive crosses.
and selfings needed for generation of a mapping population (figure 3.1).

For both dgr1 and dgr2 one back-cross cycle was performed. Mutant M3-plants (grown on soil) were backcrossed with Col-0 plants (grown on soil) at the permissive temperature (18°C). This was done under a binocular (Leica MZ 16F, Germany) using tweezers to transfer pollen from the WT Col-0 plant to the stamen of the M3-plant. Through selfing of the resulting BC1-plants, BC1:F2 seed batches were obtained. The segregation of the BC1:F2-families was tested for a 3:1 ratio of WT to mutant phenotypes. To this end, at least 50 plants were grown for seven days in vitro on MS medium at the permissive temperature and subsequently transferred to the restrictive temperature (28°C). Subsequently, BC1:F2-plants showing a mutant phenotype at the restrictive temperature were rescued by transferring them to soil and back to the permissive temperature. Self-pollination of these mutant plants led to the production of dgr1 and dgr2 lines, for which the genetic background was purified for 50 percent. The homogeneity of the resulting mutant lines was tested by growing at least 50 plants for seven days at the permissive temperature, subsequently transferring these plants to the restrictive temperature, and checking the consistency of the mutant phenotype.

Next, to generate mapping populations for both dgr1 and dgr2, backcrossed M3 mutant lines were reciprocally crossed with WT plants of the same accession (e.g. L. ericoides orange). The resulting segregating dgr1 and dgr2 F2 families served as mapping populations, used throughout the subsequent map-based cloning process. Each F2 plant was phenotyped prior to harvesting and being frozen in liquid nitrogen.

**Figure 3.1 Scheme of the different successive crosses and selfings performed for both dgr1 and dgr2 prior to the map-based cloning process.**

Firstly (A), one back-cross cycle was performed for both dgr1 and dgr2 to avoid phenotypic interference of side mutations (•), which are typically induced when using EMS as a mutagen (for more details see chapter one: §1.3.2.1 and figure 1.11). Mutant M3 plants were backcrossed with WT plants of the same accession (e.g. Col: blue; pairs of vertical lines represent homologous chromosomes). After selfing of the resulting BC1-plants BC1:F2 seed batches were obtained. A 3:1 segregation ratio of WT to mutant phenotypes was observed for both dgr1 and dgr2 BC1:F2-families. Selfing of mutant BC1:F2-plants led to the production of back-crossed dgr1 and dgr2 lines, for which the genetic background was purified for 50 percent.

Next, to generate F2 mapping populations for dgr1 and dgr2 (B), backcrossed M3 lines were crossed with the WT of a different accession (e.g. Ler: orange). Through selfing of the heterozygous F1-plants resulting from this cross F2 seeds were obtained. The resulting F2-families segregating for either dgr1 or dgr2, served as mapping populations used throughout the subsequent map-based cloning process.
3.2.4 DNA Extraction

The method for DNA preparation described by Murray and Thompson (1980) (cTAB method) was modified as described below. Individual F₂ plants were ground in separate tubes (Eppendorf; Germany). Immediately after homogenization, the tissue powder was added to 400 µL cTAB-buffer (200mM Tris-HCl pH 7.5; 2M NaCl; 50 mM EDTA; 2% (m/v) cTAB) followed by a 60 minute incubation at 65°C. Samples were subsequently put on ice for 15 minutes. 250 µL chloroform/iso-amylalcohol (24:1) was added and samples were mixed and centrifuged at full speed (13,000 rpm) for ten minutes at room temperature. The upper DNA containing aqueous phase was then transferred to a fresh eppendorf tube containing 200µL isopropanol. Samples were mixed thoroughly and centrifuged for 15 minutes at full speed (13,000 rpm). Hereafter, the supernatants was discarded and the pellet was rinsed with ethanol (70%), air-dried and suspended in 50µL of TE (0.1 mM EDTA).

3.2.5 AFLP Analysis

AFLP template preparation was carried out according to Vos et al. (1995) starting from 100 ng DNA. The restriction enzymes used were SacI and MseI. For the pre-amplifications, a selective MseI primer was combined with a SacI primer, each containing a single selective nucleotide at the 3' end. PCR conditions were as described by Vos et al. (1995). The amplification products were diluted 20-fold and 5 µl aliquots were used for final selective amplifications with SacI+2/MseI+2 PCs and Taq polymerase (Roche). The selective Sacl+2 primers were radioactively phosphorylated using [γ-33P]ATP (specific activity >92 TBq mmol⁻¹; Amersham, Buckinghamshire, UK) and the DNA fragments were separated in a Sequi-Gen GT sequencing cell (Bio-Rad, Hercules, CA). As a size marker, the SequaMark 10 base-ladder (Research Genetics, Huntsville, AL) was used. After electrophoresis, gels were dried on a Hetodry GD-1 slab gel dryer (Heto Lab Equipment, Allerod, Denmark) and visualized using a PhosphorImager 445 SI (Molecular Dynamics, Sunnyvale, CA). Markers were scored using AFLP Quantar Pro™ software (Keygene Products B.V., Wageningen, The Netherlands). AFLP is a registered trade mark of Keygene N.V.

3.2.6 Whole-Genome Mapping Using a Physically Anchored Arabidopsis AFLP Map

In this study, the physically anchored Arabidopsis AFLP map published by Peters et al. (2001) was used. This map was developed based on the genome sequence of Arabidopsis thaliana (Col) (Arabidopsis Genome Initiative, 2000) and the AFLP high-throughput marker technology (Vos et al., 1995). Briefly, for the construction of the physically anchored map, experimental and in silico AFLP were combined, using SacI and MseI as restriction enzymes. As a result, a collection of markers present in Col and absent in Ler (referred to as Col markers), and a database of in silico AFLP fragments, based on the Col-0 genome sequence, were obtained. By combining the size of an experimental AFLP fragment with the selectivity of the AFLP primers and the rough genetic position of the fragment, determined by the segregation in a limited set of Col x Ler RILs (Lister and Dean, 1993), a specific experimentally observed marker could be linked to a specific in silico AFLP fragment in the database. Using this approach, 1266 Col markers were unambiguously positioned on the genome sequence of Arabidopsis thaliana.

Based on the above described map, Peters et al. (2004) defined a set of 8 PCs, generating 85 AFLP markers that are well dispersed over the genome. This
limited set of PCs allowed us to detect the chromosomal segment containing the gene of interest (table 3.1; figure 3.2). Additional AFLP markers positioned in the identified interval were used to further delineate the region of interest. At this stage the benefits of a detailed mapping position must be balanced against the work involved in the AFLP-characterization of a larger number of F2 individuals. Usually, in the case of Arabidopsis, analyzing about 400-450 F2 individuals by the currently available set of 1266 physically mapped AFLP markers results in a region of 400-700 kb containing the gene of interest. Details on the position of the AFLP markers in the Arabidopsis genome can be found on the TAIR website (http://www.arabidopsis.org/).

3.2.7 Fine-Mapping Using Insertion/Deletion (InDel) Polymorphisms

For fine-mapping, a large population of F2-individuals (6,000-10,000) was screened with markers flanking the region containing the gene of interest. We used PCR-based InDel markers from the Cereon collection of InDel polymorphisms (Cereon Genomics, Cambridge, MA, USA; http://www.arabidopsis.org/Cereon/index.html). InDel polymorphisms were identified, and primers developed (using the software Primer3™ version 0.2; Whitehead Institute, Cambridge, UK and Howard Hughes Medical Institute, Chevy Chase, Maryland, USA), to score the flanking markers in a co-dominant way (figure 3.3). The conditions used for PCR amplification were: 2 minutes at 94°C; 10 cycles of 15 seconds at 94°C, 30 seconds at 63°C, 30 seconds at 72°C with a touch-down of 1°C per cycle; 35 cycles of 15 seconds at 94°C, 30 seconds at 53°C, 30 seconds at 72°C; 2 minutes at 72°C. Each time, two reactions were run on a single polyacrylamide gel. Depending on the size of the PCR-products, samples were loaded simultaneously or separated in time.

Since both mutations are recessive, homozygous and heterozygous WT plants could not be distinguished phenotypically in an F2 population. One way to deduce the zygosity of the WT F2 individuals is to generate F3 progeny by selfing each WT F2 individually and check the segregation. This is, however, a time-consuming process. We used an alternative and much faster approach by performing the InDel linkage analyses only on mutant F2 individuals. This way, more F2 plants have to be grown initially to obtain a sufficient number of mutant individuals, but, as opposed to the first option, it saves the time needed to complete an entire Arabidopsis life cycle.

The identified recombinants can subsequently be used to further delineate the area containing the gene of interest with additional markers from the available Cereon InDel and SNP collection, positioned within the region of interest.

3.2.8 Sequence Analysis

Genes (exons and introns) were sequenced as fragments of approximately 500 bp with a minimal overlap of 100 bp.

PCR primers were designed using the Primer3™ software (version 0.2; Whitehead Institute, Cambridge, UK and Howard Hughes Medical Institute, Chevy Chase, Maryland, USA). For amplification of the 500 bp fragments, a PCR was performed with 50 ng genomic DNA template, a final concentration of 0.5 µM for each primer and using Silverstar DNA polymerase (Eurogentec, Liège). The conditions used for the PCR amplification were: 2 minutes at 95°C; 35 cycles of 30 seconds at 95°C, 30 seconds at average Tm, 2 minutes at 72°C; 5 minutes at 72°C; 4°C. PCR products were purified using ExoSAP-IT™ (USB, Cleveland, Ohio, USA). Purified PCR products (ideal final concentration 10ng/100bp) were deposited on the wells of a semi-coated
PCR plate together with 50 µM of primer. The plate was subsequently centrifuged at 4,000 rpm in a 96-well plate centrifuge (Eppendorf, 5810R) and dried at 65°C for 0.5 to 2 hours. Sequencing mix and buffer were added and the plate was spun briefly at 4,000 rpm (Eppendorf, 5810R). A second PCR was then performed under the following conditions: 2 minutes at 95°C; 40 cycles of 10 seconds at 95°C, 10 seconds at 48°C; 2 minutes at 60°C; 4°C. Samples were sequenced with an automated sequencer (ABI3700, ABI377; Applied Biosystems Inc.).

3.3 RESULTS

3.3.1 Map-Based Cloning of dgr1

The dgr1 mutant was identified by screening an EMS-mutagenized population of Arabidopsis thaliana (ecotype Col-0) for ts growth defects at a restrictive temperature of 28°C. At the permissive temperature of 18°C dgr1 plants showed a WT phenotype.

The isolation and phenotypic characterization of the mutant have been described in detail in the previous chapter. After back-crossing dgr1 M3 plants with Col-0 WT plants, a segregation analysis of the BC1F2 families revealed the recessive nature of the mutation (chapter two: table 2.2).

F2 mapping populations were obtained by performing reciprocal crosses of dgr1 mutant plants (Col-0) with Landsberg erecta (Ler) WT plants (figure 3.1). Each F1 cross was checked for heterozygosity with the SSLP marker ‘nga8’ polymorphic between Col and Ler (figure 3.4). F2 seeds were subsequently harvested from those F1 plants confirmed to be heterozygous. The segregating F2 populations were used for the subsequent map-based cloning process (figure 3.5 and 3.6). To this end, the phenotype of each individual F2

![Figure 3.2 Distribution of the 85 Col/Ler AFLP markers over the five chromosomes of Arabidopsis (Peters et al., 2004). Markers were generated by eight AFLP PCs. On average, one Col/Ler AFLP marker is present every 1.3 Mb (6cM). Only 10% of the markers is separated by more than 3.5 Mb, whereas the maximum gap is about 5.6 Mb. Each bold horizontal line represents a marker, whereas a thin horizontal line disrupting the chromosome, stands for a sequence gap.](image)

<table>
<thead>
<tr>
<th>MARKER CODE</th>
<th>SELECTIVE NUCLEOTIDES</th>
<th>NUMBER OF AFLP MARKERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM8</td>
<td>AA</td>
<td>CT</td>
</tr>
<tr>
<td>SM57</td>
<td>AT</td>
<td>GA</td>
</tr>
<tr>
<td>SM61</td>
<td>AT</td>
<td>TA</td>
</tr>
<tr>
<td>SM205</td>
<td>TA</td>
<td>TA</td>
</tr>
<tr>
<td>SM229</td>
<td>TG</td>
<td>CA</td>
</tr>
<tr>
<td>SM233</td>
<td>TG</td>
<td>GA</td>
</tr>
<tr>
<td>SM236</td>
<td>TG</td>
<td>GT</td>
</tr>
<tr>
<td>SM240</td>
<td>TG</td>
<td>TT</td>
</tr>
</tbody>
</table>

1 according to Peters et al. (2001)
Chapter Three – Map-Based Cloning of the Temperature-Sensitive Mutations dgr1 and dgr2 in Arabidopsis

83

Figure 3.3 Screening of a Large F2 population with two flanking InDel markers (After Peters et al., 2004). Two markers, X and Y, represent InDel markers flanking the genomic region harbouring the mutation (+: WT allele; -: mutant allele). The F2 individuals that harbour a recombination event in the region flanked by X and Y can easily be scored, because in all cases but one they produce three PCR fragments instead of two (homozygous F2 non-recombinants) or four (heterozygous F2 non-recombinants). Further delineation of the region of interest is an iterative process, using markers localized more inward. The F2 genotype and phenotype belonging to non-recombinants and recombinants are represented in the bottom panel.

In an exceptional event, four fragments can also be obtained when dealing with a homozygous F2 recombinant individual.

* *, **, *** When dealing with a recessive mutation, homozygous and heterozygous WT plants cannot be distinguished phenotypically in an F2 population. Furthermore, although different in genotype at the mutant locus, homozygous and heterozygous WT plants may be identical in flanking marker genotypes (compare two WT*, two WT**, or two WT***).

In an exceptional event, four fragments can also be obtained when dealing with a homozygous F2 recombinant individual.

* *, **, *** When dealing with a recessive mutation, homozygous and heterozygous WT plants cannot be distinguished phenotypically in an F2 population. Furthermore, although different in genotype at the mutant locus, homozygous and heterozygous WT plants may be identical in flanking marker genotypes (compare two WT*, two WT**, or two WT***).

plant was determined quantitatively by measuring the growth rate of the main root as described in paragraph 2.3.4 of the previous chapter. This phenotypic analysis strategy allowed to efficiently distinguish mutant plants from WT plants that showed coincidental growth deficiencies (figure 3.5).

3.3.1.1 Whole-Genome Mapping of dgr1

In a first step of the genome-wide mapping of dgr1 a total of 126 F2 plants, of which 113 WTs and 13 mutants, together with the parental lines Ler and Col, were analyzed using the set of eight PCs outlined in table 3.1. Segregation of the 85 AFLP markers showed clear/complete linkage to a region on chromosome II (figure 3.8). Due to the distances between the AFLP markers resulting from the eight selected PCs (table 3.1; figure 3.2), the genome-wide mapping procedure is expected to determine linkage of the mutant locus to a specific chromosomal region of less than 5.6 Mb, provided that the required recombinants are present. This is indeed the case for the region containing the dgr1 mutation, which was delimited to 4.7 Mb, an interval delineated by the upper flanking marker SM240_448.2 (figure 3.6) and the lower end of chromosome II.

In a next step, 13 additional PCs, amplifying AFLP markers in the vicinity and within the defined region of interest (lower half of chromosome II), were used for further linkage analysis performed on the same 126 F2 individuals (table 3.2). This way, the region containing the dgr1 locus could be narrowed down to 1.3 Mb with SM26_495.4 as the upper flanking AFLP marker (figure 3.8). Using another eight additional AFLP markers, we did not detect any additional recombination in the 1.3 Mb region. Therefore, as a final step of the AFLP analysis, a larger number of F2 plants, consisting of 138 WTs and 46 mutants, was grown and analyzed with the remaining AFLP markers localized within the defined 1.3Mb interval. This allowed to further map the DGR1 gene to a 673 kb region delineated by the upper flanking marker SM120_153.8 and the
lower flanking marker SM190_142,6 (figure 3.9). For the two remaining AFLP markers SM61_107,7 and SM117_193,1, which are localized within the 673 kb region, complete linkage was detected. At this stage, the benefits of a detailed map position had to be balanced against the work involved in the AFLP characterization of an even larger F2 population. Therefore, fine-mapping of the dgr1 locus using individual robust PCR markers was initiated.

3.3.1.2 Fine-Mapping of dgr1

For the fine-mapping of the dgr1 locus, we grew an F2 population of 9,900 plants. The phenotype of each individual F2 plant was determined quantitatively by measuring the growth rate of the main root as described in §2.3.4 of the previous chapter. Since the DGR1 gene harbours a recessive mutation, homozygous and heterozygous WT plants could not be distinguished solely based on their phenotypes (figure 3.3).

Therefore, to avoid the time-consuming step of selfing the relevant WT recombinants and growing the F3 progeny to determine their genotype, we only used the mutant F2 individuals for further analysis. 1,317 F2 mutants (13.3%) were identified, harvested and submitted to linkage analyses using InDel markers developed from the Cereon InDel collection (table 3.3). Initially, two InDel polymorphisms (CER452389 and CER449256) flanking the previously defined 673 kb interval were identified and used to screen the F2 mutant population. Forty recombinants were identified and subsequently analyzed with six InDel markers localized in the region delineated by the two original flanking InDel markers. As a result, the interval could be narrowed down to 224 kb with CER452385 and CER452346 as the upper and the lower flanking marker, respectively. The remaining nine recombinants were screened with 11 additional InDel markers localized within the 224 kb interval. This resulted in a 35 kb region delimited by CER449813 and CER452353/54 containing the DGR1 gene. Three residual recombinants were analyzed with two remaining InDel markers located in the 35kb area but no additional recombination events could be detected (figure 3.9).

3.3.1.3 Identification of the DGR1 Gene

According to the gene annotations for chromosome II in the publicly available Arabidopsis genome database (www.arabidopsis.org; www.tigr.org), the 35 kb region comprising

![Figure 3.4 Genotyping of F1 plants resulting from reciprocal crosses between dgr1 mutant plants (Col-0) and Ler WT plants.](image)
Chapter Three – Map-Based Cloning of the Temperature-Sensitive Mutations dgr1 and dgr2 in Arabidopsis

**Figure 3.5 Segregation of an F2 mapping population derived from a cross between the Ler WT and the Col1 mutant dgr1 of Arabidopsis thaliana.** Seeds of the parental lines Ler and dgr1, and 32 F2 seeds derived from the cross between the two parental lines were sown in vitro on standard MS growth medium. Plants were germinated at the permissive temperature of 18°C and then grown for 10 days at the restrictive temperature of 28°C. A 3:1 Mendelian segregation ratio of WT to mutant (*) plants is observed. Bar=1 cm.

**Figure 3.6 Segregation pattern of an AFLP marker (SM240_448.2) showing linkage to the dgr1 locus.** Since the dgr1 mutation was generated in a Col background, complete linkage of an AFLP marker to the mutant locus is characterized by a segregation pattern showing a Col fragment (AFLP marker) that is in 25% of the F2 individuals homozygous for absence, in 25% of the F2 individuals homozygous for presence and in 50% of the F2 individuals heterozygous. Genotypic data are reflected in intensities observed after visualization of the radioactively labeled Col fragments using a PhosphorImager. These intensities are measured with the software AFLP-QuantarPro, which allows to assign a genotypic score to each intensity [A: homozygous for absence; B: homozygous for presence; H: heterozygous].

The DGR1 gene encompasses 12 putative genes. To determine which of these 12 genes carried the dgr1 mutation, all of them were sequenced in dgr1 mutants and in Col WTs. By comparing the respective gene sequences from the mutants with those from the WTs, we detected two point mutations, designated dgr1-1 (G19A) and dgr1-2 (G101A), in the gene at2g45790 (DGR1) located in BAC clone F4I18 (figure 3.9; figure 3.10). In the remaining 11 genes no mutations were detected. To further verify the presence of the two mutations, the at2g45790 gene was sequenced in 20 Col WT plants and in the 40 mutant recombinants, which were identified and used during the fine-mapping of dgr1. Both dgr1-1 and dgr1-2 were absent in the WTs and present in all 40 recombinants (data not shown).
Chapter Three – Map-Based Cloning of the Temperature-Sensitive Mutations dgr1 and dgr2 in Arabidopsis

### Table 3.2 Set of Thirteen Additional PCRs Used for the Amplification of AFLP Markers Localized in the Vicinity and Within a 4.7 Mb Region on the Lower Half of Chromosome II Containing the dgr1 Locus.

<table>
<thead>
<tr>
<th>Marker Code</th>
<th>Selective Nucleotides</th>
<th>AFLP Marker(s) of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_{AC1} + 2$</td>
<td>$M_{SE1} + 2$</td>
</tr>
<tr>
<td>SM26</td>
<td>AC</td>
<td>GC</td>
</tr>
<tr>
<td>SM33</td>
<td>AG</td>
<td>AA</td>
</tr>
<tr>
<td>SM34</td>
<td>AG</td>
<td>AC</td>
</tr>
<tr>
<td>SM34</td>
<td>AG</td>
<td>AC</td>
</tr>
<tr>
<td>SM47</td>
<td>AG</td>
<td>TC</td>
</tr>
<tr>
<td>SM54</td>
<td>AT</td>
<td>CC</td>
</tr>
<tr>
<td>SM54</td>
<td>AT</td>
<td>CC</td>
</tr>
<tr>
<td>SM54</td>
<td>AT</td>
<td>CC</td>
</tr>
<tr>
<td>SM89</td>
<td>CC</td>
<td>GA</td>
</tr>
<tr>
<td>SM117</td>
<td>CT</td>
<td>CA</td>
</tr>
<tr>
<td>SM120</td>
<td>CT</td>
<td>CT</td>
</tr>
<tr>
<td>SM120</td>
<td>CT</td>
<td>CT</td>
</tr>
<tr>
<td>SM180</td>
<td>GG</td>
<td>GC</td>
</tr>
<tr>
<td>SM190</td>
<td>GT</td>
<td>TC</td>
</tr>
<tr>
<td>SM235</td>
<td>TG</td>
<td>GG</td>
</tr>
<tr>
<td>SM239</td>
<td>TG</td>
<td>TG</td>
</tr>
<tr>
<td>SM239</td>
<td>TG</td>
<td>TG</td>
</tr>
<tr>
<td>SM244</td>
<td>TT</td>
<td>AT</td>
</tr>
</tbody>
</table>

*Bi-allelic AFLP markers (figure 3.7).*

![Bi-allelic AFLP markers (Figure 3.7) example of the F2 segregation pattern of a bi-allelic AFLP marker.](image)

The DGR1 gene consists of 2590 bp and is organized into 10 exons and 9 introns. The point mutations dgr1-1 and dgr1-2 both occur in the first exon (figure 3.9). Comparison of the WT and mutant cDNAs and their deduced amino acid sequences revealed that the missense mutation dgr1-1 involves the substitution of a glycine residue (GGA) into an arginine residue (AGA), whereas the dgr1-2 point mutation results in replacement of an arginine (codon CGA) by a glutamine (codon CAA) (figure 3.10). The deduced DGR1 protein consists of 246 amino acids with a predicted molecular mass of 27,732 Dalton and a predicted iso-electric point of 5.2.

#### 3.3.1.4 DGR1 Encodes an Eukaryotic Phosphomannomutase

The DGR1 gene (at2g45790) encodes an eukaryotic phosphomannomutase. Phosphomannomutases are phosphotransferases belonging to the recently discovered haloacid dehalogenase (HAD) superfamily of enzymes (Koonin et al., 1994; Aravind et al., 1998). Although sequence identity across the entire HAD family is less than 15%, members share three conserved sequence motifs. They furthermore possess a conserved core...
Chapter Three – Map-Based Cloning of the Temperature-Sensitive Mutations dgr1 and dgr2 in Arabidopsis

Figure 3.8 Linkage of the dgr1 locus with Col/Ler AFLP markers at the bottom of chromosome II and non-linkage with Col/Ler AFLP markers on chromosomes I, III, IV and V. 126 F2 plants, consisting of 113 WTs and 13 mutants, together with the Ler and Col parents, were analyzed using in a first step a set of eight PCs (table 3.1). After scoring the resulting 85 AFLP markers, linkage to chromosome II and non-linkage to chromosomes I, III, IV and V were observed. In a next step, 13 additional PCs, amplifying AFLP markers in the vicinity and within the defined region of interest (lower half of chromosome II), were used for further linkage analysis performed on the same mapping population (table 3.2). This way, the region containing the dgr1 locus could be narrowed down to 1.3 Mb (grey vertical bar) with SM26_495.4 as the upper flanking AFLP marker. Scoring of the markers (A: homozygous for absence; B: homozygous for presence; H: heterozygous) is performed using the software AFLP-Quantar Pro. Markers and their corresponding scores are then sorted according their physical position on the different chromosomes. By assigning a different color to each score (A: ─; B: ─; H: ─–), linkage to a chromosome 2 genomic region could easily be detected.

3.3.2 Map-Based Cloning of dgr2

The dgr2 mutant was isolated in a similar way as dgr1. The isolation and phenotypic characterization of the mutant have been described in detail in the previous chapter (chapter 2; table 2.1). F2 mapping populations
were obtained in a similar way as for dgr1 (see paragraph 3.3.2; figure 3.12 and 3.13).

3.3.2.1 Whole-Genome Mapping of dgr2

For the determination of the chromosomal segment containing the dgr2 mutation, a total of 278 F2 plants were analyzed, consisting of 211 WTs and 67 mutants. Segregation analysis of the 85 AFLP markers showed clear linkage to a genomic region on chromosome II (figure 3.12). The length of the chromosomal segment containing the gene of interest was 3.5 Mb and delineated by the lower flanking marker SM61_322.0 and the upper end of chromosome II. The 3.5 Mb interval consisted of a sequenced 38.4 kb interval, followed by a large non-sequenced area, comprising the nucleolar organizer region (NOR) and the upper telomere of chromosome II (figure 3.14).

3.3.2.2 Dgr2 is Localized in the NOR Region of Chromosome II

The identified genomic interval comprises the NOR region and telomere of chromosome II and largely lacks sequence information due to its highly repetitive nature. This made further mapping of the dgr2 locus unfeasible. At the position of the flanking marker SM61_322.0, 33 recombinants could still be identified, suggesting a genetic distance of 11.9 cM or a physical distance of \( \sim 2.7 \) Mb between this marker and the mutant locus. The 3.5 Mb region contained only a very small sequenced 38.4 kb fragment preceding the NOR region. According to annotations in the Arabidopsis genome database, this fragment comprises two ribosomal rRNAs, six retrotransposons or pseudogenes and only one protein-coding gene, at2g01031 (figure 3.14). This gene encodes a hypothetical protein, functioning in the chloroplasts. To exclude the possibility that the ts dgr2 mutation resides in at2g01031, we determined the sequence of this gene in the dgr2 mutant and in the WT. However, based on the sequence comparison no mutation was detected.
Chapter Three – Map-Based Cloning of the Temperature-Sensitive Mutations dgr1 and dgr2 in Arabidopsis

Figure 3.9 Identification of the dgr1 Gene
A random set of eight AFLP primer combinations was applied to 126 F2 individuals and identified a 47 Mb area at the bottom of chromosome II. Additional mapping with AFLP markers T6 on 184 F2 individuals delimited the DGR1 locus to 673 kb. Subsequently, 1,317 F2 mutant individuals were screened with flanking AFLP markers CER449256 and CER452389, and 40 recombinants were selected (data not shown). These recombinants were used for fine mapping by means of INDELs which delimited the locus to 35 kb flanked by the INDEL markers CER452346 and CER452353/54. (arrows indicate recombinants relevant for further definition of the region of interest). The 35 kb region contained 12 genes and was spanned by the overlapping bacterial artificial chromosome clones F14I18 and F17K2.

The gene structure of DGR1 is outlined with exon represented as boxes. The sites of the two identified dgr1 mutations and the nucleotide changes in the DGR1 gene are indicated.
Chapter Three – Map-Based Cloning of the Temperature-Sensitive Mutations dgr1 and dgr2 in Arabidopsis

3.4 DISCUSSION

Using a genome-wide AFLP-based mapping approach (Peters et al., 2004), we mapped two ts mutations, dgr1 and dgr2, affecting growth and development in Arabidopsis thaliana, to the lower and the upper arm of chromosome II, respectively. We succeeded in further fine-mapping of dgr1 using individual InDel markers derived from the Cereon Arabidopsis Polymorphism collection (http://www.arabidopsis.org/Cereon/index.html) and found that the DGR1 gene corresponds to at2g45790 (figure 3.9), a gene coding for eukaryotic phosphomannomutase in Arabidopsis.

3.4.1 THE TEMPERATURE-SENSITIVE PHENOTYPE OF dgr1 IS DUE TO A DEFICIENCY IN PHOSPHOMANNOMUTASE ACTIVITY

Arabidopsis DGR1/at2g45790 codes for eukaryotic phosphomannomutase, a phosphotransferase belonging to the recently discovered superfamily of haloacid dehalogenases (HAD) (Koonin et al., 1994; Aravind et al., 1998). This family of enzymes comprises a large and diverse set of hydrolases found in the genomes of Bacteria, Archaea and Eukarya. Besides phosphotransferases, the HAD superfamily includes phosphatases, phosphonatases, P-type ATPases, and dehalogenases, which are involved in cellular processes ranging from amino acid biosynthesis to detoxification (table 3.5).
### Table 3.4 Output of a BLAST Search (www.ebi.ac.uk/html) for the Arabidopsis DGR1 Protein/Eukaryotic Phosphomannomutase. An NCBI-BLAST2 Protein Database Query (software WU-BLAST version 2.0; EMBL-EBI) (50 best hits are shown) revealed the existence of several homologues for DGR1 in other species.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Length</th>
<th>Score</th>
<th>Identity%</th>
<th>Positives%</th>
<th>E-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. thaliana</td>
<td>Probable phosphomannomutase (EC 5.4.2.8) (PMI)</td>
<td>246</td>
<td>129</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>O. sativa</td>
<td>GLNHE393P86.16 protein</td>
<td>246</td>
<td>129</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>M. trunculata</td>
<td>Probable phosphomannomutase (EC 5.4.2.8) (PMI)</td>
<td>246</td>
<td>129</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>S. pombe</td>
<td>Man7p</td>
<td>357</td>
<td>815</td>
<td>62</td>
<td>77</td>
</tr>
<tr>
<td>X. laevis</td>
<td>Man8p</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>M. fasciculata</td>
<td>Man7p</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>Man7p</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>N. crassa</td>
<td>Probable phosphomannomutase</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>D. discoideum</td>
<td>Hypothetical protein</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>G. zea</td>
<td>Hypothetical protein</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>C. globosa</td>
<td>Hypothetical protein</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>D. domestica</td>
<td>Hypothetical protein</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>C. albicans</td>
<td>Probable phosphomannomutase (EC 5.4.2.8) (PMI)</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>C. albicans</td>
<td>Probable phosphomannomutase (PMI)</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>Similar to Arabidopsis DGR1 (C. globosa)</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>M. musculus</td>
<td>Similar to Arabidopsis DGR1 (C. globosa)</td>
<td>246</td>
<td>796</td>
<td>64</td>
<td>76</td>
</tr>
</tbody>
</table>

Figure 3.11 Alignment of the deduced amino acid sequence of the Arabidopsis DGR1 gene and several of its orthologues in other plant species, yeast, and animals. Identical, conservative and semi-conserved residues are presented in black, in dark grey and in light grey, respectively. The α/β-core domain (α/β), the cap domain (cap) and the three conserved motifs are indicated by lines above the sequence. Amino acids participating in the catalytic activity of the enzyme are coloured corresponding to their positions in catalytic loop 1 ( ), 2 ( ), 3 ( ) and 4 ( ) of the core domain. Arrowsheads indicate the positions of the mutations dgr1-1 and dgr1-2.
GA), whereas the 

A). Since ts mutations typically cause a 

LOCUS WITH 

PUBSTITUTIONS IN THE 

3.4.1.1 Dgr1-1 and dgr1-2 Cause Amino Acid 

Substitutions in the Periphery of the 

Phosphomannomutase \( \alpha/\beta \)-Core Domain 

Phosphomannomutase is a homodimeric member of the magnesium-dependent phosphotransferase family of 

HAD enzymes, which catalyze cellular 

phosphoryl transfer reactions (Knowles et al., 1980; Mildvan, 1997; Frey et al., 1987; Dzeja et al., 2003). The 

thermodynamics of these reactions vary greatly (Voet and Voet, 1995) but the kinetic barrier is consistently high. 

Phosphotransferases enhance the reaction rate by lowering the energy 

barrier for the reaction via strong binding to a transition state (Lad et al., 2003).

Solution phosphoryl-transfer reactions can proceed by three, 

mechanistically distinct, nucleophile 

substitution pathways (dissociative, 

concerted and associative) (figure 3.15). These 

differ in the timing of the formation 

of the phosphorus-nucleophile bond and 

cleavage of the phosphorus leaving- 

group bond. The prevailing chemical 

pathway is determined by the nature of the 

nucleophile, electrophile and leaving 

group, and by the solvent. There is 

evidence to support the view that all three 

types of chemical pathways are catalyzed 

by phosphotransferases, but the rules 
governing selection of a pathway for a 
given enzyme-reactant pair are not known 

(Catrina and Hengge, 2003; Grzyska et al., 

2003; Anderson, 2001; Lahiri et al., 2003).

All members of phosphotransferase 

family have structures consisting of a 
typical \( \alpha/\beta \)-core domain and a second 
cap domain, with the active site located 
at the domain-domain interface. The core 
domain, responsible for catalysis, has a 

highly conserved tertiary structure, while 

the cap domain is characterized by 
divergent sequences and sizes that confer 
specificity on the enzymes and result in the 
diverse substrate range (Wang et al., 2005; 
Morais et al., 2000). Furthermore, the 
catalytic mechanism of dephosphorylation in this family is believed to involve the formation of a 
phosphoaspartate-enzyme intermediate, which 
could be supported by the three 

conserved motifs in the active-site scaffold 

(figure 3.11; figure 3.16). Motif I has the 
sequence DXDX (T/V). 

![Figure 3.12 Linkage of the dgr2 locus with Col/ler AFLP 
markers at the top of chromosome II and non-linkage with 
Col/ler AFLP markers on chromosomes III, IV and V.](image) 

We detected two point mutations in the 

DGR1 (at2g45790) gene, designated dgr1- 

1 (G10A transition) and dgr1-2 (G101A transition) (figure 3.10). Comparison of the 

WT and mutant deduced amino acid 
sequences revealed that dgr1-1 is 

a missense mutation involving the 

substitution of a glycine residue (codon 

GGA ) into an arginine residue (codon 

AGA), whereas the dgr1-2 point mutation 
results in the replacement of an arginine 

(codon CGA) by a glutamine (codon 

CAA). Since ts mutations typically cause a 
conformational change of the protein in 
the mutant at restrictive (usually higher) 
temperatures, it is plausible that dgr1-1 and/or dgr1-2 could affect the structure of the 

DGR1 protein.
in which the first absolutely conserved aspartate residue forms an intermediate with the substrate (Collet et al., 1998). Motif II contains a conserved serine or threonine residue that forms a hydrogen bond with the phosphoryl oxygen atom (Wang et al., 2001). Motif III consists of a conserved lysine residue followed by a number of less conserved residues and a strictly conserved aspartate residue, which helps to coordinate the magnesium ion required for the enzyme activity (Morais et al., 2000; Wang et al., 2001). Examination of the sequence lengths of the cap domains shows that there are three distinct structural subgroups within the HAD family. The first subgroup includes proteins that contain an inserted cap domain between motif I and motif II, the second, to which phosphomannomutase belongs, includes proteins with a large inserted cap domain between motif II and motif III, and the third contains proteins with little to no inserted domain.

As depicted in figures 3.12 and 3.16, dgr1-1 and dgr1-2 result in amino acid substitutions in the periphery of the α/β-core domain. The changed amino acids will therefore not directly and specifically affect the regions that are essential for substrate binding (substrate-specific loop in cap domain) or catalytic activity (catalytic scaffold in core domain).

What might be expected when dealing with ts mutations is indeed a more subtle interference, since the effect is only observed under certain conditions.

Therefore, we suggest that, at elevated temperatures, the induced amino acid substitutions might disrupt the highly conserved tertiary structure of the core domain, which is responsible for catalysis. This is primarily supported by the fact that both mutations cause relatively drastic amino acid substitutions: the dgr1-1 point mutation causes a change of a neutral aliphatic glycine residue into a positively charged hydrophilic arginine, whereas dgr1-2 results in the substitution of a positively charged basic arginine into a negatively charged acidic glutamine. These changes of charge and/or hydrophilic nature of amino acid residues could affect interactions, needed to maintain the active enzyme conformation but also unbalance enzyme-solvent bonds, needed to maintain optimal enzyme activity. The first is often the case with buried residues, while the latter usually concerns exposed amino acids, positioned at the outer surface of the protein. Although both mutations could equally contribute to the ts phenotype, our results suggest that dgr1-2 is possibly solely responsible. Support can be found in figure 3.11, which depicts the alignment of Arabidopsis phosphomannomutase to a number of its orthologues in other plant species, yeast and animals. It is clear that glycine at position 7 (dgr1-1) is highly variable, whereas arginine at position 37 (dgr1-2) is conserved in all shown species. More extensive BLAST searches revealed that in rare cases the arginine residue is replaced by a lysine, a similar positively charged basic residue.

**Figure 3.13 Genotyping of F1 plants resulting from reciprocal crosses between dgr2 mutant plants (Col-0) and Ler WT plants.** Each pollination of a particular cross was checked by analyzing DNA from leaf material from one F1(y) plant, resulting from the yth pollination in a cross of a Ler(x) parent plant with a dgr1(x) parent, with the Col/Ler SSR marker nga8 (Bell and Ecker, 1994). F2 seeds were subsequently harvested from those F1 plants that were confirmed to be heterozygous by the microsatellite analysis. The resulting segregating F2 populations were used for the map-based cloning process (figure 3.5).
Figure 3.14. The dgr2 gene was mapped to a chromosomal region of 3.5 Mb on the top of chromosome II, consisting of a sequenced (red) 38.5 kb region, the Nor2 (blue) and the upper telomere (dark blue). Grey lines represent AFLP markers used for the mapping of dgr2 on the chromosomal segment. Genes are presented as black boxes. The diagram at the bottom of the figure depicts the tandem arrangement of rRNA genes in the Nor region and the organization of one gene in the array.
Using the software tool SIFT ("Sorting Intolerant From Tolerant"), which predicts, based on sequence homology, whether amino acid substitutions will be tolerated or deleterious for protein function, our empiric observations were confirmed (Pauline and Henikoff, 2001; 2003; http://blocks.fhcrc.org/sift/SIFT.html).

The output also indicated the existence of natural DGR1 homologues, carrying an arginine at position 7.

Based on these arguments that are supported by the results, we suggest that the ts phenotype of dgr1 is due to a conformational change in the tertiary structure of the phosphomannomutase α/β-core domain at elevated temperatures. Furthermore, we strongly...
believe that this conformational change is possibly solely caused by the substitution of the positively charged basic arginine residue at position 37, which is proven to be highly conserved among species, by a negatively charged acidic glutamine residue.

### 3.4.1.2 Biological Functions of the DGR1 Gene Encoding Phosphomannomutase and Effects of Deleterious Mutations

Phosphomannomutase catalyzes the interconversion of D-mannose-6-phosphate and D-mannose-1-phosphate and is an enzyme that participates in cell wall biosynthesis and protein glycosylation. Furthermore, it is believed to catalyze a step in the plant ascorbic acid biosynthetic pathway proposed by Wheeler et al. (1998) (figure 3.17). Despite the importance of these processes, the plant enzyme has only been partially purified in a few species, such as *Amorphophallus konjac* and *Zea mays*.

According to the Arabidopsis microarray database Genevestigator (https://www.genevestigator.ethz.ch/af/), at2g45790 shows a moderate expression in the entire plant. Only in the primary root, the gene is highly expressed. The DGR1 gene is highly conserved among plants (~90% amino acid sequence similarity) and between plants and other eukaryotes (~
75% amino acid sequence similarity) (table 3.4; figure 3.11). It is a single copy gene, as can be expected from a gene with an essential function, and hence has no paralogues. Therefore, we can reasonably assume that the DGR1 gene encodes a phosphomannomutase protein, which enzymatically converts D-mannose-6-phosphate to D-mannose-1-phosphate. The essential nature of the DGR1 gene is consistent with the essential role of the activated form of mannose-1-phosphate, GDP-mannose, a sugar that functions in a number of diverse cellular processes.

Cell wall biosynthesis

Both prokaryotes and eukaryotes utilize GDP-mannose in the synthesis of complex structural carbohydrates. Mannose is a component of microbial lipopolysaccharides and exopolysaccharides. Similarly, GDP-mannose also contributes to the synthesis of three different structural carbohydrates in plant cell walls. First, hemicellulose polymers such as glucomannans and galactomannans, contain D-mannose that is derived from its activated form. Mannans are major wall-based carbohydrate stores in some seeds (Bewley et al., 1997). Second, GDP-mannose is the substrate for GDP-D-mannose-4,6-dehydratase, an enzyme that catalyzes the first step in GDP-L-fucose biosynthesis, which is encoded by the MUR1 gene of A. thaliana (Bonin et al., 1997). L-Fucose is present in both plant cell walls and on glycoproteins. Finally, L-galactose, a minor component of plant cell walls, is synthesized from GDP-mannose (Wheeler et al., 1998).

Protein Glycosylation

In addition to a major function in structural carbohydrate biosynthesis, mannose also plays a key role in protein glycosylation in eukaryotes. D-mannose, the major carbohydrate component of both N- and O-linked glycoproteins, is derived from GDP-mannose during the glycosylation process. Glycoprotein mutants with known or predicted phosphomannomutase deficiencies have been described in yeast and humans.

In yeast, a temperature-sensitive mutant, generated by gene disruption of the SEC53 gene encoding phosphomannomutase, was arrested in growth at the restrictive temperature (Kepes and Schekman, 1988).

In humans, a recessive genetic disease referred to as carbohydrate-deficient glycoprotein syndrome type Ia (CDG-Ia; Jaeken syndrome) results in underglycosylation of secretory proteins, lysosomal enzymes and membrane glycoproteins due to a defect in lipid-linked oligosaccharide biosynthesis (Powell et al., 1994; Krasnewich et al., 1995). It has been reported that cells from patients with this disease are severely deficient in phosphomannomutase activity (Van Schaftingen and Jaeken, 1995). Phosphomannomutase has two isoenzymes in humans with 66% sequence homology: PMM1 and PMM2. They catalyze the same reaction at approximately equal rates, but PMM2 is more substrate-specific (Pirard et al., 1997; 1999). PMM2 is the defective enzyme in CDG-Ia. The clinical features vary with age, and usually go through four stages: the infantile alarming multisystem stage, the childhood ataxia-mental retardation stage, the teenage leg atrophy stage and the adult hypogonadic stage. The overall mortality rate is approximately 20%, mostly occurring in the first years of life due to infection, multiple organ failure or heart tamponade (Jaeken et al., 1991; 2001). Up to now, more than seventy mutations in the PMM2 gene, resulting in CDG-Ia, have been identified, most of them being of the missense type. The absolute absence of patients harbouring two null alleles points to an essential role of PMM2 in normal cellular functions, and suggests that such zygotes are lost in early pregnancy.
L-Ascorbic Acid Biosynthesis

Phosphomannomutase also plays a key role in plant L-ascorbate biosynthesis by converting D-mannose-6-phosphate to D-mannose-1-phosphate in the Smirnoff/Wheeler pathway (figure 3.17 step 3). In plants, a diversity of physiological roles can be ascribed to the antioxidant and cellular reductant vitamin C (L-ascorbic acid) (Smirnoff, 1996).

We suggest that a possible ascorbate deficiency in dgr1 causes accumulation of reactive oxygen species at the higher temperature, leading to a stress-sensitive phenotype.

Decreased growth, observed in the dgr1 mutant, was also seen in the vtc-1 mutant, an ozone-sensitive ascorbic acid mutant, deficient in GDP-mannose pyrophosphorylase (Conklin et al., 1996; 1997; figure 3.17 step 4), and might reflect an effect on the supply of C-skeletons for cell wall biosynthesis. On the other hand, in view of the many correlations between ascorbate and growth and the putative role of ascorbate in the plant cell cycle, part of the effect of low ascorbate could also be due to the disruption of control mechanisms involved in cell division and/or elongation (Noctor and Foyer; 1998a). One mechanism may be accelerated cross linking of cell wall components, which limits cell expansion in the mutant. This notion is supported by observations of Veljovic-Jovanovic et al. (2001) that show a substantial increase in total leaf peroxidase. In addition to the effect of peroxidase induction, the lower extracellular ascorbate concentration in the mutant creates an environment that markedly favors cross linking.

Given the many varied roles of ascorbic acid both as an anti-oxidant and a cofactor, one presumes that mutant plants totally devoid of ascorbic acid would be inviable. This is what we observe for the dgr1 mutant at the restrictive temperature and this was also seen with the embryo-lethal cyt1-2 mutant (allelic with vtc-1), which contained an embryonic lethal mutation that creates a stop codon in the GDP-D-mannose pyrophosphorylase gene (Nickle and Meinke, 1998). However, as mutations such as dgr1 and cyt1-2 are predicted to be pleiotropic, it is difficult to say whether cell wall defects, protein glycosylation alterations and/or ascorbic acid deficiency causes the lethality.

**Figure 3.17 Proposed pathway for ascorbic acid biosynthesis in higher plants (Wheeler et al., 1998).** The enzymatic step catalyzed by the VTC1 gene product is shown as well as branch points to the synthesis of cell wall carbohydrates and glycoproteins. Enzymes: GPI, phosphoglucose isomerase; PMI, phosphomannose isomerase; PMM, phosphomannomutase; GMPP, GDP-D-mannose pyrophosphorylase; GME, GDP-D-mannose-3,5-epimerase; GDH, L-galactose dehydrogenase; GLDH, L-galactono-1,4-lactone dehydrogenase.

### 3.4.2 The dgr2 Mutation Is Localized in the Upper Subtelomeric Region of Chromosome II

Using a genome-wide mapping approach, a 3.5Mb region containing the DGR2 gene was defined at the top of chromosome II. Due to the lack of informative polymorphisms in this region that discriminate ecotype-specific alleles in a mapping population, the mapping of the DGR2 gene could not be continued.

Basically, the chromosomal segment at the top of chromosome II, comprising dgr2, consists of four distinct regions: a very small sequenced region
(38.5kb), the non-sequenced NOR2 (~3.5Mb), the NOR2-telomere junction and the upper telomere (~2.5kb).

First of all, it is highly improbable that the mutant locus resides within the small 38.5kb sequenced fragment. For the flanking marker SM61_322.0, which delineates the 3.5Mb region, 33 recombinants (within a population of 278 F₂ individuals) were identified, implying a genetic distance of 11.9cM or a physical distance of 2.7Mb between this marker and the mutant locus. Furthermore, the sole protein-coding, potential candidate gene in this region, at2g01031, was sequenced and proved to be free of mutations.

A more obvious reasoning would be that dgr2 is localized in NOR2, which largely occupies the 3.5Mb region. In Arabidopsis, approximately 4 million basepairs are devoted to NOR2. At this site (and similarly at NOR4 at the top of chromosome 4), the genes encoding the precursor transcripts of 18S, 5.8S and 25S ribosomal RNA are clustered. These genes are arranged head-to-tail with each transcription unit separated by an intergenic spacer (Brown and Dawid, 1969; Reeder, 1974; figure 3.14). The ~375 Arabidopsis rRNA genes at each NOR evolve rapidly. However, unlike most other gene families that evolve independently, every rRNA gene remains virtually identical in sequence to every other rRNA gene within the individual and even within a population, a phenomenon known as concerted evolution.

Mutations in bacterial, mitochondrial and chloroplast rRNAs resulted in mutant phenotypes and helped to elucidate the molecular mechanisms of translation. However, in all of these cases, only a limited number of rDNA copies are present (Gregory et al., 2005; Murgola et al., 1988; Shen and Fox, 1989; Thompson et al., 2001; Yassin et al., 2005). Triman et al. (1989) isolated ts 16S rRNA mutants of Escherichia coli, causing a disruption of base-pairing in the secondary structure of 16S rRNA at the restrictive temperature. There was a lag period of about two generations between a shift to the restrictive temperature and cessation of growth, implying that the structural defects caused impairment of ribosome assembly. One eukaryotic organism, Tetrahymena, is also well-suited for the isolation of rDNA mutations (Spangler and Blackburn, 1985), since a single copy of the rDNA is present in the germline micronucleus and is amplified in the somatic macronucleus. However, in general, mutations in eukaryotic rRNA genes do not result in a mutant phenotype because of the highly repetitive nature of eukaryotic rDNA. Therefore, it is unlikely that the dgr2 point mutation is localized in one of the rRNA repeats, thereby causing defects in protein synthesis, which would subsequently be expressed in the dgr2 growth phenotype.

Another possibility would be that the DGR2 gene corresponds to an essential protein-coding gene, localized in the NOR2-telomere junction region. Using telomere-specific and 18S rRNA-specific primers Copenhaver and Pikaard (1996) were able to clone the Arabidopsis NOR4-telomere junction. Their results clearly revealed that the NOR4 directly abuts the telomere of chromosome IV. A similar situation has been described in Neurospora, Giardia and Tetrahymena, where telomeric repeats abruptly join the rDNA (Adam et al., 1991; Yu and Blackburn, 1991). The phenomenon is thought to be caused by chromosome breaks that are ‘healed’ by telomerase, which is capable of adding consensus telomere repeats de novo to broken chromosome ends (Blackburn, 1991). Until now, the Arabidopsis NOR2-telomere junction has not been characterized. Therefore, we cannot exclude the possibility that an essential gene, corresponding to DGR2, could reside within this region.
A remaining possible hypothesis is that the ts phenotype of dgr2 is caused by a point mutation in the telomere. Telomeres play an important role in cell growth and senescence and their architecture is well conserved across evolution, consisting of tandem G-rich 6bp (-GGATT- in Arabidopsis) repeats. The 3' terminus of the G-rich strand forms a single-strand overhang, which is used by the enzyme telomerase for the synthesis of unit repeat sequences by the addition of nucleotides with dNTPs as substrates (McEachern et al., 2000).

Because of the repetitiveness in the sequence of telomeres, individual single point mutations generally do not cause a phenotypic discrepancy (Blackburn, 1991). However, exceptions are point mutations induced at the very tip of the telomeres. This has to do with the mechanism by which telomeric length is regulated. Much data support a dynamic model, in which telomerase and the telomere act together as a homeostatic system consisting of both negative and positive regulators of telomere length. Hence, there is a continual dynamic turnover at the telomere tip, but the overall telomere length will stay constant. This makes that, even single base pair substitutions, at the very ends of telomeres can cause loss of regulation of telomeres and loss of the capability of telomeres to cap chromosomes (figure 3.18).

An example of such a mutation was the acc mutation in Saccharomyces cerevisiae, a point mutation in the binding site for Rap1 (a protein regulating telomeric length) in the telomeric repeat. This mutation caused dropping of the binding affinity by a couple of orders of magnitude (Krauskopf et al., 1996). As soon as one or two acc repeats were added to the ends of otherwise WT telomeres, length control was completely lost, and these telomeres became very long and highly deregulated (figure 3.18), which was reflected in a slow growth rate. This study demonstrates that, at least for yeast cells, it is possible to obtain a mutant growth phenotype by inducing point mutations at the very ends of telomeric repeats. However, these findings cannot simply be compared with results obtained in more complex organisms, like Arabidopsis.

Therefore, it is not yet clear in which part of the top of chromosome II dgr2 resides and how it thereby affects the plant phenotype.

**Figure 3.18 Dynamic Turnover at the Telomere Tip (Blackburn, 1991).** Point mutations (X) at the very ends of telomeres can cause loss of regulation of telomeres and loss of the capability of telomeres to cap chromosomes.

### 3.5 Conclusions

In summary, the map-based cloning of the ts mutant dgr1 resulted in the identification of a gene essential for the growth and development of Arabidopsis. DGR1 or at2g45790 encodes eukaryotic phosphomannomutase, a protein that plays a crucial role in several fundamental cellular processes, including cell wall biosynthesis, glycoprotein synthesis and plant L-ascorbic acid production. Molecular complementation of the dgr1 mutant with the wild type
gene is currently ongoing and will provide us with the ultimate proof that DGR1 corresponds to the at2g45790 gene, encoding eukaryotic phosphomannomutase. In the future, dgr1 will serve as an excellent tool in the genetic dissection of these processes. Particularly, our dgr1 mutant will add greatly to our understanding of how plants synthesize ascorbic acid, a pathway which is still poorly understood. Only recently, it gained new attention with the isolation of a series of ozon-sensitive ascorbic acid deficient mutants, which are currently used to further unravel this metabolic pathway.

The DGR2 gene was mapped to a chromosomal region of 3.5Mb on the top of chromosome II, consisting of NOR2 and the upper telomere. Despite the fact that these regions largely consist of non-protein coding genes, they fulfill important functions in overall plant growth and development. Since, due to the lack of informative polymorphisms, no exact map position could be assigned to dgr2, we have tried to characterize this mutant using an alternative approach: the molecular phenotyping of the dgr2 mutant using gene expression analyses will be the subject of chapter four.

With the extensive genome sequence, currently available for Arabidopsis, phenotyping and genotyping remain the critical limiting variables in the process of map-based cloning. The genotyping problem is currently dealt with by hybridizing labeled genomic DNA to the Affymetrix Arabidopsis ATH1 GeneChip (Gresham et al., 2006). Using this approach several high heritability development and circadian clock traits were mapped via direct hybridization of mutant DNA to ATH1 expression arrays (Hazen et al., 2005). Direct detection of small mutations, such as an ethyl-methane sulfonate derived single base substitutions, are still limited by array coverage and sensitivity. However, deletions for example caused by fast neutron mutagenesis are easily detected.
In the present study, we made a first attempt to characterize the temperature-sensitive growth mutant *dgr2* by means of gene expression analyses. We previously mapped the *DGR2* gene to a chromosomal region of 3.5Mb on the top of chromosome II. However, difficulties encountered during the fine-mapping of the *DGR2* gene prompted us to characterize this mutant using an alternative approach. A gene expression analysis using cDNA micro-arrays was performed, allowing us to monitor the transcriptional responses to elevated temperature in both the wild type and the *dgr2* mutant. This study revealed that a diverse set of genes involved in various fundamental cellular processes is affected by the presence of the *dgr2* mutation. We observed that genes encoding proteins involved in protein synthesis, cell wall biosynthesis and cell differentiation are primarily down-regulated in the mutant, whereas those functioning in energy metabolism (respiration), cell death and environmental stress are mainly upregulated. Further characterization of the *drg2* mutant will be required to define the exact nature of the *DRG2* gene. Nevertheless, our results illustrate the power of expression profiling for functional gene characterization in plants, comparable to what has been previously shown by others working on yeast (Hughes et al., 2000) and tumour cells (Van ‘t Veer et al., 2002).
4.1 INTRODUCTION

Whole-genome expression profiling, facilitated by the development of DNA microarrays (Schena et al., 1995; Lockhart et al., 1996), represents a major advance in genome-wide functional analysis. In a single assay, the transcriptional response of each gene to a change in cellular state can be measured, whether it is disease, a process such as cell division, or a response to a chemical or genetic perturbation (Cho et al., 1998; DeRisi et al., 1997; Galitski et al., 1999; Heller et al., 1997; Holstege et al., 1998; Spellman et al., 1998). Because the relative abundance of transcripts is often tailored to specific cellular needs, most expression profiling studies have focused on the genes that respond to conditions or treatments of interest. In the same way, the cellular state can be inferred from the expression profile, and the idea that the global transcriptional response itself constitutes a detailed molecular phenotype has begun to receive attention (Eisen et al., 1998, Gray et al., 1998; Holstege et al., 1998; Marton et al., 1998; Roberts et al., 2000). Provided that the cellular transcriptional response to disruption of different steps in the same pathway is similar, and that there are sufficiently unique transcriptional responses to the perturbation of most cellular pathways, systematic characterization of novel mutants can be carried out using this strategy (Sommerville and Sommerville, 1999).

In the present study, we used this approach to expand our knowledge on the temperature-sensitive (ts) dgr2 mutant of Arabidopsis. To this end, we tried to map biological processes through the identification of sets of genes that showed a differential expression in the mutant relative to the wild type (WT). This way, the collection of genome-wide expression patterns, consisting of thousands of individual observations, results in a molecular phenotype or a transcriptional fingerprint, characteristic for the mutant. During the past five years, the power of this method has been shown by several studies, for instance in yeast (Hughes et al., 2000) and tumors (Van 't Veer et al., 2002). However, the approach has never been applied for the characterization of ts mutations. Therefore, this approach represents an efficient, however challenging, way to characterize unknown mutations, such as dgr2.

4.2 EXPERIMENTAL PROCEDURES

4.2.1 GROWTH CONDITIONS

Seeds were sterilized by a ten-minute incubation with a 1.5% sodiumhypochlorite solution (Merk, Germany). Sterile seeds were sown on MS germination medium (Murashige and Skoog, 1962) in square Petri dishes (120mm x 120mm, Greiner bio-one, Frickenhausen, Germany). After sowing, Petri dishes were placed vertically in a growth chamber (temperature specified below) under continuous light provided by V.H.O. Sylvania fluorescent white tubes at an irradiance of 50 µmol PAR photons m⁻².s⁻¹ (Osram, Milan, Italy) or by Lumilux Cool White light tubes (Osram, Milan, Italy) at an irradiance of 50 µmol PAR photons m⁻².s⁻¹.

4.2.2 PLANT MATERIAL

Mutant dgr2 (Col-0) and WT (Col-0) plants were grown in vitro for nine days at the permissive temperature of 18°C. At day 10 mutant and WT test plants were transferred to the restrictive temperature of 28°C, whereas control plants continued to grow at 18°C.

For the microarray analysis, samples (five to ten shoots per sample) were taken at six consecutive time points...
after the temperature shift (0hrs, 12hrs, 36 hrs, 60hrs, 84hrs, 96 hrs). Three days after the temperature-shift (at time point 84hrs) test plants were rescued by transferring them back to the permissive temperature. A pool of WT control plants served as a reference. A biological repeat was performed, which served at the same time as a color swap.

For the cDNA-AFLP analysis, samples (five to ten shoots per sample) were taken at nine consecutive time points after the temperature shift (0hrs, 0.25hrs, 0.5hrs, 1hr, 2hrs, 4hrs, 6hrs, 8hrs, 10hrs). No rescue step was incorporated. Both a pool of WT (ten shoots) and a pool of mutant (ten shoots) control plants served as a reference.

4.2.3 RNA EXTRACTION AND cDNA SYNTHESIS

Total RNA was prepared from each sample using the LiCl precipitation protocol (Sambrook, 1988).

First strand cDNA synthesis was done by reverse transcription using a biotinylated oligo-dT25 primer (Genset) and SuperscriptII (Gibco-BRL). Second strand synthesis was performed by strand displacement with E. coli ligase (Gibco-BRL), DNA polymerase I (USB) and Rnase-H (USB). Further cleaning of the cDNA occurred by elution through DNA Spin columns (Amersham Pharmacia Biotech).

4.2.4 cDNA AFLP ANALYSIS

Double stranded cDNA (500 ng) was used for AFLP analysis based on the method described by Vos et al. (1995) and Bachem et al. (1996) with the below described modifications. The restriction enzymes used were BstYI and Msel (New England Biolabs) and the digestion was done in two separate steps. After the first restriction digest with BstYI, the 3’ end fragments were collected on Dynabeads (Dynal) by means of their biotinylated tail, while the other fragments were washed away. After digestion with Msel, the released restriction fragments were collected and used as templates in the subsequent AFLP steps. The adapters and primers used are:

| Enzyme | Adapter Primer 1 | Adapter Primer 2 | Primer
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BstYI</td>
<td>5’ – CTCGTAGACTGGGAGTG – 3’</td>
<td>5’ – GATCACTACGCAGTCTAC – 3’</td>
<td>5’ – GACTGCGTAATGATC(T/C)N1-2 – 3’</td>
</tr>
<tr>
<td>Msel</td>
<td>5’ – GACGATGAGTCCTGAG – 3’</td>
<td>5’ – TACTCAGGACTCAT – 3’</td>
<td>5’ – GATGAGTCCTGAGTAAN1-2 – 3’</td>
</tr>
</tbody>
</table>

For pre-amplifications, an Msel-primer without selective nucleotides was combined with a BstYI-primer containing either a T or a C as 3’ most nucleotide. PCR conditions were applied as described (Vos et al., 1995). The obtained amplification mixtures were diluted 600-fold and 5 µL was used for selective amplifications using a P32-labeled BstYI-primer and the AmpliTaq-Gold polymerase (Roche) following the described procedure (Vos et al., 1995). Amplification products were separated on 5% polyacrylamide gels using the Sequigel system (Biorad). Sodium acetate (22 mg/L) was added to the buffer in the bottom tank to avoid run-off of the smallest fragments. Dried gels were exposed to Kodak Biomax films and in an additional approach scanned in a Phosphor-Imager apparatus (Molecular Dynamics).

Bands corresponding to differentially expressed transcripts were cut out from the gel, the DNA was eluted in TE buffer and re-amplified using the same conditions as for selective amplification.

Sequence information was obtained by direct sequencing of the re-amplified PCR product using the selective BstYI-primer. The obtained sequences were compared against nucleotide and protein sequences present in the publicly available databases by BLAST sequence alignments (Altschul et al., 1997).

Scanned Molecular Dynamics gel images were quantitatively analyzed using AFLP-QuantarPro™ (Keygene, The Netherlands). This software has been
designed and optimized for accurate lane definition, fragment detection and quantification of band intensities. All visible AFLP fragments were scored and individual band intensities were measured per lane. The obtained data were used to determine the quantitative expression profile of each transcript. Raw data were corrected for differences in total lane intensities, which may occur due to loading errors or differences in the efficiency of PCR amplification with a given primer combination (PC) for one or more time points. The correction factors were calculated based on constant bands throughout the time course. For each PC, a minimum of ten invariable bands was selected and the intensity values were summed per lane. Each of the summed values was divided by the maximal summed value to give the correction factors. Finally, raw data generated by AFLP-QuantarPro were divided by the correction factors.

4.2.5 Microarray Analysis

Microarray hybridizations were performed at the MAF (MicroArray Facility, VIB, Leuven; http://www.microarray.be/). We used the VIB 6K Arabidopsis microarray, spotted with 6008 Arabidopsis PCR-amplified cDNA-clones or EST’s with a size ranging from 0.5 to 2.5 kb. Each clone was spotted twice.

Test samples of both the WT and the mutant were labeled with the green fluorescent dye Cy3, whereas the reference was labeled with the red fluorescent dye Cy5. According to a reference design each test sample was then hybridized with the reference in a one to one ratio on a slide. Finally, Cy3 and Cy5 intensities were measured for each spot on the slide. A biological repeat was performed and served at the same time as a color swap (Cy5=test en Cy3=control). This resulted for each time point in four observations per genotype.

For the statistical analysis, spot ratios between Cy3- and Cy5-values were calculated. These ratios were subsequently Lowess-fit normalized (Yang et al., 2002). This normalization step corrects the systematically higher Cy3-intensities, especially occurring at low total spot intensities.

Genes with a statistically significant positive expression signal at more than half of the time points in both the mutant and the WT were selected. For each array, the cut-off (95-percentile) for a positive signal was calculated based on 96 (Cy3- and Cy5-) values of 24 Bacillus clones, spotted at different positions on the 6K array. 4797 genes (80%) met to this criterium and were further analyzed. Subsequently, a sequential ANOVA (Wolffinger et al., 2001) was performed, which accounts for experiment-wide systematic effects (e.g. array effects) and extracts appropriate information about the treatment effects, i.e. temperature, genotype, and the interaction which is of primary interest for each gene. We tested for significant differences between these effects by using the Wald-statistic, which is comparable with an F-statistic.

Cluster analysis of the expression data was performed using two methods. In the first method, a hierarchical, average linkage clustering was performed using the Cluster software developed by Eisen and co-workers (1998). For the visualization of the results, the Treeview software (Eisen et al., 1998) was applied. Both software packages can be downloaded for free at http://rana.lbl.gov/EisenSoftware.html. The resulting cluster shows a hierarchical tree, in which each line represents a gene. By clicking on a node, the corresponding cluster can be viewed in more detail. The expression values are presented with color codes for each measured time point. The second method is Adaptive Quality Based Clustering (De Smet et al., 2002) and is accessible for use on http://www.esat.kuleuven.ac.be/~thijs/Work/Clustering.html. As an output, genes are grouped in different clusters and are presented in graphs, where their average
expression pattern is shown in function of time. Both methods require the input expression data to be submitted in a text format (.txt), where a given row indicates data for one single gene and each column represents an expression value for a time point.

4.3 RESULTS

To expand our knowledge about the biological processes that are affected by the ts dgr2 mutation, we performed a gene expression analysis using cDNA microarrays. This study allowed us to identify the transcriptional response to high temperature in the WT and the dgr2 mutation.

4.3.1 MICROARRAY ANALYSIS (1): TRANSCRIPTIONAL RESPONSE TO ELEVATED TEMPERATURES IN WILD TYPE PLANTS

As a first step in the analysis of the microarray data we specifically focussed on the effect of the temperature treatment. Samples of WT plants (five to ten shoots per sample) were taken at six consecutive time points after the temperature shift (0hrs, 12hrs, 36 hrs, 60hrs, 84hrs, 96 hrs). Three days after the temperature-shift (at time point 84hrs) test plants were rescued by transferring them back to the permissive temperature. A pool of WT control plants served as a reference. A biological repeat was performed, which served at the same time as a color swap. We submitted the WT dataset to an ANOVA (analysis of variance) to capture the effects of the temperature shift on gene expression. As a result, 1098 genes were differentially expressed (p<0.001).

The selected genes were categorized by function according to the classification used by MatDB (MIPS Arabidopsis thaliana database). Additional information regarding (potential) gene function was gathered from the literature. A function could be assigned to approximately 40% of the genes (422 genes).

Among these 422 differentially expressed genes all major functional categories were represented (figure 4.1). Almost 50% of the stress and defense genes represented on the microarray were differentially expressed upon the temperature shift. Furthermore, 36% of the genes involved in carbohydrate metabolism, 30% of those functioning in amino acid metabolism and 32% playing a role in energy metabolism were significantly up- or downregulated as a consequence of the temperature changes. The remaining functional classes were less represented and each comprised between 12-18% of the genes present on the array.

We clustered the collection of 1098 differentially regulated genes based on the gene expression profiles using hierarchical (Eisen et al., 1998) and adaptive quality based clustering (De Smet et al., 2002) (figure 4.2). Both clustering methodologies resulted in three distinct groups of genes, characterized by the transcriptional response to the temperature up- and/or downshift (=rescue-step).

A first group of genes shows a rapid and immediate transcriptional response to both temperature changes. Within this group, we were able to distinguish a first subcluster showing induced and a second subcluster showing repressed gene expression at the higher temperature. The first subcluster consists of 59 genes, which are all stress- or defense related. The second subcluster comprises approximately 200 genes belonging to various functional classes, although a vast majority plays a significant role in transcription and energy metabolism.

A second group of genes shows a gradual transcriptional response to the temperature upshift. Again two subclusters were identified, characterized by induced and repressed gene expression.
Chapter Four – Expression Analysis of the Temperature-Sensitive dgr2 Mutant of Arabidopsis thaliana

**Figure 4.1** Functional classification of the genes that were differentially expressed (422) upon temperature increase in WT plants in a microarray experiment. For each functional class, blue bars indicate the percentage of genes represented on the microarray that are differentially expressed upon temperature elevation. For each functional class, red bars represent the percentage of Arabidopsis genes represented on the microarray. Almost 50% of the stress and defense genes represented on the microarray were differentially expressed upon the temperature shift. Furthermore, 36% of the genes involved in carbohydrate metabolism, 30% of those functioning in amino acid metabolism and 32% playing a role in energy metabolism were significantly up- or downregulated as a consequence of the temperature changes. The remaining functional classes were less represented and each comprised between 12-18% of the genes present on the array.

**Figure 4.2** Hierarchical and adaptive quality based clustering of expression profiles modulated by high temperature in WT plants. Both clustering methodologies resulted in three distinct groups of genes, characterized by the transcriptional response to the temperature up- and/or downshift. A first group of genes [...] shows a rapid and immediate transcriptional response to both temperature changes. Within this group, we were able to distinguish a first subcluster showing induced and a second subcluster showing repressed gene expression at the higher temperature. A second group of genes [...] shows a gradual transcriptional response to the temperature upshift. Again two subclusters were identified, characterized by induced and repressed gene expression, respectively. A third group of genes [...] shows a rapid transcriptional response (exclusively) to the temperature downshift. Expression is downregulated in the first and upregulated in the second subcluster.

**Figure 4.3** A plot of the up- and down regulation of genes in response to a temperature shift in WT plants, classified according to function. Gene expression ratios, obtained at a representative time point (60 hrs after the temperature shift), were plotted and sorted by gene function. The category of detoxification and cell defense is mainly characterized by induced gene expression, whereas transport, energy, and carbohydrate metabolism mainly consist of down-regulated genes. Each individual gene is represented by a vertical black bar.
respectively. Both subclusters contain genes acting in a diversity of biological processes.

A third group of genes shows a rapid transcriptional response (exclusively) to the temperature downshift. Expression is downregulated in the first and upregulated in the second subcluster, with the latter consisting of a significant number of cold responsive genes.

Finally, to assess whether up- or down regulation of genes was specific for particular functional classes, we plotted the gene expression ratios, obtained at a representative time point (60 hrs after the temperature shift), and sorted these by the function of the corresponding genes (figure 4.3). The category of detoxification and cell defense is mainly characterized by induced gene expression, whereas transport, energy, and carbohydrate metabolism consist of a vast number of down-regulated genes. Other classes show a similar degree of induced and repressed gene expression.

4.3.2 Microarray Analysis (2): Transcriptional Response to dgr2 in Mutant Plants

In a second microarray analysis, we made a first attempt to map the biological processes, specifically affected by the dgr2 mutation. Mutant samples (five to ten shoots per sample) were taken at six consecutive time points after the temperature shift (0hrs, 12hrs, 36 hrs, 60hrs, 84hrs, 96 hrs). Three days after the temperature-shift (at time point 84hrs) test plants were rescued by transferring them back to the permissive temperature. A pool of WT control plants served as a reference. A biological repeat was performed, which served at the same time as a color swap. To this end, we identified sets of genes that were differentially expressed in the mutant relative to the WT. This way, the collection of genome-wide expression patterns results in a molecular phenotype or a transcriptional fingerprint, characteristic for the mutant.

We selected 463 (approximately 8%) genes differentially expressed (p<0.001) in the mutant compared to the WT. Again, a functional classification was performed according to the MatDB. An accurate function could be assigned to approximately 35% (161 genes) of the selected genes.

To determine which functional classes were significantly over- or underrepresented in our collection of 161 genes, a $\chi^2$-test was performed. For none of the functional classes statistically significant $\chi^2$-values were obtained. Therefore, we tried to assign a more specific function to each gene by subdividing each general category in several subcategories. Seven subclasses appeared to be significantly (p<0.05) over- or underrepresented: intracellular transport, respiration, cell differentiation, peptide transporters, environmental stress, protein synthesis and RNA localization (table 4.1). A broad diversity of processes appear to be significantly affected by the mutation. After calculating the percentages of induced and repressed gene expression, we observed that the categories protein synthesis, cell wall biosynthesis and cell differentiation primarily comprised down-regulated genes, whereas all the other classes mainly consisted of upregulated genes.

4.3.3 Early Transcriptional Response to dgr2 in Mutant Plants Detected by cDNA AFLP Analysis

In order to identify very early effects of the mutation on gene expression, we performed a limited cDNA AFLP analysis. Samples were taken at nine consecutive time points within a time course of 12 hours after the temperature shift.

132 differentially expressed cDNA fragments were visually selected and 106 of these could be efficiently sequenced.
Chapter Four – Expression Analysis of the Temperature-Sensitive dgr2 Mutant of Arabidopsis thaliana

**Table 4.1** Functional Classes of Genes, Found to be Significantly Affected by the dgr2 Mutation After Microarray Gene Expression Analysis.

<table>
<thead>
<tr>
<th>Functional Class</th>
<th>Number of Genes</th>
<th>$\chi^2 (0.05;1)$-Value</th>
<th>Up-/Down-Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell death</td>
<td>4</td>
<td>3.45</td>
<td>Up</td>
</tr>
<tr>
<td>Cell wall biosynthesis</td>
<td>15</td>
<td>3.15</td>
<td>Down</td>
</tr>
<tr>
<td>mRNA transcription</td>
<td>51</td>
<td>3.82</td>
<td>Up</td>
</tr>
<tr>
<td>Proteolytic degradation</td>
<td>15</td>
<td>3.33</td>
<td>Up</td>
</tr>
<tr>
<td>Intracellular transport</td>
<td>13</td>
<td>7.19</td>
<td>Up</td>
</tr>
<tr>
<td>Respiration</td>
<td>15</td>
<td>5.32</td>
<td>Up</td>
</tr>
<tr>
<td>Cell differentiation</td>
<td>15</td>
<td>4.96</td>
<td>Down</td>
</tr>
<tr>
<td>Peptide transporters</td>
<td>4</td>
<td>5.18</td>
<td>Up</td>
</tr>
<tr>
<td>Environmental stress</td>
<td>20</td>
<td>5.46</td>
<td>Up</td>
</tr>
<tr>
<td>Protein synthesis</td>
<td>6</td>
<td>8.10</td>
<td>Down</td>
</tr>
<tr>
<td>RNA localization</td>
<td>8</td>
<td>4.70</td>
<td>Up</td>
</tr>
</tbody>
</table>

$\chi^2 (0.05;1) = 3.84$
The obtained sequences were compared against nucleotide and protein sequences present in the publicly available databases by BLAST sequence alignments.

The corresponding expression profiles were hierarchically clustered (Eisen et al., 1998). This resulted in two types of clusters (figure 4.4).

A first cluster, mainly comprising stress-related genes, shows an induction of gene expression at elevated temperatures both in the mutant and the WT. Within this cluster, an early, brief response and a late, constitutive response were distinguished.

In a second cluster the expression of genes is specifically induced in the mutant (but not in the WT). An up-regulation of genes in this cluster is observed both at high and, although to a lesser degree, at low temperatures. This is in agreement with what we expect when dealing with a ts mutation: the molecular deficiencies are moderately expressed at the permissive temperature, whereas at the restrictive temperature severe molecular defects are induced, giving rise to a visible mutant phenotype. Also in this cluster, an early and a late, constitutive response are distinguished. The first subcluster consists primarily of stress related genes, whereas the second comprises genes involved in transcription and cell wall biosynthesis (figure 4.4).

### 4.4 Discussion

In the present study, we made an attempt to map the biological processes that are affected by the ts dgr2 mutation. A genome-wide expression analysis using cDNA microarrays was performed, which allowed us to identify the transcriptional responses to both the elevated temperatures in the WT and to the dgr2 mutation in the mutant.

#### 4.4.1 Genome-Wide Transcriptional Response to Elevated Temperatures in Wild Type Arabidopsis Plants

1098 of the 6008 genes present on the microarray were differentially expressed in WT plants upon temperature shift. In this collection, genes from all the major functional categories are represented. The large number of stress regulated transcriptome changes observed in our study underscores the difficulty of understanding the global context of a particular stress response. Extrapolating this to the entire genome of Arabidopsis, approximately 5000 genes are potentially affected in their expression upon higher temperature. Because all aspects of plant physiology are impacted by stress, we consider a major change in the transcriptome to be reasonable.

Genes encoding proteins involved in stress/defense were generally upregulated and comprised 22% (91 genes) of the collection of 422 genes to which a function could be assigned. This category included genes encoding proteins of the antioxidant response (gluthathione-S-transferases, peroxidases), the SAR response (pathogenesis-related genes) and cell rescue (heat shock proteins, heat shock transcription factors). Genes that are reported to be induced in response to other stresses, such as drought (Gosti et al., 1995), and dehydration (Kiyosue et al., 1993; Koizumi et al., 1993; Nylander et al., 2001) were also identified in our stress collection of differentially expressed genes, underscoring the overlap between the different stress responses.

Furthermore, genes involved in carbohydrate metabolism are highly represented in our collection of differentially expressed genes (9% or 39 genes). Soluble sugars, especially sucrose, glucose, and fructose, play an obviously central role in plant structure and metabolism at the cellular and whole-organism levels. They are involved in the responses to a number of stresses, and they act as nutrient and metabolite
signalling molecules that activate specific or hormone-crosstalk transduction pathways, thus resulting in important modifications of gene expression patterns. Various metabolic reactions and regulations directly link soluble sugars with the production rates of reactive oxygen species, such as mitochondrial respiration or photosynthesis regulation, and, conversely, with anti-oxidative processes, such as the oxidative pentose-phosphate pathway and carotenoid biosynthesis. Moreover, stress situations where soluble sugars are involved, such as heat shock, chilling, herbicide injury, or pathogen attack, are related to important changes in reactive oxygen species balance. These converging or antagonistic relationships between soluble sugars, reactive oxygen species production, and anti-oxidant processes are generally confirmed by current transcriptome analyses, and suggest that sugar signalling and sugar-modulated gene expression are related to the control of oxidative stress (Couée et al., 2006; Sulmon et al., 2006).

Genes encoding proteins involved in amino acid metabolism constituted (together with carbohydrate metabolism) the second largest group (9% or 39 genes) in the collection. A majority of the genes in this category are involved in proteolysis. Proteolysis of important regulatory proteins is a key aspect of cellular regulation in eukaryotes (del Pozo et al., 1999) and there is evidence that the ubiquitin-proteasome pathway is important in implementation of the plant defense response (Azevedo et al., 2002; Austin et al., 2002; Liu et al., 2002). Proteasome subunit genes are induced in response to stresses and some regulatory subunits were identified in the collection, such as RING finger proteins, which are key components of the ubiquitin-proteasome pathway (Saurin et al., 1996). The RING finger motif is thought to mediate protein-protein interactions and E3 ligase complex assembly. Certain RING finger proteins are rapidly induced by elicitors in Arabidopsis and may be involved in the rapid degradation of regulatory proteins during early stages of pathogen attack (Salinas-Mondragon et al., 1999). Four different genes encoding RING/RING-H2-finger proteins were identified in our collection.

In summary, the collection of high temperature-induced genes comprises a broad repertoire of stress-responsive genes encoding proteins that are involved in both the initial and subsequent stages of the physiological response to stress.

### 4.4.2 Genome-Wide Transcriptional Response to the Temperature-Sensitive dgr2 Mutation

In a second analysis, we identified genes that showed a differential expression in the dgr2 mutant relative to the WT. This way, the collection of expression patterns results in a molecular phenotype or a transcriptional fingerprint, characteristic for the mutant. In the mutant 463 genes (~8%) were differentially expressed (p<0.001).

A very diverse set of fundamental cellular processes is affected by the presence of the dgr2 mutation. Genes involved in protein synthesis, cell wall biosynthesis and cell differentiation are primarily down-regulated, whereas the categories of energy metabolism (respiration), cell death and environmental stress mainly comprise upregulated genes. The large number and diversity of these cellular processes might suggest that the observed gene expression profile corresponds to a secondary physiological response to the mutation, which makes it very difficult to pinpoint the processes that are primarily affected by the mutation.

Our previous results outlined in chapter three show that dgr2 maps to a 3.5 Mb region on the upper part of chromosome II. We postulated different hypotheses concerning the location of the gene in the NOR region, in the NOR-telomere junction or in the upper telomere of chromosome II. In higher eukaryotes, no rRNA mutants have been isolated (or functionally characterized) until now due
to the high copy number of rDNA in these species. Furthermore, in most eukaryotic species, the NOR-telomere junction region has been poorly characterized. However, telomere-deficient mutants in yeast (Saccharomyces cerevisiae) have been described by Nautiyal et al. (2002). They characterized these mutants by monitoring the genome-wide transcriptional response, whereby sets of genes were defined that are involved in cellular responses to altered telomere structure or function. When comparing our expression data from the dgr2 mutant with the results described in this report, some coincidence is observed. Nautiyal et al. report a marked up-regulation of energy production genes, more specifically involved in oxidative phosphorylation. Furthermore, they observed an up-regulation of approximately 50% of the genes involved in the environmental stress response. Finally, a vast majority of the genes involved in protein synthesis was reported to be down-regulated.

These results are similar to what we have observed in the dgr2 mutant. Additionally, genes involved cell differentiation and cell wall biosynthesis are down-regulated in dgr2, which is also known to be the case when telomere function is disturbed and which eventually results in senescence and cell death. The latter process was also significantly represented in the collection of differentially expressed genes in dgr2.

Despite the similarities between the transcriptional responses of the dgr2 mutant and the yeast telomere-deficient mutants, we have to take caution making this comparison since the observations were made in different systems and under different conditions.

4.4.3 Early Effects of the dgr2 Mutation on the Transcriptional Response in the Mutant

Limited cDNA-AFLP analysis allowed to identify induced gene expression in the mutant as early as 30 minutes after the temperature shift.

Based on the identity of the selected differentially expressed genes, we observed that a very early and a late general stress response are induced upon temperature increase in both the mutant and the WT. This can be considered as the onset of the general response to heat stress. Furthermore, mutant-specific stress-related genes are induced, which reveals that the mutant is more stress-sensitive than the WT. Further conclusions concerning gene expression patterns based on this limited study would not be accurate and should be based on a more detailed genome-wide analysis, using the early time points that we have defined as informative. This way fewer secondary effects will be encountered.

4.5 Conclusion

We applied genome-wide expression profiling to identify the biological processes affected by the dgr2 mutation, and hence, indirectly obtain a vision on the potential function of the DGR2 gene. The results we obtained do indeed show that several crucial cellular processes are affected by mutating the DRG2 gene. It is the first time that this approach has been applied to the characterization of temperature-sensitive mutations. Whilst the expression data did not allow us yet to assign a clear-cut function to the DGR2 gene, our results already give an indication of the power of expression profiling for functional gene characterization, comparable to what has been previously shown by others working on yeast or tumor cells. Further characterization of the dgr2 mutant will be required to define the exact nature of the DRG2 gene and to pinpoint the processes that are primarily affected by the dgr2 mutation.
CHAPTER FIVE

CONCLUSIONS AND PERSPECTIVES
CONCLUSIONS AND PERSPECTIVES

In this work we have studied a collection of 12 newly isolated temperature-sensitive Arabidopsis mutants and performed a detailed molecular characterization of two of these mutants, named dgr1 and dgr2. These two mutants, which are both defective in overall growth at the restrictive temperature, were chosen to test the feasibility of using map-based cloning to identify novel genes essential for the growth and development of the plant. Whereas dgr1 was selected for its irreversible defective phenotype, dgr2 was chosen as a representative of a mutant subclass with a phenotype that is reversible when temperature is switched back to normal levels.

We were successful in fine-mapping and cloning the DGR1 gene, and have provided evidence that it encodes a eukaryotic phosphomannomutase, an enzyme that had not yet been functionally identified in Arabidopsis. Homologues of this enzyme in other eukaryotic species have been shown to play a crucial role in several fundamental cellular processes, including in cell wall biosynthesis and glycoprotein synthesis. The plant enzyme is thought to be involved in L-ascorbic acid production. Molecular complementation of the dgr1 mutant with the wild type gene is currently ongoing and should provide the ultimate proof that DGR1 corresponds to the at2g45790 gene, encoding eukaryotic phosphomannomutase. In the future, functional enzymatic characterization should allow to assign a role to at2g45790 in cell wall biosynthesis, vitamin C production and/or protein glycosylation. If this is the case, dgr1 will serve as an excellent tool in the genetic dissection of these processes. Particularly, our dgr1 mutant will add greatly to our understanding of how plants synthesize ascorbic acid, a pathway which is still poorly understood. Only recently, it gained new attention with the isolation of a series of ozon-sensitive ascorbic acid deficient mutants, which are currently used to further unravel this metabolic pathway.

The DGR2 gene mapped to a chromosomal region of 3.5Mb on the top of chromosome II. This 3.5Mb region comprises one of the few segments of the Arabidopsis genome that has not been sequenced because it consists mainly of rDNA repeats, the NOR region. Because of the lack of sequence information there are no molecular markers available in this region, which hampered further fine-mapping of the DGR2 gene. Difficulties encountered in the mapping of the DGR2 gene prompted us to explore an alternative approach to characterize the dgr2 mutant. We applied genome expression profiling using cDNA micro-arrays to identify the biological processes affected by the dgr2 mutation, and hence, indirectly obtain a clue of the potential function of the DGR2 gene. A genome-wide expression analysis using cDNA micro-arrays was performed comparing the transcriptional responses to elevated temperatures in both wild type and the dgr2 mutant. The results we obtained do indeed show that a diverse set of genes involved in various fundamental cellular processes is affected by the dgr2 mutation. Since the genome-wide expression data did not allow us yet to assign a clear-cut function to the DGR2 gene, further characterization of the dgr2 mutant will be required to define the exact nature of the DGR2 gene and to pinpoint the processes that are primarily affected by the dgr2 mutation.

Ultimately, to be able to fully characterize dgr2 gene, the underlying mutation would have to be identified by DNA sequencing. The ~3.5Mb region on the top of chromosome II, to which the dgr2 mutation was mapped, comprises the non-sequenced NOR2 region, the NOR2-telomere junction and the upper telomere. Sequencing the NOR2 region would not
only be tremendously time-consuming and expensive, but also unlikely to succeed because it would require finding single base differences between wild type and mutant within a 3.5 Mb segment, which is technically not feasible. Alternatively, the TILLING approach (Targeting Induced Local Lesions in Genomes) might provide a way to determine whether a point mutation is present in this area or not. TILLING is a method that allows a directed identification of mutations in a specific gene or nucleotide sequence, and could be used to search for differences between wild type and mutant. Also this approach may not succeed because of the highly repetitive nature of the sequence. Furthermore it is difficult to imagine how a mutation in one of the ~300 ribosomal DNA repeats could lead to a temperature-sensitive phenotype. Consequently, the most likely hypothesis is that the dgr2 mutation lies in either in the NOR2-telomere junction region or the telomere itself. Until now, there is no experimental evidence for the presence of protein coding genes in NOR-telomere junction regions. However, to exclude this possibility a fragment bridging the NOR2 and the telomere could be sequenced using ribosomal repeat and a telomeric repeat specific primer. The upper telomere, being ~5kb in length, could be sequenced using telomere specific primers, designed based on the sequence of the telomeric repeats.

Our work has illustrated that the identification of novel essential genes can be carried out using the classical forward genetics approach of characterizing temperature-sensitive mutants. Especially for organisms such as Arabidopsis, for which sufficient genetic resources including high density genetic maps and a wealth of DNA polymorphisms are available, map-based cloning of genes is nowadays a relatively straightforward procedure. Nonetheless, map-based cloning of randomly induced mutations remains very laborious and time consuming, and as we experienced with dgr2, does not always succeed. The principal rate-limiting steps in map-based cloning are the requirement for (multiple) back-crosses, the generation of large mapping populations, and ultimately the sequencing of the candidate genes. Consequently, despite the tremendous technical progress, map-based cloning is not yet a high-throughput approach that can be used to characterize large numbers of genes. With the elucidation of the Arabidopsis genome sequence and the development of novel reverse genetic tools, one of the ultimate goals of Arabidopsis research is the systematic characterization of the function of all genes in a high-throughput manner. In this perspective, the high throughput identification of genes that are essential for growth and development remains an important target. For this, the recently developed inducible or conditional reverse genetic systems, such as conditional RNAi induced inhibition of gene expression, could provide a more efficient way to identify novel essential genes in Arabidopsis in the future.
CHAPTER SIX

SAMENVATTING
Gedurende dit onderzoek hebben we een collectie van 12 nieuwe temperatuursensitieve Arabidopsis mutanten geïsoleerd. Twee van deze mutanten, namelijk dgr1 en dgr2, werden geselecteerd voor een verdere gedetailleerde moleculaire karakterisering. Deze twee mutanten, beide verstoord in hun algemene groei en ontwikkeling bij de restrictieve temperatuur, waren een ideaal onderzoeksmiddel om nieuwe genen te identificeren die essentieel zijn voor groei en ontwikkeling van een gehele gedifferentieerde plant. Dgr1 werd specifiek geselecteerd voor zijn irreversibel fenotype, terwijl dgr2 werd gekozen als representatief exemplaar voor een subklasse van mutanten die alle een reversibel groeifenotype vertoonden wanneer ze terug naar een normale temperatuur werden overgeplaatst.

We zijn erin geslaagd het DGR1 gen te karteren en vonden dat het gen codeert voor een eukaryotisch fosfomannomutase. Het is de eerste keer dat dit enzyme functioneel werd gekarakteriseerd in Arabidopsis. Homologen van dit enzyme in andere eukaryoten waren gekend en spelen een cruciale rol in diverse fundamentele celulaire processen, zoals de celwandbiosynthese, glycoproteïnesynthese en vitamine C produktie in planten.

Het DGR2 gen werd gekarteerd tot een chromosomaal segment van 3.5 Mb op de top van chromosoom II. Dit gebied bevatte slechts een zeer klein gesequeneerd 38.5 kb fragment voorafgaand aan de NOR regio en het telomeer van chromosoom II, wat verdere fijnkartering van het DGR2 gen onmogelijk maakte. Daarom poogden we om deze mutant te karakteriseren, gebruik makende van een alternatieve strategie. We voerden een genoomwijde expressie analyse met cDNA microarrays uit, wat ons toeliet de globale transcriptionele respons tengevolge van temperatuursverhoging te observeren in het wild type en in de dgr2 mutant. Deze studie toonde aan dat een diverse set van genen, betrokken in allerlei fundamentele celulaire processen, werd geaffecteerd door de aanwezigheid van de dgr2 mutatie. Deze geneexpression-data op zich lieten ons nog niet toe om de specifieke biologische processen die in eerste instantie zijn geaffecteerd door de dgr2 mutatie, te capteren. Verdere (genoomwijd) karakterisering van de dgr2 mutant is dan ook vereist om de exacte aard van het gemuteerde gen te definiëren.

Ons onderzoek benadrukt de voordelen van het combineren van methoden uit de klassieke genetica en uit het moderne genomics tijdperk. Traditionele technieken, zoals het screenen van EMS mutanten voor specifieke fenotypes, waren vroeger reeds gekend als succesvolle methoden om genen te identificeren, betrokken in fundamentele processen van ontwikkeling. Verder heeft het gebruik van temperatuursensitieve mutanten ons toegelaten om planten, die ernstig verstoord waren in groei en ontwikkeling, op te pikken en te bestuderen. Door gebruik te maken van klassieke positionele klonering, een methode die in Arabidopsis aanzienlijk werd vergemakkelijkt door de beschikbaarheid van grote hoeveelheden aan sekwentie-informatie en genetische kaarten met hoge resolutie, kon het DGR1 gen worden gekloneerd. Door de moeilijkheden bij het karteren van het DGR2 gen, werden we gedwongen een alternative strategie te exploreren om de dgr2 mutant te karakteriseren. We voerden een genoomwijd expressie-analyse uit gebruik makende van cDNA microarrays om de biologische processen die geaffecteerd werden door de dgr2 mutatie te identificeren en zodanig onrechtstreeks een indicatie te verkrijgen over de mogelijke functie(s) van het DGR2 gen. Onze resultaten tonen inderdaad aan dat verschillende essentiële processen worden beïnvloed in de dgr2 mutant bij hoge temperatuur. Het is de eerste keer dat deze aanpak wordt gebruikt voor het karakteriseren van relatief complexe temperatuursensitieve mutaties. Desalniettemin illustreren onze
resultaten reeds de kracht van expressie-analyse om mutanten in planten te karakteriseren op moleculair niveau, vergelijkbaar met wat reeds aangetoond werd door anderen in diverse organismen.


Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stem, D., Winkler, J., Lockhart, D.J., Morris, M.S.,

8, 549-546.

interactions between, wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis.
Plant Physiology 129, 661-667.


Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L.,
Stammers, M., Federspiel, N., et al. (1999). Genome-wide mapping with biallelic markers in Arabidopsis


Arabidopsis thaliana. Plant J. 16, 735-743.

phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX T/V motif.


GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis. Plant Biology 96, 4198-4203.

mutants. Genetics 154,847-856.

Cooke, R. et al. (1996). Further progress towards a catalogue of all Arabidopsis genes: analysis of a set of 5000 non-
redundant ESTs. The Plant Journal 9(1), 101-124.


References


References


References


References


References


