

Institut für Biochemie und Biologie
Arbeitsgruppe Prof. Dr. Bernd Müller-Röber

Identification of
growth-related tonoplast proteins
in *Arabidopsis thaliana*

Dissertation
zur Erlangung des akademischen Grades
„doctor rerum naturalium“
(Dr. rer. nat.)

in der Wissenschaftsdisziplin „Molekularbiologie“

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von

Samuel Janne Arvidsson
geboren am 15.12.1979 in Örkelljunga, Schweden

Potsdam, Oktober 2010

Published online at the
Institutional Repository of the University of Potsdam:
URL <http://opus.kobv.de/ubp/volltexte/2011/5240/>
URN <urn:nbn:de:kobv:517-opus-52408>
<http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-52408>

Declaration

I hereby declare that this Ph.D. thesis is the result of my own work carried out between February 2007 and September 2010 in the group of Prof. Dr. Bernd Müller-Röber at the University of Potsdam in Potsdam-Golm, Germany, written by me alone using solely the cited sources and described methods. Additionally I declare that I have not until now tried to submit or defend a Ph.D. thesis, neither at the University of Potsdam nor at any other university.

Potsdam, 08.10.2010

Samuel Arvidsson

Acknowledgements

First of all, I would like to express my gratitude to Prof. Dr. Bernd Müller-Röber, who made it possible for me to come to do my Ph.D. work in his lab, and for his motivation and supervision during the years here. Also, thanks to Dr. Katrin Czempinski, I was able to take part in the VaTEP project, which she organized with excellence; thanks to this project I have had many possibilities to interact with and visit other research groups as well go to many interesting conferences. Also, I'd like to thank the reviewers of this thesis for taking the time and effort.

Another big thanks goes to the greenhouse personnel, especially Christiane Schmidt and Doreen Mäker, for taking good care of the plants and offering a lot of help for my work. I'd also like to thank the Greenteam at the MPI, mainly Karin Köhl and Linda Bartetzko for their help with the plants I grew there. Also, without the work of Sebastian Kopp, Franziska Jeschke, Paul Saffert and Elisa Schulz, who helped out with genotyping and various other lab tasks, I wouldn't have managed to obtain all the data presented in this thesis.

Thanks to the EU-funded project VaTEP ('Vacuolar transport equipment for growth regulation in plants', MRTN-CT-2006-035833) I had funding for the first three years of Ph. D. studies, and thanks to the BMBF-funded project GoFORSYS ('Potsdam-Golm BMBF Forschungseinrichtung zur Systembiologie. Photosynthesis and Growth; a Systems Biology Based Approach', FKZ 0313924), I had time to finish up the experiments and thesis writeup.

I would also like to thank Prof. Dr. Diego Mauricio Riaño-Pachón, Dr. Mirosław Kwaśniewski, Dr. Paulino Pérez-Rodríguez, Flavia Vischi Winck, Magda Łotkowska and Muhammad Arif for the nice collaborations and contributions to the work presented in this thesis, and all other current and former Mü-Rös, especially Dr. Luiz Gustavo Guedes Corrêa and Dr. Fernando Alberto Arana Ceballos for the nice time in and outside of the lab.

Last, but certainly not least, I would like to thank my family for their support during these and former years and especially my wife Sandra, for her continuous love, support and understanding for the long working days, the late nights and weekends in front of the computer writing up this thesis.

Also, I would like to thank all others who have helped me and were not mentioned here.

Table of contents

Chapter	Title	Page
	Declaration	i
	Acknowledgements	iii
	Table of contents	v
	List of figures	vii
	List of tables	viii
	List of abbreviations	ix
	Summary	xi
1	General introduction	1
1.1	Introduction to plant vacuoles	1
1.2	Diversity of plant vacuoles and their different roles	3
1.3	The tonoplast and its constituents	4
1.4	The role of the vacuole and the tonoplast in growth	5
1.5	Leaf cell elongation and leaf organ size control	6
1.6	Expression analysis with real time RT-qPCR	7
1.7	Leaf growth phenotyping	12
1.8	Aims and structure of the thesis	15
2	High-throughput RT-qPCR primer design	17
2.1	Authors' contributions	17
2.2	Abstract	18
2.3	Background	18
2.4	Implementation	19
2.5	Results	28
2.6	Discussion	29
2.7	Conclusion	29
2.8	Methods	29
2.9	Availability and requirements	31
2.10	Authors' contributions	31
2.11	Acknowledgements	31
2.12	References	31
3	A growth phenotyping pipeline for <i>Arabidopsis thaliana</i>	33
3.1	Summary	33

Table of contents

3.2	Introduction	33
3.3	Results	35
3.4	Discussion	50
3.5	Materials and methods	53
4	Identification of leaf growth-related tonoplast protein genes	57
4.1	Summary	57
4.2	Introduction	57
4.3	Results	58
4.4	Discussion	69
4.5	Materials and methods	72
5	General discussion and outlook	79
5.1	Overview with summary	79
5.2	QuantPrime – a tool for improved RT-qPCR platform design	79
5.3	Plant growth phenotyping	81
5.4	Identification of growth-related tonoplast protein genes in <i>A. thaliana</i>	84
6	References	87
	Supplementary material	97
	Allgemeinverständliche Zusammenfassung	147
	List of publications	149
	<i>Curriculum vitae</i>	151

List of figures

Figure	Title	Page
1.1	A schematic representation of the plant cell.	1
2.1	'Primer finding' in QuantPrime.	21
2.2	'Results' in QuantPrime.	22
2.3	'Primer pair details' in QuantPrime.	23
2.4	Overall work flow of primer pair design and specificity testing.	24
2.5	Work flow overview of the primer pair design algorithm.	25
2.6	Work flow overview of the primer pair specificity testing algorithm.	26
3.1	The imaging chamber.	36
3.2	Example of a plant image before and after processing.	37
3.3	Plot of CV of technical replicate measurements against mean rosette area.	39
3.4	Screenshots of the annotation tool.	40
3.5	Smoothed fits for three phenotypic parameters using four different predictor variables.	41
3.6	Linear mixed-effects model fit for rosette area and RGR of <i>sex4-3</i> .	44
3.7	Development times box plots.	46
3.8	Development stage plots for phenotypes.	48
3.9	Linear mixed-effects model fit for rosette area and RGR of <i>grf9</i> .	50
4.1	Schematic of leaf samples used for expression analysis.	61
4.2	Area and RGR modeling for <i>nhx4</i> mutants.	66
4.3	Area and RGR modeling for <i>exp3</i> , <i>exp6</i> , <i>grf9</i> and <i>vha-e3</i> mutants.	66
4.4	Photos of representative mutant plants at 14 DASE.	67
4.5	Plant development comparison for <i>nhx4</i> mutants.	68

List of tables

Table	Title	Page
2.1	Examples of transcriptome annotations available on the public QuantPrime server	20
2.2	Results of <i>in silico</i> benchmarking of QuantPrime	27
2.3	Experimental results of primer pairs designed with QuantPrime	28
3.1	Comparison of phenotyping capabilities and plant densities for different tray types.	37
3.2	Comparison of observed phenotype values for <i>sex4-3</i> and WT between five experiments.	45
3.3	Development times for <i>sex4-3</i> and WT.	47
3.4	Rosette areas of <i>sex4-3</i> and WT at different growth stages.	48
4.1	A summary of the RT-qPCR platform.	60
4.2	Scoring template for growth-association of genes.	61
4.3	The genes with the highest scores for growth-association.	62
4.4	Knockout lines for which homozygous progeny was obtained and which were screened with growth genotyping.	64
4.5	Growth effects for genotypes showing significant effects when modeling area over time.	65
4.6	Significant plant development effects for genotypes showing significant effects – early developmental stages.	67
4.7	Significant plant development effects for genotypes showing significant effects – late developmental stages.	68

List of abbreviations

Abbreviation	Full name
ANOVA	analysis of variance
API	application programming interface
BBCH	Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und Chemische Industrie
BLAST	basic local alignment search tool
CaMV35S	cauliflower mosaic virus 35S protein
cDNA	complementary deoxyribonucleic acid
C _q	quantification cycle
CSV	comma-separated values
CV	coefficient of variation
DAG	days after germination
DAS	days after sowing
DASE	days after seedling establishment
E	amplification efficiency
ER	endoplasmatic reticulum
EST	expressed sequence tag
EXP	expansin
gDNA	genomic deoxyribonucleic acid
GRF	growth regulating factor
H ⁺ -PPase	proton-translocating pyrophosphatase
HSV	hue, saturation and value
LOESS	locally weighted scatterplot smoothing
LV	lytic vacuole
MB	megabytes
MIQE	minimum information for publication of quantitative real-time PCR experiments
mRNA	messenger ribonucleic acid
NA	not available
<i>n</i>	number of individuals
NHX	sodium-proton-exchanger
<i>p</i>	probability value (p-value)

List of abbreviations

PAR	photosynthetically active radiation
PCR	polymerase chain reaction
PO	plant ontology
PSV	protein storage vacuole
qPCR	quantitative real time polymerase chain reaction
RDML	real-time PCR data markup language
REST	representational state transfer
RGB	red, green and blue
RT	reverse transcription
RT-qPCR	reverse transcription quantitative real time polymerase chain reaction
SD	standard deviation
SE	standard error
SOAP	simple object access protocol
SUBA	the Arabidopsis subcellular database
TAIR	the Arabidopsis information resource
TIP	tonoplast intrinsic protein
T _m	melting temperature
UTR	untranslated region
V-ATPase	vacuolar-type adenosine triphosphatase
VHA	vacuolar H ⁺ -ATPase
WT	wild type

Summary

In a very simplified view, the plant leaf growth can be reduced to two processes, cell division and cell expansion, accompanied by expansion of their surrounding cell walls. The vacuole, as being the largest compartment of the plant cell, plays a major role in controlling the water balance of the plant. This is achieved by regulating the osmotic pressure, through import and export of solutes over the vacuolar membrane (the tonoplast) and by controlling the water channels, the aquaporins. Together with the control of cell wall relaxation, vacuolar osmotic pressure regulation is thought to play an important role in cell expansion, directly by providing cell volume and indirectly by providing ion and pH homeostasis for the cytoplasm.

In this thesis the role of tonoplast protein coding genes in cell expansion in the model plant *Arabidopsis thaliana* is studied and genes which play a putative role in growth are identified. Since there is, to date, no clearly identified protein localization signal for the tonoplast, there is no possibility to perform genome-wide prediction of proteins localized to this compartment. Thus, a series of recent proteomic studies of the tonoplast were used to compile a list of cross-membrane tonoplast protein coding genes (117 genes), and other growth-related genes from notably the growth regulating factor (GRF) and expansin families were included (26 genes). For these genes a platform for high-throughput reverse transcription quantitative real time polymerase chain reaction (RT-qPCR) was developed by selecting specific primer pairs. To this end, a software tool (called QuantPrime, see <http://www.quantprime.de>) was developed that automatically designs such primers and tests their specificity *in silico* against whole transcriptomes and genomes, to avoid cross-hybridizations causing unspecific amplification.

The RT-qPCR platform was used in an expression study in order to identify candidate growth related genes. Here, a growth-associative spatio-temporal leaf sampling strategy was used, targeting growing regions at high expansion developmental stages and comparing them to samples taken from non-expanding regions or stages of low expansion. Candidate growth related genes were identified after applying a template-based scoring analysis on the expression data, ranking the genes according to their association with leaf expansion.

To analyze the functional involvement of these genes in leaf growth on a macroscopic scale, knockout mutants of the candidate growth related genes were

Summary

screened for growth phenotypes. To this end, a system for non-invasive automated leaf growth phenotyping was established, based on a commercially available image capture and analysis system. A software package was developed for detailed developmental stage annotation of the images captured with the system, and an analysis pipeline was constructed for automated data pre-processing and statistical testing, including modeling and graph generation, for various growth-related phenotypes. Using this system, 24 knockout mutant lines were analyzed, and significant growth phenotypes were found for five different genes.

1. General introduction

1.1 Introduction to plant vacuoles

Plant vacuoles (see **Figure 1.1**) are multifunctional organelles that occupy most of the cell volume. They were discovered during the early uses of microscopy and named 'vacuoles' since they were incorrectly believed to be empty cell space, devoid of cytoplasmic matter (De, 2000). Although this cell compartment is not directly involved in most parts of the central metabolism and housekeeping of the cell, indirectly it plays a major role in keeping the plant alive on the cellular as well as on the organ and organism level (Taiz, 1992; Marty, 1999). It has been repeatedly shown that a plant cell cannot survive without a vacuole, contrasted with the fact that there are plant cell types which can survive without other common organelles like mitochondrion and Golgi bodies, and that vacuoles quickly reappear after evacuating cells (De, 2000).

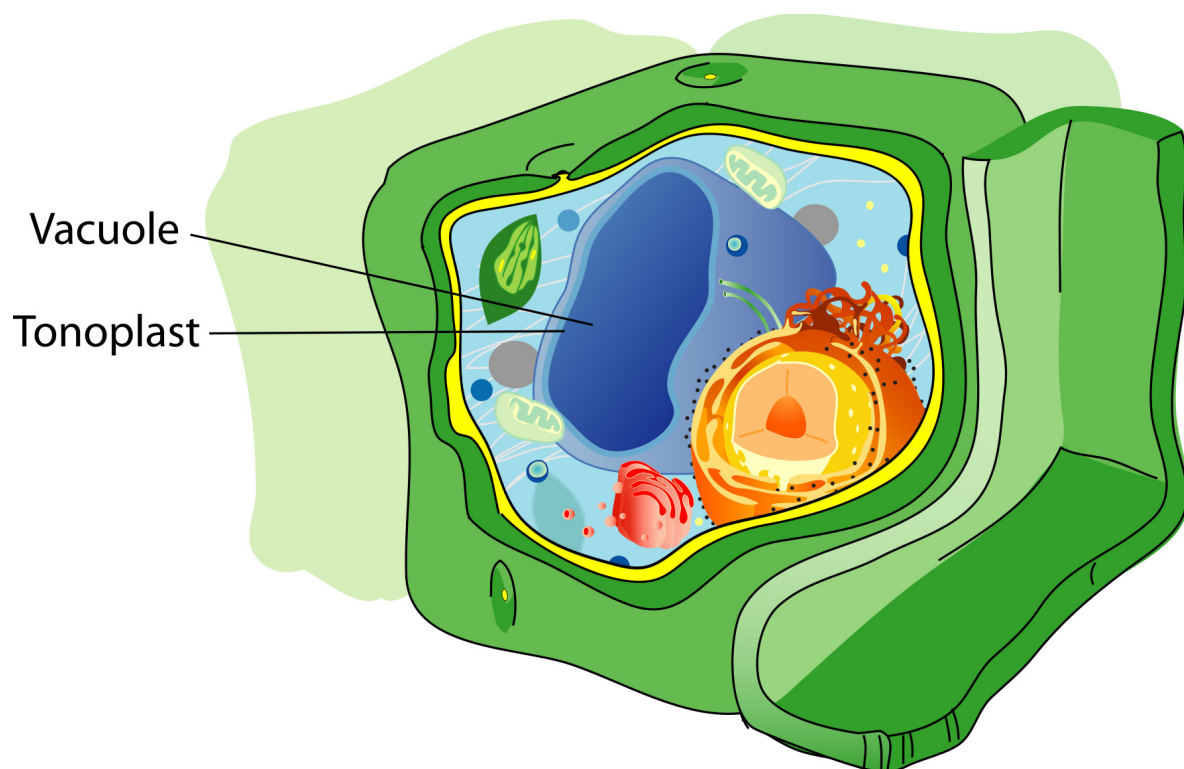


Figure 1.1: A schematic representation of the plant cell (source: *Wikipedia: 'Vacuole'*). The vacuole and the tonoplast are marked. Note that the vacuole typically occupies a larger part of the cell volume in reality than in this figure.

1. General introduction

Plant vacuoles share many properties with vacuoles in yeast and algae as well as some with lysosomes in animal cells, although these organelles are generally much smaller than the plant vacuoles. Also, the tasks assumed by animal lysosomes are in plants divided between their own lysosomes and the vacuoles (Marty, 1999; De, 2000). Plant vegetative cell vacuoles have a unique main function, which is to maintain a large cellular volume at a low energy cost, and thus maintain the stiffness of the tissue by pushing the plasma membrane against the cell wall. This is possible as the plant central vacuole generally contains high concentrations of small solutes (mostly inorganic) and thus has a high osmotic pressure. As long as the water level of the plant is high enough, water will consistently try to enter the cell and increase the volume. The cytoplasm is squeezed into a thin layer between these two main cellular membranes, resulting in a high ratio of membrane surface to cytoplasm volume. If the water level of the plant is too low, plasmolysis occurs, the cells cannot maintain turgor pressure on the cell walls and thus the rigidity of the plant is compromised, causing wilting (De, 2000).

Apart from the important role in providing structural support to the plant organs, the plant vacuole is also involved in several other physiological processes; mainly storage, waste disposal, protection and growth (Marty, 1999; De, 2000). In order to fulfill these tasks the solution contained by the vacuole is markedly different from the cytosol. Besides the higher concentration of small solutes that was already mentioned, it also often contains high concentrations of pigments, tannins and carbohydrates. These compounds impart much of the taste, odor and in some cases toxic properties of the plant leaves as well as distinct colors to the flower petals. This way the plant can e.g. attract insects for pollination, through bright pigmented flowers or an aromatic scent, or dissuade herbivores from eating its leaves by releasing bitter or toxic compounds when chewed. Many such compounds synthesized by the plants for protection or attraction have found uses for humans, some examples being natural rubber, garlic flavoring, opium and flower pigments for coloring textiles (De, 2000).

Vacuoles are also involved in the protection of the plant to abiotic stresses. One major way for a plant to counter phytotoxicity of many natural and synthetic chemicals (xenobiotics) is to covalently link them to the endogenous tripeptide glutathione and then export the complex from the cytoplasm into the vacuole (Coleman *et al.*, 1997). The same pathway is also used for sequestering secondary

1. General introduction

metabolites (e.g. pigments) into the vacuole. Also, since the vacuole serves as a reservoir of many mineral and organic ions, it can support ion and pH homeostasis of the metabolically active parts of the cell even under strong water stress (drought or high soil salinity), temporary nutrient deficiency or metabolism shifts; day/night shifts, sink to source shifts and senescence progression (De, 2000).

1.2 Diversity of plant vacuoles and their different roles

The generally accepted model states that that vacuole biogenesis occurs through fusion of smaller vesicles originating from the endoplasmic reticulum (ER) and the Golgi apparatus (De, 2000). However, several different types of vacuoles exist and they can be very diverse in form, size, content and display different functional dynamics. The following classification into two major types of vacuoles is commonly made:

1.2.1 *The central lytic vacuole (LV)*

This type of plant vacuole shares several features with the yeast vacuole and the animal lysosome, has an acidic pH (typically 5.0-5.5) and contains hydrolytic enzymes (similar to the lysosomal enzymes in animal cells) that are active in protein degradation. This type of vacuole is forming the large central vacuole in most plant cells (occupying more than 90% of the total cell volume) and fulfills most of the physiological roles described in the previous section (De, 2000).

The tonoplast for this type of vacuole contains the specific aquaporin γ -TIP (tonoplast intrinsic protein, the group of the most abundant tonoplast proteins) – in *Arabidopsis thaliana* these are the members of the TIP1 subfamily (Frigerio *et al.*, 2008; Wudick *et al.*, 2009), and antibodies raised against TIPs from this group, or fluorescent protein fusions with the transmembrane parts of γ -TIPs are commonly used to identify such vacuoles.

1.2.2 *The protein storage vacuole (PSV)*

This type of vacuole is markedly different from the LV by its lack of proteolytic activity. Thus it can function as a protein reservoir in specialized cells in seeds, fruits and some root cells, as well as certain cells in vegetative tissue (Frigerio *et al.*, 2008). The PSVs are more diverse than the LVs, and more difficult to specifically mark using antibodies raised against TIP isoforms. However, it is generally accepted that the

1. General introduction

PSV tonoplast in seeds contains the aquaporin α -TIP (in Arabidopsis the TIP3 subfamily) and that vegetative protein storage vacuole tonoplast contains δ -TIP (the TIP2 subfamily) (Frigerio *et al.*, 2008). Also, proteins accumulate in ER-derived vesicles in the endosperm of cereal grains. These organelles are vacuole-like in size but are not derived from a vacuolar compartment.

Both main types of vacuole can assume functions in storage and degradation at different stages of their life cycle. It has also been observed that two or more vacuoles of different types can co-exist in the same cell. As many observations of this type were reported, it was proposed that this would be the case for most plant cells. In this model, the course of cell development would lead to the fusion of two smaller parental vacuoles, forming the large central lytic vacuole in a mature cell. The fused vacuole would then fulfill the role of both parental compartments, which earlier in development served two different purposes; storage and degradation. However, recent findings have raised doubts on this generalized theory and it is currently a matter of debate (Frigerio *et al.*, 2008; Wudick *et al.*, 2009).

1.3 The tonoplast and its constituents

The tonoplast is a lipid bilayer that separates the vacuolar lumen from the cytoplasmic space (see **Figure 1.1**). Although it has been shown that considerable amounts of tonoplast originates from the ER and passes through the Golgi apparatus (Herman *et al.*, 1994), it does not generally share many common features with those or other endomembranes (De, 2000). Remarkably, the tonoplast has the ability to grow very rapidly, and although the exact origin of the lipids necessary for such fast growth is unknown, they are assumed to originate from the smooth ER. However, it shares some common features with the plasma membrane, notably the high permeability of water facilitated by the abundant aquaporins and the existence of a highly active ATPase, which maintains the acidic pH in the vacuolar lumen and extracellular space, respectively.

The main enzyme constituents of the main central vacuole tonoplast in Arabidopsis are the previously mentioned aquaporins (TIPs), vacuolar-type ATPase (V-ATPase) and the proton-translocating pyrophosphatase (H^+ -PPase). Apart from these main enzymes there is a series of transporters supporting cytoplasmic ion and sugar homeostasis (Jaquinod *et al.*, 2007).

1. General introduction

Although sorting signals are known for protein localization to the nuclear membrane, the endoplasmic reticulum membrane, the Golgi apparatus, the mitochondrion and the chloroplast membranes, no protein localization signal for the tonoplast is known. Thus, protein localization predictions to this compartment are limited to homology comparisons with experimentally proven tonoplast localized proteins in plants or other organisms. The lipid composition of this membrane is only slightly different from that of the plasma membrane, against which the tonoplast is typically pressed. Thus it is not straightforward to distinguish the correct localization of a protein between these two compartments, especially when the protein studied is not evenly distributed over the membrane. Due to these difficulties, it is hard to verify which transporters and other membrane associated proteins are present in the tonoplast. However, due to the rising interest in the plant vacuole, the protein content of the vacuole, including the tonoplast, has recently been object of four proteomic studies on Arabidopsis (Szponarski *et al.*, 2004; Shimaoka *et al.*, 2004; C. Carter *et al.*, 2004; Jaquinod *et al.*, 2007) which has contributed a great deal in confirming earlier predictions and providing information about previously unknown tonoplast-localized proteins. In **Chapter 4** these findings are summarized and treated in more detail.

A few online resources for Arabidopsis are also helpful in providing summaries of published proteomics and other experimental localizations of tonoplast protein, notably the Arabidopsis Subcellular Database (SUBA) (Heazlewood *et al.*, 2007) and The Arabidopsis Information Resource (TAIR) at the Carnegie Institute of Science's Department of Plant Biology at Stanford University (Swarbreck *et al.*, 2008).

1.4 The role of the vacuole and the tonoplast in growth

As described above, the vacuole maintains the plant rigidity by providing cellular volume. Thus, by controlling the import and export of ions, the total cellular volume can be regulated. Guided by ectopical relaxation of the cell wall, cell expansion proceeds in a controlled manner (Hamant and Traas, 2010). When the cell reaches mature size, the cell wall is not relaxed anymore and the turgor pressure only acts to maintain the rigidity of the structure, not to promote growth. Thus, by studying the activity of the tonoplast proteins, it should be possible to identify key regulators of growth.

It has been shown that vacuolization coincides with rapid expansion of root tip cells (Brumfield, 1942; Beemster *et al.*, 2003). There are also reports on major tonoplast

proteins being associated with cell expansion, e.g. expression of TIP1;1 was shown to be associated with cell elongation in *Arabidopsis* (Ludevid *et al.*, 1992). Based on a study carried out with reduction of TIP1;1 transcripts in *Arabidopsis* (Ma *et al.*, 2004), it was initially believed that functional TIP1;1 would be essential for plant survival. However, later studies on knockout mutants for TIP1;1 and 1;2 refuted this statement (Schüssler *et al.*, 2008), although it could be recently shown that root growth is impaired in the *tip1;1* genotype under water limiting conditions (Beebo *et al.*, 2009).

1.5 Leaf cell elongation and leaf organ size control

It is currently far from certain how the final leaf size in a plant is determined. Local regulation of cell division rates and organism-level regulation mechanisms of cell division and expansion have been thoroughly studied and discussed in the last decade (Tsukaya, 2002, 2008; Beemster *et al.*, 2003; Harashima and Schnittger, 2010). The observation of compensation has been spurring this discussion, which is the possibility for a leaf to compensate with an increased cell size when cell numbers are decreased, e.g. through disruption of cell cycle related genes, to maintain a similar final organ size (Tsukaya, 2002). This phenomenon has been extensively studied (Tsukaya, 2008), and conclusions about the directionality of this compensation effect could be made. A reduction in cell numbers often result in an increase in cell size, however the inverse situation is not true (an increase in cell number only seldom results in smaller cells) and the molecular details of this control has not been solved (Tsukaya, 2008).

However, studies of core cell cycle gene activity and the spatial distribution of cell sizes and expansion rates in *Arabidopsis* rosette leaves have given clear answers as to where and when cell division and expansion occurs (Donnelly *et al.*, 1999; Beemster *et al.*, 2005).

A time line for cell division and expansion for the primary rosette leaves in *Arabidopsis* development was established by kinetic analysis of cell sizes and numbers (Beemster *et al.*, 2005). In this timeline, the phases where cell division and expansion are predominant could be determined, although the authors also show that lower rates of cell division still occur when expansion is predominant. The spatial distribution of cell division and expansion zones evolves with time (Donnelly *et al.*, 1999). In small leaves cell division takes place over the entire leaf blade, however

with time the cell cycling activity can be described with a basipetal gradient and the cell division zone retracts towards the base of the leaf, while cell expansion still goes on across the complete leaf blade. When the leaf has reached a certain size, cell division almost completely stops, and then cell expansion also takes on a basipetal gradient and eventually retracts towards the base as the leaf reach maturity (Donnelly *et al.*, 1999). These studies have focused on epidermal cells, but it is suggested that they can be generalized to the whole leaf blade for studies where differentiation of cell type is impossible, such as nucleic acid or metabolite extraction from whole leaf samples (Beemster *et al.*, 2005).

In recent tempo-spatial studies of leaf growth in *Arabidopsis* (Wiese *et al.*, 2007), the findings of a basipetal expansion gradient could be shown on a macroscopic level for mid-sized leaves (between 30 % and 50 % of their final sizes). Also, the temporal pattern suggests that the main growth period of the leaf is the end of the night and the first few hours of light, with the lowest growth reported in the beginning of the night.

Also, it has been demonstrated that rising ploidy levels (due to endoreduplication) in cells across the growth gradient in leaves is associated with cell expansion (Beemster *et al.*, 2005), although cell ploidy level is not directly correlated with its size nor the speed of expansion, and above a certain threshold organism polyploidy results in smaller leaf sizes (Tsukaya, 2008).

1.6 Expression analysis with real time RT-qPCR

The quantitative real time polymerase chain reaction (qPCR) is a method for DNA quantification based on the ubiquitous PCR technology (Saiki *et al.*, 1985; Mullis *et al.*, 1986). One of its major uses is in gene expression analysis, for which it is then commonly confusingly referred to as qRT-PCR (in which it is unclear whether RT then is short for 'reverse transcription' or 'real time', and is used for both meanings); the less ambiguous RT-qPCR notation, for reverse transcription quantitative PCR, is preferred (Nolan *et al.*, 2006). In RT-qPCR, reverse transcription (RT) is necessary to convert the messenger RNA (mRNA) species into complementary DNA (cDNA), which can then be used as a template for qPCR.

The advantage of RT-qPCR when comparing it with other expression analysis techniques, such as Northern blotting or microarrays, is the wide dynamic range (it is generally considered to give a linear response over 7 orders of magnitude), the low

1. General introduction

detection limit (10-100 molecules) and thus low template consumption and high precision (results are technically very reproducible) (Bustin, 2000; Bustin *et al.*, 2005). However, as it is a technique based on enzymatic amplification, accuracy can be a problem if proper controls are not made, since slight differences in concentrations or enzyme activity will cause exponentially growing errors. Also, the time and complexity of setting up the PCR reactions rise linearly with the number of targets to quantify (one or few measurements are done in each reaction vessel), whereas with microarray techniques a thousands of sequences are quantified in one hybridization. Additionally, alternate transcript splice variants can only be detected if one is specifically designing primers for this purpose, while total mRNA size differences can be observed in Northern blotting.

The underlying assumption of qPCR is that by observing the increasing quantity of DNA during PCR cycling (in 'real time') one can determine the initial amount of the template DNA, or at least compare the relative quantity between two samples (Higuchi *et al.*, 1993). The major difference to traditional RT-PCR quantification, called 'end point' semi-quantitative PCR (where the end product is usually separated, stained, and quantified on an agarose gel), is the ability to track the amplification over many cycles, and differences in amplification efficiencies or the presence of inhibitors are easier to detect and correct for.

In order to quantify the increasing amount of DNA in the reaction mixture, there are two main types of qPCR in use today, both of which can be used in most brands of qPCR machines, which are essentially PCR thermocyclers with integrated fluorometers.

1.6.1 Intercalating dye qPCR

This was the first qPCR type described and uses a DNA intercalating dye for quantification. In this method normal PCR primer pairs flanking the sequence targeted for amplification are used, and as the double-stranded DNA concentration rises during cycling more and more dye is bound and increases the fluorescence of the reaction mixture.

Initially ethidium bromide, which is commonly used for staining DNA and RNA electrophoresis gels, was used but later replaced by SYBR Green I (N',N'-dimethyl-N-[4-[(E)-(3-methyl-1,3-benzothiazol-2-ylidene)methyl]-1-phenylquinolin-1-ium-2-yl]-N-propylpropane-1,3-diamine). SYBR Green I has better properties for qPCR, due to its

1. General introduction

lower background fluorescence when free in solution or bound to single-stranded DNA such as the oligonucleotide primers in the reaction mixture (C T Wittwer *et al.*, 1997). It also shows better resistance against degradation at higher temperatures, and lower inhibition of the DNA polymerase, two other properties crucial for high efficiency in PCR. Recently, new dyes have been presented as alternatives SYBR Green I, some of which show better properties, mainly in fluorescence performance and lower inhibition of PCR; most notably EvaGreen (Sang and Ren, 2006).

1.6.2 qPCR using oligonucleotides linked to fluorescent dyes

There are several different methods using different ways to covalently link fluorescent reporter dyes to primers and oligonucleotide probes which is being used in qPCR and it is outside of the scope to describe them all here. The most common in gene expression analysis, however, is the hydrolysis probe technique (a.k.a. TaqMan™ from Applied Biosystems) using normal PCR primer pairs targeting the template sequence and an oligonucleotide probe, covalently linked to a fluorescent reporter dye and a quencher, prohibiting the dye from fluorescing as long as it is in proximity. The oligonucleotide probe is designed to bind in the middle of the amplicon, between the priming sites, and when the DNA polymerase reaches the probe in the extension step of the PCR protocol, its 5'-exonuclease activity will cleave the bond between the fluorescent dye and the probe. Once released into solution, the reporter dye is no longer quenched and fluoresces. Thus, the fluorescence in the reaction mixture is relative to the number of extended DNA molecules.

When comparing the two types, intercalating dye qPCR has the advantage of simplicity, since standard PCR primers can be used for amplification. Also, at the end of the qPCR run, the accumulated amplicons can be studied by melting curve analysis, by gradually increasing the temperature and recording the fluorescence. Thus, it is possible to verify that one specific amplicon was produced, which is not possible in TaqMan qPCR. However, the major disadvantage is that all DNA, also single stranded species such as the primers, will be bound by the intercalating dye. This causes higher background fluorescence as well as problems if primer-dimers or other non-specific products are formed during PCR cycling. For the same reason it is also impossible to perform multiplex qPCR, i.e. quantification of several distinct template sequences within a single reaction vessel, using intercalating dyes. Using the TaqMan method, several primers and probes can be present in the same reaction

1. General introduction

mixture, using different fluorescent dyes with different emission wavelengths for each target sequence. However, TaqMan has the disadvantage of complexity of chemistry and in the design of probes, which makes it generally more expensive than intercalating dye qPCR, especially when high numbers of transcripts are targeted.

Regardless of the type of detection chemistry used, similar data are produced. Generally, the fluorescence caused by intercalation or free dye release rises over the background when sufficient signal is produced, typically after 15-25 PCR amplification cycles. Then, the signal increases exponentially until the PCR reaction becomes inhibited (typically by end products) and the signal starts to flatten out and eventually reach a plateau. In some detection chemistries for intercalating dye qPCR an independent dye, with similar chemical properties as the detection dye but which is not binding to DNA, is used as a normalization agent and can correct for thermal degradation of the dye or changes in concentration caused by water evaporation. Also, in order to correct for differing background fluorescence signals in different thermocycler wells, a background removal algorithm is applied which typically uses the mean signal from the first few cycles as a background.

For quantification, a certain fluorescence threshold is used (detection threshold), at which the theoretical cycle for each reaction is calculated. This value is called 'C_T' or 'C_P', for cycle at threshold or crossing point, respectively, depending on the instrument software. Here we will use 'C_q' (for quantification cycle) as recommended in the RDML (Real-time PCR Data Markup Language) specifications (Lefever *et al.*, 2009). Here, typically a threshold is selected which is crossed during the exponential phase in all reactions.

For absolute quantification, when determining DNA molecule count, the C_q value of each well is compared to a standard curve, where known titers of the sequence are present over the dynamic range. Usually several different dilutions of the sample are quantified to increase the certitude of the measurement.

For relative quantifications, where differences between samples are quantified, there is a plethora of different methods. The most common method is the so-called $2^{-\Delta\Delta C_q}$ method (Livak and Schmittgen, 2001), which is simple and applicable when the amplification efficiencies allow it. In this method the difference in C_q (ΔC_q) between the transcript of interest and a reference transcript is calculated for each sample. The reference transcript is assumed to be equally expressed in all samples, so typically a transcript from a 'housekeeping' gene is used, where the expression is known to be

1. General introduction

stable. Then, the ΔC_q values are compared between the samples (giving $\Delta\Delta C_q$ values) and the difference in expression (as fold change) is defined as $2^{-\Delta\Delta C_q}$. Thus, we effectively normalize by setting the differences in the C_q for the reference transcript to zero and thus assume that this operation evens out any differences in input cDNA concentration (and originally mRNA concentration) and any in-between operations (such as reverse transcription, dilutions etc.). Also, sometimes more importantly, we assume PCR amplification efficiency (defined as of the amount of molecules duplicated in one cycle) to be constant for the all quantifications with the same primer pairs. Also, since we normalize with the reference transcript C_q , we assume that this quantification is performed with the same PCR efficiency as the transcripts of interest.

To avoid bias by differing efficiencies between samples, methods have been suggested that correct for such efficiency differences when calculating relative quantities (Pfaffl, 2001). However, well-founded criticism has been raised whether the efficiency can be estimated well enough to avoid introduction and propagation of further errors, and it has been suggested that the efficiency of each individual reaction should be assessed for such a correction, e.g. through log-linear regression of the amplification curve (Ramakers *et al.*, 2003). This remains a matter of debate, and methods using efficiency corrections are mostly successfully applied to qPCR results where high numbers of technical and biological replicates have been carried out, to provide a good basis for accurate efficiency estimations. Additionally, it has been suggested to use several reference genes when normalizing RT-qPCR data (Vandesompele *et al.*, 2002), to avoid bias by fluctuations in the expression of a single reference gene. Although this method has received a high acceptance in the qPCR community, much experimental data are still published where only single reference genes are used for normalization, in some cases genes which are known to be fluctuating considerably (Bustin, 2010). Recently, a method for applying reference gene free normalization (using quantile normalization) has been reported (Mar *et al.*, 2009), which seems promising for datasets where many transcripts are studied – the authors suggest at least 50 transcripts.

As a conclusion, the recent intense activity in technical and analysis development for qPCR techniques shows that even though it can be considered to be a 'gold standard technique' for expression measurements, there is a need for maturation of the software used in data analysis. Also, there is a need to bring standard good

practices, concerning experimental design and data analysis, to the attention of the end users, and standardized procedures should be more carefully adhered to (Udvardi *et al.*, 2008; Bustin, 2010; Pfaffl, 2010). Notably, proper primer design is critical in obtaining high quality RT-qPCR results (Udvardi *et al.*, 2008). This subject is treated in detail in **Chapter 2**.

1.7 Leaf growth phenotyping

The word “phenotyping” originates from the noun phenotype, meaning observing a visible characteristic of an organism. When phenotyping a plant, this means to observe and collect visible traits for that plant, usually with the intention of relating these to a certain plant genotype; meaning the genetic make-up of that specific plant (Mahner and Kary, 1997). For example, in crop breeding research it is common to screen crosses of plant ecotypes and closely observe the progeny for a desired phenotypic trait, such as increased biomass or increased tolerance to biotic or abiotic stresses. When an observed phenotype can then be clearly linked with a specific genotype, it is then possible to specifically look for progeny stably carrying this genotypic trait. It is thus up to the phenotyping method and analysis to provide such clear links.

In vegetative leaf growth phenotyping, meaning how and to what extent biomass accumulates in this organ, it is possible to employ invasive or non-invasive methods.

1.7.1 Invasive methods

Typical methods include the determination of whole shoot or single leaf fresh and dry weight (by weighing) and leaf detachment and subsequent length or area analysis by manual measurement or scanning, all of which are relatively easy to perform with inexpensive lab equipment. However, as the studied sample number increases, leaf harvest and measurements become highly labor-intensive and difficult to perform in narrow time-windows, which is sometimes required for representative comparisons. Also, small samples (such as *Arabidopsis* seedlings) can be difficult to precisely weigh. On the other hand, combinations with other kinds of quantitative invasive analysis methods, such as gene expression, metabolite, and ion concentration profiling are possible and can contribute greatly to the biological understanding of the status of the tissue studied.

1.7.2 *Non-invasive methods*

The major benefit of non-invasive methods for growth phenotyping is their inherent characteristic that the studied object is available for repeated measures. Since biomass accumulation in plants generally follows the compound interest law, $m_1 = m_0 \times e^{rt}$ (Blackman, 1919), individual variations early in development have an impact on later development. Thus, it is possible to reduce the number of individual plants necessary in order to establish accurate relationships between a phenotype and genotype. This is especially advantageous when the variability between individual plant subjects is high.

Until recently, there were few non-invasive systems feasible for efficient leaf growth phenotyping of smaller plants such as *Arabidopsis*. However, in the last few years several platforms for image-based leaf growth phenotyping have emerged (Granier *et al.*, 2006; Wiese *et al.*, 2007; Walter *et al.*, 2007; Jansen *et al.*, 2009). These systems employ the combination of digital image capture with automated image analysis, and provide various levels of automation using robotic arms and sensors for plant identification, pot or tray handling and camera movements. In some cases automated weighing and watering is provided (Granier *et al.*, 2006). When fully controlling the imaging conditions, such as keeping stable light intensity with homogeneous distribution, fixed object distances and camera settings, it is possible to automate image analysis, which is typically done in the following steps:

1. Background separation: The leaves are separated from the rest of the image through one or a sequence of color and intensity filters, using color separation with RGB (red, green and blue color channels) or HSV (hue, saturation and value channels) coordinates, where appropriately set thresholds create a binary image mask.
2. Mask correction: Holes inside detected objects the foreground (leaf) mask are filled to correct for small image artifacts or highly reflective parts of leaves (such as trichomes). In some cases smoothing is applied to reduce edge effects.
3. Conversion into leaf area: The area in computer picture elements (pixels) is counted and converted into area units according to image scale.

1. General introduction

A different approach for more specific studies is used in the DISP system (Wiese *et al.*, 2007) where monochromatic light, at a wavelength not influencing the plant physiology, is used for imaging a single leaf day and night over several days. Subsequently, the images are analyzed for the detection of small moving objects, making it possible to identify temporal and spatial growth patterns in the leaf over a diurnal cycle. Due to the complicated setup, this method is only applicable to smaller replicate series and limited leaf size ranges, and thus limiting the throughput and applicability.

However, when the goal is to quantify the actual biomass (as fresh or dry weight) for smaller objects there are few suitable direct non-invasive methods. Therefore the different measures obtained with a non-invasive method are usually correlated with a reduced set of measures performed using an invasive method, i.e. creating a standard curve. It has been reported that Arabidopsis total rosette area correlates well with shoot fresh and dry weight (Walter *et al.*, 2007). However, such correlations should be applied with caution and exhaustive controls should be made when e.g. applying stresses which change the water status of the shoot (thus affecting the fresh weight), and/or induces accumulation of stress-related compounds (thus affecting the dry weight).

1.7.3 *Developmental stage annotation for growth phenotyping*

In order to describe the stage of development of a plant when performing phenotypic analysis and comparisons, different scales are being used. For Arabidopsis plants data are typically presented based on the time (in days) after sowing or germination (DAS and DAG, respectively).

In order to describe the exact stages for various phases of the life cycle of a plant, there are only a few scales have been developed for whole plants, instead many scales are limited to a certain tissue or organ type. Notably, the BBCH scale (for Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und Chemische Industrie; alternatively meaning BASF, Bayer, Ciba-Geigy, Hoechst for the companies involved in creating the scale) was created to apply a universal decimal code describing the developmental stage of mono- and dicotyledonous plants (Lancashire *et al.*, 1991). This scale was adopted for Arabidopsis phenotyping (Boyes *et al.*, 2001) and has been extensively used (> 260 citations to date). Also, the plant ontology (PO) for plant growth and development stages (Avraham *et al.*,

2008) was created, based partly on the BBCH scale, although this ontology is more comprehensive and therefore more complex. By using a standardized scale for developmental stages when performing growth phenotyping, it is easier to make comparisons between data from different experiments and laboratories performing them. Also, it makes data mining approaches of published data easier, since the developmental stage information is provided in a computer-decodable form.

Suitable developmental stage annotation is discussed in **Chapters 3** and **4**, together with the presentation of a novel pipeline for automated, non-invasive growth phenotyping pipeline for Arabidopsis.

1.8 Aims and structure of the thesis

This thesis aims at the identification of tonoplast protein genes relevant for growth in the model plant *Arabidopsis thaliana*. The underlying hypothesis is that the vacuole dominates the cell volume, and thus is a primary determinant of cellular expansion, which must in turn be controlled by the activity of the proteins at its membrane, the tonoplast. The main assumption is that this regulation should be visible at the expression level (at least in part), and thus detectable when comparing transcripts between growing and non-growing tissue. The genes coding for the differentially expressed transcripts could then be identified and functionally studied.

The main methods used in the work were high-throughput RT-qPCR and automated leaf growth phenotyping. In order to provide reliable and biologically relevant results using these methods, a great deal of effort was spent on developing methods for preparing experiments as well as handling and analyzing the data provided by them. These development efforts are described in detail in **Chapter 2** and **3**. The first method (**Chapter 2**) was published in BMC Bioinformatics in 2008 (Arvidsson *et al.*, 2008) and deals with the design of reliable, specific primers for high-throughput RT-qPCR. The second method (**Chapter 3**) describes a novel way to integrate automated image capture and analysis with development stage annotation of plants to gain deeper biological understanding when observing differential growth phenotypes. The overall aim of the thesis is then treated in **Chapter 4**, where the developed methods in RT-qPCR primer design and leaf growth phenotyping were applied to the case of putatively growth related tonoplast protein coding genes. In the general discussion in **Chapter 5** the main work is summarized and further collaborative applications of the developed methods are commented.

2. High-throughput RT-qPCR primer design

As briefly mentioned in the introduction, high-throughput RT-qPCR was extensively used in the work underlying this thesis. In order to efficiently design the primer pairs necessary for the tonoplast RT-qPCR platform (presented in detail in **Chapter 4**) and other locally planned RT-qPCR platforms, there was a need for a robust design pipeline. Therefore 'QuantPrime', a RT-qPCR primer design and specificity pipeline, was developed and eventually presented in a publication and provided to the general public as a free web service (<http://www.quantprime.de>). The paper was published in *BMC Bioinformatics* (Arvidsson *et al.*, 2008) and was shortly after publication considered as a 'highly accessed' paper according to the number of views in a certain time frame. In this chapter the paper is included as a part of this thesis.

2.1 Authors' contributions

Samuel Arvidsson designed and programmed QuantPrime, carried out most of the primer testing and drafted the manuscript. Mirosław Kwaśniewski designed the graphics for the user interface and contributed to the design of the program, carried out the primer tests with barley and revised the manuscript. Diego Mauricio Riaño-Pachón helped out to design the program, prepared sequence databases, installed and administrates the public server and revised the manuscript. Bernd Müller-Röber supervised the group, helped out with the design and testing of the program and helped drafting the manuscript.

Software

Open Access

QuantPrime – a flexible tool for reliable high-throughput primer design for quantitative PCR

Samuel Arvidsson^{1,2}, Mirosław Kwasniewski^{1,2,3}, Diego Mauricio Riaño-Pachón^{1,2} and Bernd Mueller-Roeber*^{1,2}

Address: ¹Max-Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany, ²Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Straße 24-25, Haus 20, 14476 Potsdam-Golm, Germany and ³Department of Genetics, University of Silesia, Jagiellonska 28, 40032, Katowice, Poland

Email: Samuel Arvidsson - arvid@uni-potsdam.de; Mirosław Kwasniewski - kwasniew@us.edu.pl; Diego Mauricio Riaño-Pachón - Riano@mpimp-golm.mpg.de; Bernd Mueller-Roeber* - bmr@uni-potsdam.de

* Corresponding author

Published: 1 November 2008

Received: 4 August 2008

BMC Bioinformatics 2008, 9:465 doi:10.1186/1471-2105-9-465

Accepted: 1 November 2008

This article is available from: <http://www.biomedcentral.com/1471-2105/9/465>

© 2008 Arvidsson et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Medium- to large-scale expression profiling using quantitative polymerase chain reaction (qPCR) assays are becoming increasingly important in genomics research. A major bottleneck in experiment preparation is the design of specific primer pairs, where researchers have to make several informed choices, often outside their area of expertise. Using currently available primer design tools, several interactive decisions have to be made, resulting in lengthy design processes with varying qualities of the assays.

Results: Here we present QuantPrime, an intuitive and user-friendly, fully automated tool for primer pair design in small- to large-scale qPCR analyses. QuantPrime can be used online through the internet <http://www.quantprime.de/> or on a local computer after download; it offers design and specificity checking with highly customizable parameters and is ready to use with many publicly available transcriptomes of important higher eukaryotic model organisms and plant crops (currently 295 species in total), while benefiting from exon-intron border and alternative splice variant information in available genome annotations. Experimental results with the model plant *Arabidopsis thaliana*, the crop *Hordeum vulgare* and the model green alga *Chlamydomonas reinhardtii* show success rates of designed primer pairs exceeding 96%.

Conclusion: QuantPrime constitutes a flexible, fully automated web application for reliable primer design for use in larger qPCR experiments, as proven by experimental data. The flexible framework is also open for simple use in other quantification applications, such as hydrolyzation probe design for qPCR and oligonucleotide probe design for quantitative *in situ* hybridization. Future suggestions made by users can be easily implemented, thus allowing QuantPrime to be developed into a broad-range platform for the design of RNA expression assays.

Background

The use of real-time quantitative PCR (qPCR) [1] in medium – (hundreds of transcripts) to large-scale (thousands of transcripts) profiling experiments is growing.

While in a large number of experiments qPCR is still mainly used to confirm results obtained by microarray-based hybridization experiments, the number of high-throughput discovery experiments is growing steadily

[2,3], especially for the quantification of transcripts of low abundance (e.g. those coding for transcription factors), due to the low detection limit of the method [4].

There are surprisingly few free software packages available to the academic research community suitable for the design of primer pairs for such high-throughput projects, for online use or download, including Osprey [5], Primique [6] and a few interfaces to Primer3 [7] such as Primer3Plus [8], AutoPrime [9], BatchPrimer3 [10]. Additionally, some databases of pre-computed primers, RTPrimerDB [11], PrimerBank [12], qPrimerDepot [13], AtRTPrimer [14] and DATFAP [15], have been established. There are numerous commercial and free software packages available for low-throughput design of primers, some of which are highly configurable and well suited for the design of primer pairs for qPCR.

However, none of the available packages combines all the important features (strict parameters for primer design, strict specificity checking and targeted design to avoid problems with contaminating genomic DNA) into a simple pipeline. Instead, with currently available computational tools, users have to either manually move information (such as identifiers, transcript sequences, primer sequences and others) between software packages or perform some steps completely on their own, such as specificity checking using an alignment package like BLAST [16]. Such manual steps make researchers lose valuable time, increase the risk of mistakes (e.g. labeling and sequence errors), and force them to take important decisions based on their personal interpretation of complex problems regarding large amounts of data (such as BLAST alignment sets), which either require expert knowledge or introduce bias into the results. With respect to the available primer pair databases, they are usually of limited scope. Often, only few species are covered (human and mouse being clearly over-represented), few transcripts of the species are represented (especially in databases based on submitted or published primer pairs), or inappropriate primer design parameters for combined analysis were used, requiring time-consuming optimization of PCR amplification conditions.

Here we developed QuantPrime, a program for design and specificity testing of primer pairs for qPCR, designed to meet the needs of the average or advanced user in low- to high-throughput transcript profiling experiments, while keeping the user interface very simple and yet providing important features missing in other available software packages and web services.

Implementation

QuantPrime includes a relational database for information storage, scripts containing the procedures to perform

primer pair design and specificity testing, scripts for sequence installation and maintenance, scripts for command line user interface used in high-throughput design, and a web interface as the main user interface for low- to medium-throughput primer design. For academic users we currently offer web access to the public QuantPrime server (available at <http://www.quantprime.de/>) or, on demand, compiled scripts for local installation. Commercial users are requested to get in contact with the authors to develop a license agreement.

The public QuantPrime server is currently set up with publicly available transcriptome and genome annotations from 295 different eukaryotic species. Table 1 gives examples of supported species with included features and references. The list can be easily extended according to user requests.

User interface

The web interface is designed for maximum simplicity and convenience for the user. Users have to register at the first time they visit the website. The registration step allows users to return at a later time to check the results of longer runs. Their gene lists and jobs are kept confidential, i.e. no information is relayed to other users. Furthermore, registration eases the even distribution of computing resources among users and it is the main mechanism to verify academic affiliation. An account with access to limited computing resources is available for testing purposes.

The work flow starts with the generation of a 'Project' that is associated with the annotation of a species and a certain quantification protocol. The quantification protocol implies certain parameters for primer design and specificity testing; four standard protocols for typical situations are provided:

1. SYBR Green-based real-time qPCR (accept splice variant hits): typical parameters for real-time qPCR are used, such as 50–150 bp amplicon length, 60°C annealing temperature and strict primer criteria for G/C content and melting temperature (T_m). The specificity testing will allow amplicons present in splice variants of the transcript (more details in the 'Work flow' section).
2. SYBR Green real-time qPCR (no splice variant hits): as 1, but no amplicons in splice variants of the transcript are allowed.
3. End-point semi-quantitative PCR (accept splice variant hits): similar to 1, except that longer amplicons are preferred (350–1500 bp) for easier in-gel quantification.

Table 1: Examples of transcriptome annotations available on the public QuantPrime server

Species	Annotated features included in QuantPrime			Annotation source	Reference
	Genomic sequences	Splice variants	Keyword search		
254 different species or crosses	No	No	Yes	TIGR plant transcript assemblies	[22]
91 different species or crosses	No	No	Yes	UniGene	[23]
<i>Arabidopsis thaliana</i>	Yes	Yes	Yes	TAIR release 7	[24]
<i>Aspergillus niger</i>	Yes	No	No*	JGI assembly v1.0	Non-published data
<i>Bos taurus</i>	Yes	No	Yes	NCBI RefSeq	[25]
<i>Chlamydomonas reinhardtii</i>	Yes	No	No*	JGI assembly v3.1	[26]
<i>Danio rerio</i>	Yes	No	Yes	NCBI RefSeq	[25]
<i>Drosophila melanogaster</i>	Yes	Yes	Yes	FlyBase release 5.4	[27]
<i>Homo sapiens</i>	Yes	No	Yes	NCBI RefSeq	[25]
<i>Homo sapiens</i>	Yes	Yes	Yes	H-Invitational Database 5.0	[28]
<i>Mus musculus</i>	Yes	No	Yes	NCBI RefSeq	[25]
<i>Oryza sativa ssp japonica</i>	Yes	Yes	Yes	TIGR release 5	[29]
<i>Ostreococcus lucimarinus</i>	Yes	No	No*	JGI assembly v2.0	Non-published data
<i>Physcomitrella patens ssp patens</i>	Yes	No	No*	JGI assembly v1.1	[30]
<i>Populus trichocarpa</i>	Yes	No	No*	JGI assembly v1.1	[31]
<i>Rattus norvegicus</i>	Yes	No	Yes	NCBI RefSeq	[25]
<i>Saccharomyces cerevisiae</i>	Yes	No	Yes	Saccharomyces Genome Database	[32]
<i>Selaginella moellendorffii</i>	Yes	No	No*	JGI assembly v1.0	Non-published data
<i>Vitis vinifera</i>	Yes	No	No	Genoscope assembly	[33]
<i>Xenopus tropicalis</i>	Yes	No	Yes	NCBI RefSeq	[25]

The latest versions of the annotations were added, and are updated regularly as updates become available.

* Protein IDs are searchable.

4. End-point semi-quantitative PCR (no splice variant hits): as 3, but no amplicons in splice variants of the transcript are allowed.

Users are allowed to change any parameter and create custom protocols; see Additional file 1 for a list of all possible parameters.

Next, users should create a list of transcript identifiers in the project for which primer pair design is planned. This list can either be entered manually (using the identifiers of the chosen annotation), or can be created from a similarity-based search using BLAST and a starting query sequence. Additionally, for certain annotations, keywords describing the gene(s) can be used in a text search for identifiers.

Once the list of identifiers is ready, users may proceed to 'Primer finding' (Figure 1), which when started will continue completely in the background; in the meantime users can continue to look at resulting primer pairs or add new transcripts to the list. Larger primer finding projects may take longer time to process, therefore users may close the web browser and return at a later time to check the status of their jobs.

Successful primer pairs are displayed in the 'Results' page (Figure 2), where users can inspect primer pairs in detail (T_m, G/C content, positions within transcript sequence etc., see example in Figure 3) and do bulk export of the primer data (in delimited plain text format) for ordering or local storage.

Users may return at a later time to access their data, as lists of transcripts and primer pairs are automatically saved into their corresponding projects. On the public server, projects are kept for at least a month after the latest update, and may then be deleted by the administrator for space limitation reasons. Thus, users are recommended to export primer data and store locally for reference purposes.

Work flow

QuantPrime employs a fully automated work flow for design and specificity testing of primer pairs, a process that does not require any intermediate intervention by the user. Once users have added the transcript identifiers to the project, selecting the 'Start' button will initiate the whole primer selection process, and the identified primer pairs will automatically be displayed in the 'Results' page when the process is completed.



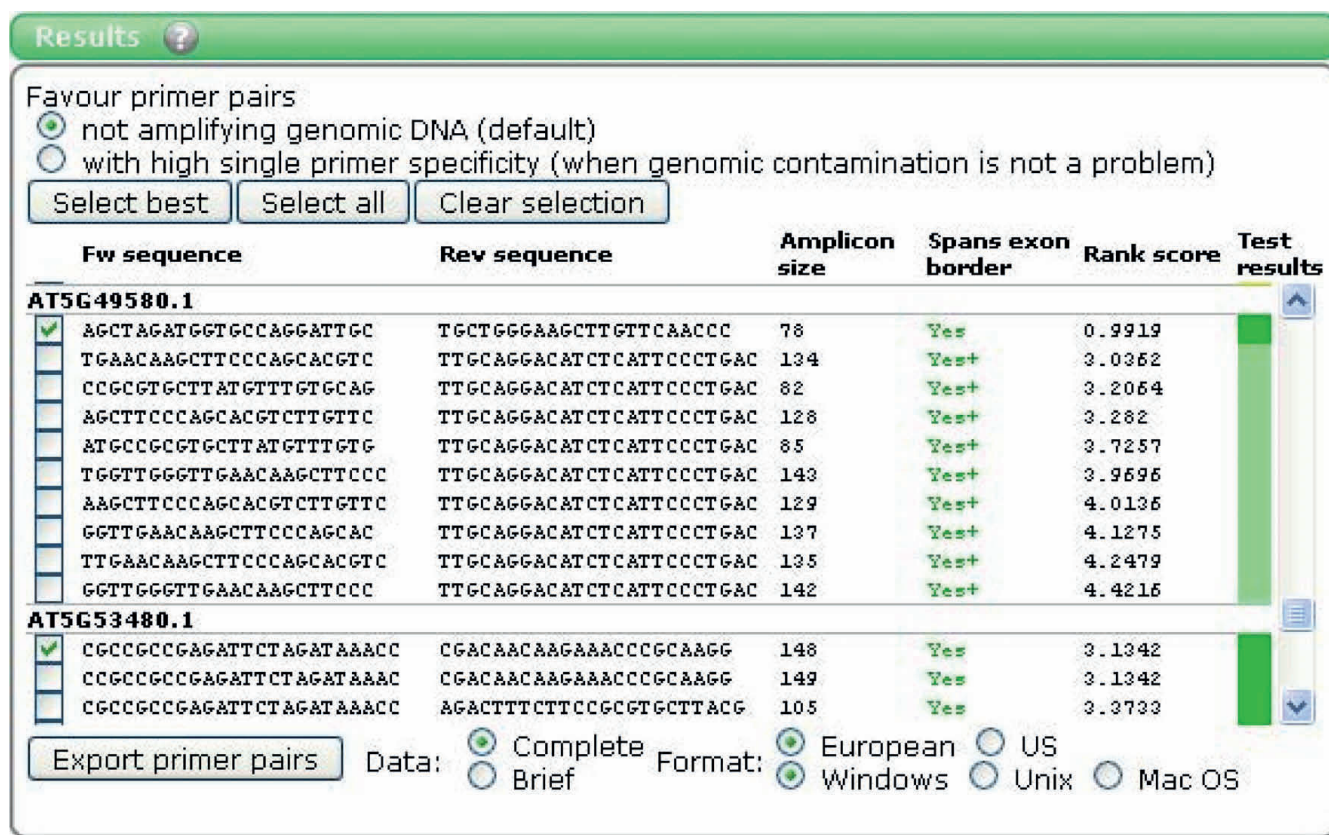
Figure 1
'Primer finding' in QuantPrime. The figure shows an example of the QuantPrime user interface for primer finding (A: up to nine transcripts, B: ten or more transcripts). The progress and success of the finding can be followed closely for small number of transcripts, for larger batches the time estimation helps users to estimate when the primer pairs will be ready.

The overall work flow of QuantPrime is sketched in Figure 4. It has two main algorithms, one for primer pair design and one for specificity testing, which are accessed by worker threads which check the output of each algorithm and decide upon the measures to be taken. The worker threads operate independent of the web server, processing submitted jobs according to defined load balancing principles (distributing computing power equally between users and projects). Due to the loosely bound system architecture it is straightforward to attach additional computing nodes to the central database allowing for high user loads. For testing purposes, a developer machine was set up to work as a computing node for the public server. With rising demand on the public server, local computing resources can be quickly mobilized to avoid long waiting times for the end users.

The primer pair design algorithm uses the Primer3 software to design primer pair candidates; a graphical representation can be found in Figure 5.

The Primer3 design parameters can be specified by the user when setting up the project; default settings are as follows:

- Primer length: 20–24 bases
- Amplicon size: 60–150 bp
- Primer melting temperatures (T_m): 64 +/- 3 °C (for optimal annealing around 60 °C) (using nearest neighbor thermodynamics [17]), maximum 2 °C T_m difference between forward and reverse primers

**Figure 2**

'Results' in QuantPrime. The figure shows an example of the 'Results' page. Primer pairs successfully identified for the examined transcripts are presented. The following information is provided: the sequences (5' to 3') of the forward and reverse primers; the amplicon size (in bp); whether at least one primer spans an exon-exon junction ('Yes' in all cases in the example shown); the rank score (as calculated by Primer3); and the color code of the specificity rank given to the primer pair (see text for details). When clicking the primer pairs, more details are shown (see Figure 3).

- Amplicon melting temperature: 75–95 °C
- G/C content: 45–55%
- Max. repetition of a nucleotide: 3
- G/C-clamp: last 3' base of each primer must be a G or a C

In addition to the Primer3 selection criteria, the primer pair candidates are filtered through the following steps:

- Extended G/C clamp options: to avoid mispriming, it is often appropriate to avoid too many G/C bases within the 3' region of the primer. This cannot be controlled by Primer3; therefore we introduced a parameter that allows the user to define a maximum number of G/C bases to be present in the last 3' bases. The default setting is maximum three G/C bases in the last five bases of a primer.

- Amplicon bias at 3' end of transcript: primers for amplicons at the 3' end of the transcript (the last 1000 bp) are favored. For the common user this is often wanted as cDNA preparations primed with oligo-d(T)_x generally exhibit 3' region bias. For those using random hexamers for cDNA synthesis, this parameter can be switched off.

- Skip 3' UTR: in cases where multiple polyadenylation signals exist in the 3' UTR it might be desirable to avoid priming in this region, as it could lead to biased quantification. This option can be switched on for custom design protocols.

- Exon-exon junction in primers: as RNA preparations may contain some genomic DNA even after digestion with DNase I, such primers can successfully distinguish between cDNA and genomic DNA. When possible (i.e., when a genomic sequence with one or more intron(s) is available), primers that span an exon-exon

Primer pair information
✕

Transcript identifier: AT1G76130.1 - ATAMY2, AMY2, AMY2/ATAMY2 (ALPHA-AMYLASE)
alpha-amylase, chr1:28566175-28568970 FORWARD

Forward primer
Sequence: ACTTACTCATCCCGGCATTCCC (22 b)
Melting temperature: 62.5 °C
G/C content: 54.5 %

Reverse primer
Sequence: TTGTCGCCTCCTAATGTCAATCAG (reverse complement: CTGATTGACATTAGGAGGCGACAA)
Melting temperature: 61.3 °C
G/C content: 45.8 %

Amplicon
Size: 106 b
Melting temperature: 81.1 °C
G/C content: 43.4 %
Optimal annealing temperature: 60.3 °C

Alignment with transcript sequence

```

GTGTGATTAGCTTCTTCGTTTCCTTGGTAATCGTTCTGTTTTTCGTTGATTACTGTTTGGAGGTTTTTCATTGAATCTTG
TCTCGACATGGGCTACTATAACAATGTCTTTGATGAATGCAACGACCAAACCTGATATCGGTCGGGTTATACGCCGATGGAA
GGGAAGTCATTCTCCAAGCATATAATTGGGAATCTCATAAATATGATTGGTGGAGAACTTGGATGGTAAAGTTCCTGAC
ATCGCGAAATCCGGCTTTACTTCTGCATGGTTGCCACCACCATCTCAGTCTCTTGCACCGGAAGGTTATCTTCCACAGGA
CCTTTATTCACTAAACTCAGCATATGGCTCTGAGCATCTATTGAAATCCTTACTGCGCAAGATGAAACAATACAAAGTTA
GAGCTATGGCTGATATAGTCATCAATCATCGTGTGGGACAACCTAGAGGACATGGTGAATGTATAACCGTTTATGATGGA
ATTTCAATTACCGTGGGATGAACACGCCGTGACTTCTTGTACCGGAGGACTGGGTAACCGAAGCACAGGGGATAATTTCAA
TGGAGTTCAAAATGTTGATCACACTCAGCATTITGTTAGGAAAAGATATAATTGGATGGCTTCGTTGGTTACGCAACACCG
TCGGGTTTCAAGATTTCCGTTTTGACTTTGCTAGGGGTTATTGAGCAAACCTATGTGAAGGAATACATTGGCGCAGCGAAA
CCATTATTCTCGGTTGGAGAATGTTGGGATCTTGCAATTACAATGGCCATGGTCTAGACTATAATCAAGATAGCCATAG
ACAGCGTATAATCAGTTGGATCGATGCCAGGGACAGATCTCTGCTGCATTTGACTTCACAACATAAAGGAATTCTGCAGG
AAGCCGTAAGGGTCAAGTATTGGCGTTTTATGTGATGCCAAGGGGAAGCCACCGGGTGAATGGGATGGTGGCCTTCAAGA
GCTGTACATTCCTTGATAAACCATGACACTGGCTCTACTCAGGCTCATTGGCCGTTCCCTTACACCACGTTATGGAGGG
CTATGCATATATACTTACTCATCCCGGCATTCCCTGTGTGTTCTACGATCACATTTTACGATTGGGGAAAGCTCAATACATG
ATCAGATTGTCAAACT*GATTGACATTGGGAGGCGACAAGATATCCACAGTAGTCAACTGTCCGTGTTTTTAAAAGCAGA
ATCTAACTTATATGCAGCCATTGTTGGTGAGAAAATATGTATGAAGCTTGGAGACGGCTCTTGGTGCCTTCTGGTAGAG
ACTGGACTCTAGCAACAAGTGGCCATCGCTATGCCGTCTGGCACAAGTAAATCTTTTGTGTCTATCTAGCTTTGAGTCCA
AAGAAAAATATATGTAATTCATTTTCTGTACCGGGTTTTACTATAATGTCATGCCCTTACGTCTTCATGACGTAAGATTT
TATAAGTTATATTTATGAATGTAATAATTTGCATCTTTCGTGTGAATGTTTTG

```

Specificity test results
Overall score: Perfect
cDNA specificity: Good
Single primer specificity: Good
Amplifies genomic DNA: No

Figure 3
Primer pair details in QuantPrime. The figure shows an example of the 'Primer pair information' page. The page provides details about the selected primers and the amplicon. Positions to which the primers anneal within the target sequence are indicated in blue or green; the amplicon is highlighted by gray shadowing. Primers shown in blue anneal to an exon, whereas primers shown in green anneal across an exon-exon junction (the position of the intron is indicated by a red arrow head). In the 'Specificity test results' section, details about the specificity of the primer pair can be seen. If specificity problems exist, more details can be found here concerning the other possible amplicons.

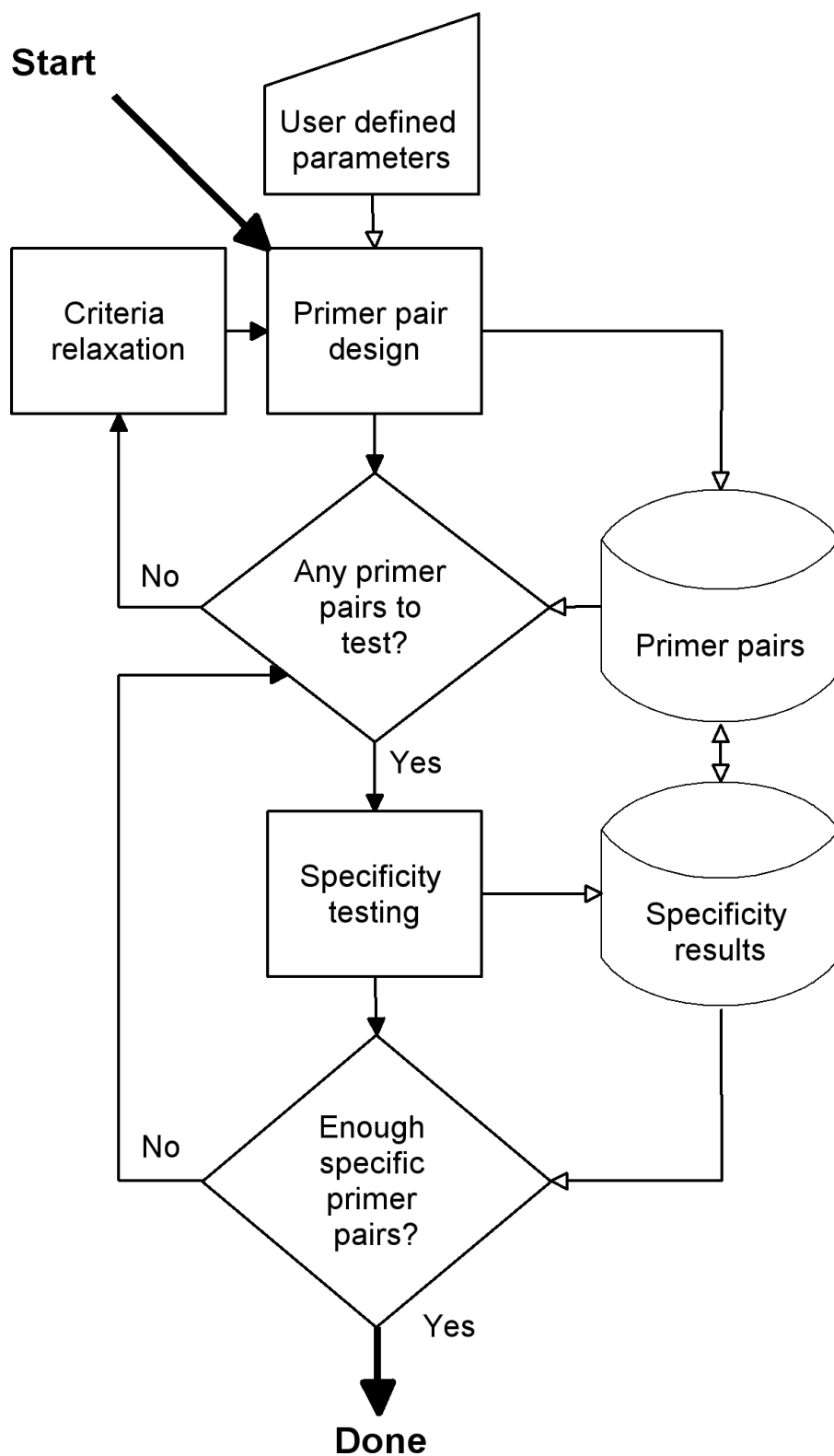


Figure 4
Overall work flow of primer pair design and specificity testing. Filled arrows symbolize logical flow while open arrows symbolize data flow.

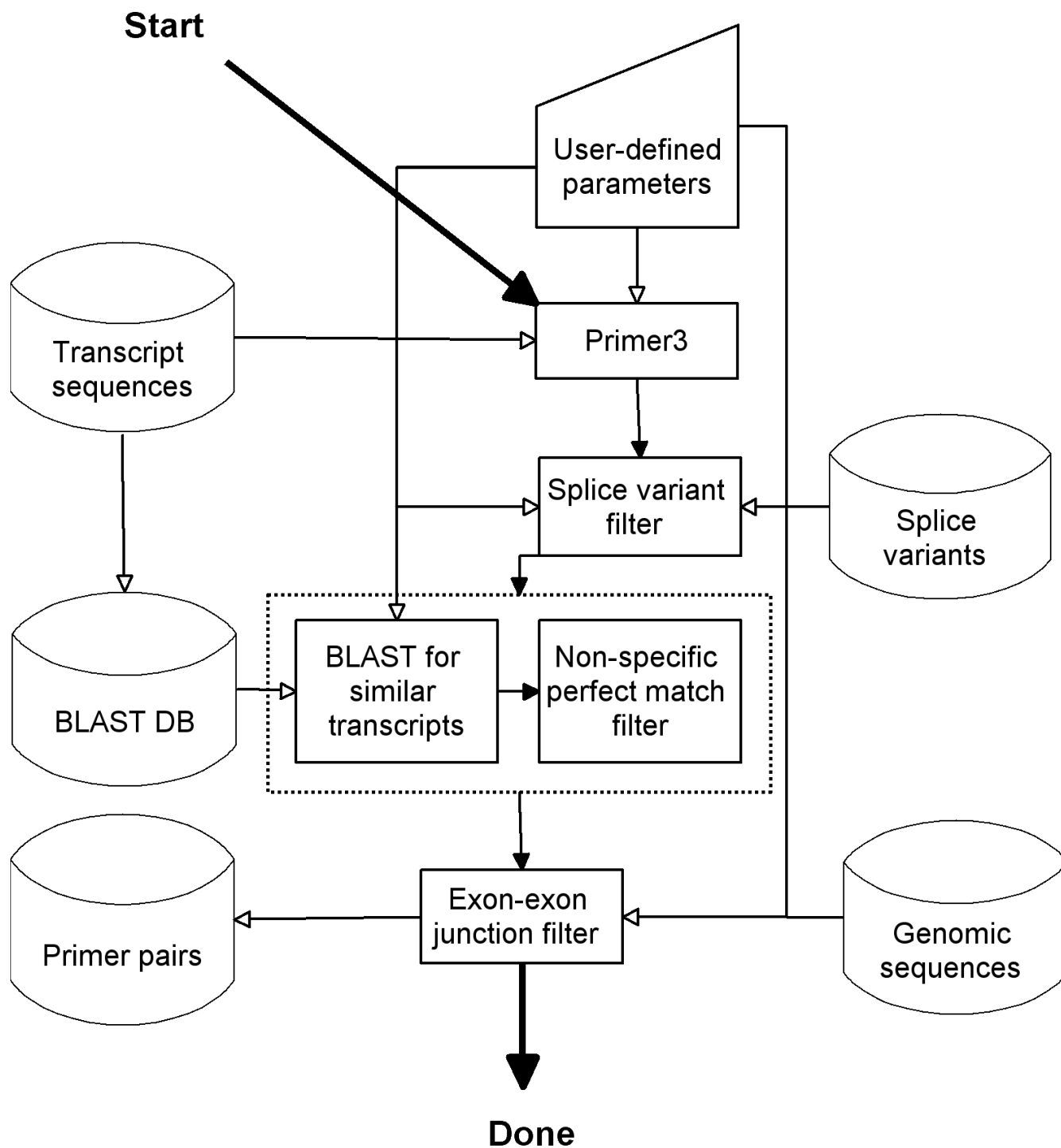


Figure 5
Work flow overview of the primer pair design algorithm.

junction are favored, especially when the junction occurs at the 3' end of the primer, to further decrease the probability of extendable annealing to genomic DNA.

● Specificity pre-filtering: in order to save workload for the specificity testing algorithm, obvious un-specific primer pairs are removed at this step. This is achieved by finding transcripts that are similar to the

target transcript using BLAST (blastn of transcript against the whole transcriptome with an e-value = 1) and filtering out the primer pair candidates annealing perfectly to any of those sequences. Three configurations of the filter are possible; one that forces the algorithm to find primer pairs amplifying all splice variants of the transcript (for annotations containing such information), one that forces it to find only those specific to a certain splice variant, and one that allows

(but does not force) them to amplify other splice variants (default setting).

The successful primer pairs are saved to the database, and the algorithm reports the number of designed primer pairs back to the calling worker thread. If it was possible to find primer pairs, the next step is specificity testing, described below (an overview is shown in Figure 6):

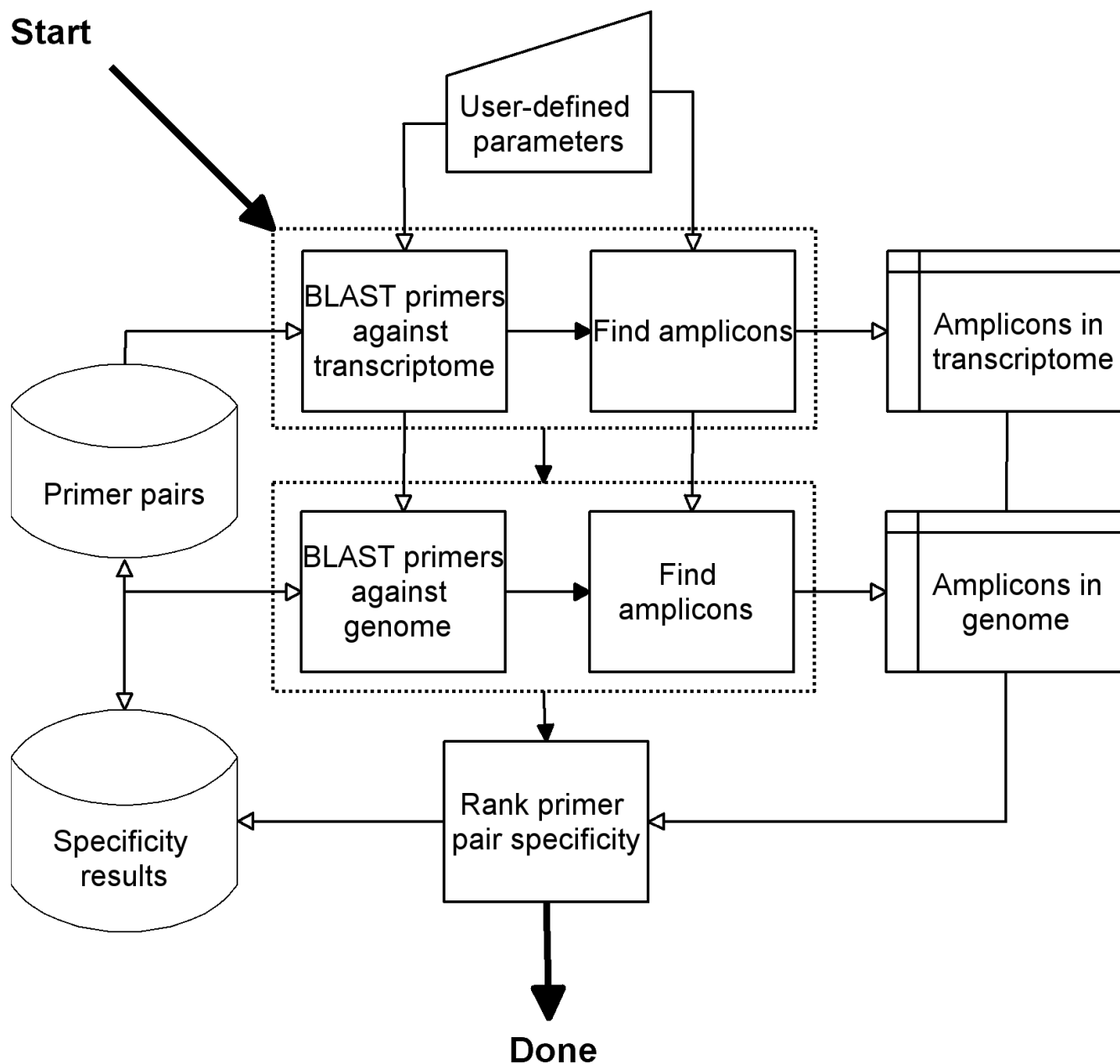


Figure 6
Work flow overview of the primer pair specificity testing algorithm.

The primer pair specificity determination algorithm is based on the interpretation of BLAST results (with default settings: e-value = 200, word size of 7), using each primer as a query towards the transcriptome and, when available, against the genome. To identify unspecific amplicons in a transcriptome or a genome, the following (configurable) criteria are applied to the BLAST hits:

- Last two bases of the 3' region of each primer must be identical to the BLAST hit.
- Amplicons of up to 1500 bp are considered for SYBR Green protocols, and 3500 bp for end-point protocols.

Even though the primer pairs cannot give rise to an unspecific amplicon, it is generally preferred that they should be as specific as possible to the target sequence. This is approximated by checking whether a single primer in the pair has a significant (the default setting is 75%) identity to another cDNA sequence, and where the last 3' base is identical (which can be configured).

The information from the above procedures is assembled and saved into the primer pair database. Based on this specificity information, QuantPrime labels the tested primer pairs with one out of four specificity ranks: bad, acceptable, good or very good. They are defined as follows:

1. Bad (shown in red in the web interface): might amplify a non-specific cDNA fragment.
2. Acceptable (yellow): amplifies only the specific sequence, but one primer has a high similarity with a non-target sequence and the primer pair might amplify genomic DNA.

3. Good (light green): amplifies only the target sequence, but one primer has a high similarity with a non-target sequence or the pair might amplify genomic DNA. This is the highest possible rank for primer pairs designed for species without a genome annotation.

4. Very good (dark green): amplifies only the target sequence, both primers are highly specific to this sequence and will not amplify genomic DNA.

The list of designed primers is worked through until enough (the default setting is 10) of at least acceptable (rank 2) primer pairs are found. The worker thread then decides whether it is possible to find higher-ranking primer pairs (e.g., when more primer pairs spanning exon-exon junctions can be designed); if so it continues until it is successful or until a certain primer pair threshold is reached (default setting is 500 primer pairs).

The work flow implemented on the web server only performs automated relaxation in amplicon 3' bias and exon-exon junction criteria; the Primer3 parameters are not relaxed. Thus, for certain transcripts, QuantPrime will fail to find specific primer pairs; with the default settings, we arrived at a failure rate of 2–9% (see Table 2). If the user wishes to relax the Primer3 parameters to be able to find specific primers for such problematic transcripts, a new project has to be created with different primer design parameters. Some users might find this procedure cumbersome, but we chose this design to prevent primer pairs with heterogeneous design parameters to be mixed within an assay. We are open for user suggestions to introduce certain configurable relaxations in future versions of QuantPrime.

Table 2: Results of *in silico* benchmarking of QuantPrime

Species	Transcripts	Total search time	Average search time	Primer pair specificity ranking ¹		
				Acceptable ²	Good ³	Very good ⁴
<i>Arabidopsis thaliana</i>	5000	20:22:06	15 s	4916 (98%)	4323 (86%)	2534 (50%)
<i>Vitis vinifera</i>	5000	50:45:33	37 s	4765 (95%)	3927 (78%)	2315 (46%)
<i>Drosophila melanogaster</i>	5000	13:48:45	9.9 s	4894 (97%)	4075 (81%)	3096 (61%)
<i>Chlamydomonas reinhardtii</i>	5000	12:11:07	8.8 s	4568 (91%)	3999 (79%)	2349 (46%)
<i>Oryza sativa ssp japonica</i>	5000	83:31:12	60 s	4658 (93%)	3821 (76%)	1984 (39%)
<i>Hordeum vulgare</i>	23078	22:56:59	3.6 s	22145 (95%)	21564 (93%)	-

Primer pairs designed for hypothetical high-throughput experiments, for random transcripts of each species. The numbers of successfully designed primer pairs for the different specificity ranks and the search times are reported for each species (percentages refer to the total number of transcripts tested).

¹ Percentages indicate for how many of the transcripts primer pairs of at least the rank given were identified. ² 'Acceptable' amplifies only the specific sequence, but one primer has a high similarity with a non-target sequence and the primer pair might amplify genomic DNA. ³ 'Good' amplifies only the target sequence, but one primer has a high similarity with a non-target sequence or the pair might amplify genomic DNA. This is the highest possible rank for primer pairs designed for species without a genome annotation. ⁴ 'Very good' amplifies only the target sequence, both primers are highly specific to this sequence and will not amplify genomic DNA.

Results

Experimental testing of primers designed through QuantPrime

To verify the experimental usefulness of the primer pairs designed with QuantPrime, we tested it to design primers for a medium-sized expression profiling experiment for *Arabidopsis thaliana* (for 128 transcripts of various genes), carried through by fellow researchers in our group. The default settings for design and specificity testing (SYBR Green protocol, splice-variant hits allowed) were used and the highest ranking primer pairs were chosen. As can be seen in Table 3, we experienced a success rate of 96%, meaning unique amplicons of predicted size and amplification efficiencies (E) = 1.8 (see Methods for details). Over 88% of the primers were predicted not to amplify genomic DNA. For five out of 128 transcripts we obtained non-satisfying results. For those, good primer pairs could be obtained by testing one or two alternative primer pairs designed by QuantPrime, without having to perform any PCR optimization (results not shown).

We also designed primer pairs for 33 transcripts (cell cycle genes) from *Chlamydomonas reinhardtii* and tested them in the same way as above. In this case transcripts of four genes could not be detected, and as the primer pairs for these transcripts spanned exon-exon junctions, we could not test them on genomic DNA. However, only one of the primer pairs of the detectable transcripts did not pass the quality control (having multiple products seen on gel), giving a success rate of 97%. Seventy-three percent of the designed primer pairs were predicted not to amplify genomic DNA.

Additionally, primer pairs for 30 different barley (*Hordeum vulgare*) transcripts were tested. For two primer pairs, no product could be detected, but only one of the detectable transcripts did not pass the quality control (low amplification efficiency), yielding a success rate of 96%. As no whole-genome sequence is available for barley, no predictions for genomic amplicons could be made.

In these three experiments, we thus observed a success rate > 96%. Examples of primer pairs and PCR amplification products separated on agarose gels can be found in Additional file 2.

To assess QuantPrime's accuracy of prediction of genomic DNA amplification, 173 primer pairs from an existing qPCR platform for tonoplast-related transcripts of *A. thaliana* (to be published elsewhere) were tested *in silico* with QuantPrime and experimentally with genomic DNA in real-time PCR. QuantPrime predicted 95 of these as 'gDNA-unsafe', while in real-time PCR measurable amplification was recorded for 88 of the primer pairs (data not shown). Twelve primer pairs (6.9%) were falsely predicted as 'gDNA-unsafe', and 19 (11%) falsely as 'gDNA-safe'.

In silico benchmarking of QuantPrime

In order to assess the success rate and speed of QuantPrime for larger expression profiling projects, hypothetical high-throughput assays were designed for six different species. Five assays consisted of respectively 5000 randomly selected transcripts from current genome annotations of five species (*Arabidopsis thaliana*, *Vitis vinifera*, *Drosophila melanogaster*, *Chlamydomonas reinhardtii* and *Oryza sativa ssp. japonica*), while the sixth assay consisted of the whole UniGene collection of barley (*Hordeum vulgare*) transcripts. As seen in Table 2, the success rates (primer pairs ranked as 'acceptable' or better by specificity testing) varied between 91 and 98%, which correlates relatively well with the status and complexity of the annotations. For the higher specificity ranks rather high variation between species was observed, ranging from 76–93% for the rank 'good', and 39–61% for the rank 'very good'. Since the barley annotation lacks genomic information, 'good' is the highest possible rank. Primer pair identification speed varied between 3.6 (barley) and 60 (rice) seconds per transcript, correlating roughly with the size of the sequence sets to be searched by BLAST.

We also did preliminary tests with data sets from larger transcriptomes/genomes (human, mouse), for which the

Table 3: Experimental results of primer pairs designed with QuantPrime

Experiment	Predicted gDNA-safe	Quality control passed ¹	Quality control passed ¹ for detectable transcripts ²
<i>A. thaliana</i>	113/128 (88.3%)	117/128 (91.4%)	117/122 (95.9%)
<i>C. reinhardtii</i>	24/33 (72.7%)	28/33 (84.8%)	28/29 (96.6%)
<i>H. vulgare</i> ³	-	27/30 (90.0%)	27/28 (96.4%)
	137/161 (85.1%)	172/191 (90.1%)	172/179 (96.1%)

The primer pairs were designed for wet-lab quantification experiments. The number of primer pairs passing strict quality control checks (melting curve analysis, agarose gel separation and efficiency check) are reported in the table.

¹ Melting curve analysis, gel analysis and efficiency check ($E \geq 1.8$) passed. ² Undetectable transcripts ($C_t > 40$) removed from the statistics. ³ TIGR Transcript Assembly annotation used, no genomic sequences available.

design speed dropped (data not shown). This is due to a higher memory demand of the BLAST runs that can be offered in the future, when requests for the service rise.

Discussion

Our experimental results show that the primer pairs designed by QuantPrime can be directly used with a high success rate (> 96%) in qPCR applications, without a need for experimental optimization of individual reaction conditions. When running tests in parallel on a standard desktop computer, the speed is enough to design primers for high-throughput projects for small- to medium sized transcriptomes as shown by the *in silico* tests.

To our knowledge, there are no other web-based tools directly comparable to QuantPrime, although programs like Osprey [5] and Primique [6] offer possibilities for batch primer pair design. In those two other applications, however, the user has to supply the database against which primer pair specificity is tested, but the upload capacity is limited to 10 MB which does not suffice for most transcriptomes. QuantPrime always tests the primer pairs against the whole transcriptome of the annotation used, and additionally offers a richer user interface, exon-exon junction design of primers to avoid genomic DNA amplification, and a high degree of customization of parameters, features not available in the other software packages. Most annotations are already included in QuantPrime; in the case that users have special annotations not available on the public server, they can contact us for adding it there, or they can run QuantPrime locally. A more exhaustive comparison of QuantPrime with other available primer design software can be found in the Additional file 3.

For some species pre-computed databases of primers exist. An example is AtRTPrimer [14] containing primer pairs for most genes of *A. thaliana*. When looking at the available primers in this resource one will find that the parameters for design, especially amplicon size, make the primer pairs unsuitable for real-time PCR, and due to the differences in T_m between different primer pairs exhaustive PCR optimization would be necessary for using them in high-throughput. The authors report a success rate of 93%, however only 21 primer pairs offered by the database were experimentally validated. In comparison, QuantPrime offers complete customization of parameters for different quantification methods, and we see higher success rates (> 96% for the three species tested here, $n = 191$). Another example is the PrimerBank [12], which covers primer pairs for human and mouse transcripts, which could be useful for high-throughput purposes (due to strict design criteria), even though amplicon sizes vary. Those two databases are limited to specific species; there are a couple of databases covering more species, notably

RTPrimerDB [11], which however cover very few non-human genes. Another database containing primer pairs for plant transcription factors is DATEFAP [15], which however is based on EST sets, which is questionable for *A. thaliana* and *O. sativa* for which good genome annotations are available. It therefore lacks information about possible genomic sequences amplified by the primer pairs; additionally T_m values vary widely between primer pairs, which might require exhaustive PCR optimization.

The parameter flexibility for design and specificity testing offered in QuantPrime makes it straightforward to employ it for the design of oligonucleotides for a number of other quantification applications, such as qPCR with hydrolyzation probes (e.g. TaqMan probes, Scorpion primers), quantitative *in situ* hybridization of mRNA and others. Such protocols will be added to QuantPrime as we gather experimental data and feedback from users.

Conclusion

The QuantPrime website offers a unique service to the scientific community, with ease-of-use, flexibility of parameters and a broad scope of transcript databases and genomic annotations, which should make it a very useful tool for primer design. No other publicly available tool offers the same services. Overall, the speed of computation and the quality of the designed primer pairs as shown experimentally make QuantPrime (on the public web server or as standalone software) a suitable system for primer design in low- to high-throughput transcription profiling projects.

We are open for suggestions from the scientific community to further develop QuantPrime in the future. Upon request we may for example include further transcript databases and genome annotations, sets of parameters for other quantification protocols and applications, or improve the applied specificity testing algorithms. Institutions wanting to host mirrors of the QuantPrime public web server or supply additional computing power are encouraged to contact the authors.

Methods

General

Standard molecular techniques were performed as described [18]. Oligonucleotides were obtained from MWG (Ebersberg, Germany). Unless otherwise indicated, other chemicals were purchased from Roche (Mannheim, Germany), Merck (Darmstadt, Germany), or Sigma (Deisenhofen, Germany).

Growth conditions

Arabidopsis thaliana (L.) Heynh accession Col-0 plants were grown in growth chambers with an 8-h day length provided by fluorescent light at $120 \mu\text{mol m}^{-2} \text{s}^{-1}$ (50%

intensity during the first and last 30 minutes of the light period) and a day/night temperature of 20/16°C and relative humidity of 60/75%. Whole, young plants (four weeks after germination) including washed roots were harvested 2 hours after lights-on, snap-frozen in liquid nitrogen and stored at -70°C until RNA extraction. *Chlamydomonas reinhardtii* CC503 cw92 mt+ was grown under continuous light (100 μmol m⁻² s⁻¹) at 21°C in HEPES-based medium as described [19]. *Hordeum vulgare* (Karat variety) plants were grown as previously described [20], and parts of roots from seven days-old seedlings were used for total RNA extraction.

RNA extraction and cDNA synthesis

After grinding of plant/algal material in liquid nitrogen, total RNA was isolated with Trizol reagent (Invitrogen, Karlsruhe, Germany) or RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturers' specifications. RNA quality was determined spectrometrically ($A_{260}/A_{280} > 1.8$) using a NanoDrop ND-1000 spectrometer (NanoDrop, Detroit, USA) and by visual inspection of separated bands on agarose gels.

After isolation, genomic DNA was digested using Turbo DNA-free recombinant DNase I (Applied Biosystems Applera, Darmstadt, Germany) following the manufacturer's specifications. The level of remaining genomic DNA contamination was measured by diluting the samples to the same concentration as the final cDNA samples (10 ng μl⁻¹) and performing real-time PCR using primers for a genomic sequence (*UBQ10*: Fw 5'-GGCCTTG-TATAATCCCTGATGAATAAG-3', Rev 5'-AAAGAGATAACAGGAACGGAAACATAGT-3'). Samples with consistent cycle threshold (Ct) values below 35 were re-treated with DNase I or new RNA extractions were performed.

Two μg of total RNA was used in 20-μl reactions for cDNA synthesis, using RevertAid R-minus cDNA synthesis kit (Fermentas, St. Leon-Rot, Germany), following the manufacturer's specifications. The cDNA was then diluted 1:10 in order to reduce the effect of RNA isolation and cDNA synthesis buffer on the subsequent PCRs.

Real-time quantitative PCR

qPCR was carried out in technical triplicates or quadruplicates using 0.5 or 1 μl of diluted cDNA in 5- or 10-μl reactions, 2 or 4 μl of 500 nM primer pairs and 2.5 or 5 μl of 2× Power SYBR Green PCR Master Mix (Applied Biosystems). The following PCR protocol was used on Applied Biosystems 7300 (96-well plates) and 7900HT (384-well plates) real-time PCR systems: 10 min at 95°C, 15 sec at 95°C, and 1 min at 60°C repeated in 50 cycles, followed by melting curve analysis. When testing primer pairs, the

PCR products were then separated on a 2% agarose gel and visualized with ethidium bromide, using 50 bp DNA ladder (Invitrogen) for size determination.

Cycle threshold (Ct) values for each reaction were calculated using Applied Biosystems SDS software, with baseline set to cycle 3–15 and threshold to 0.2 Rn, recorded from the SYBR Green I dye signal normalized against the ROX dye signal.

Real-time PCR amplification efficiencies were calculated using the LinRegPCR tool [21], using the best-fit method for 4 to 6 points. This tool uses linear regression on log-values of normalized fluorescence data from individual reactions to calculate E in the equation for PCR kinetics, $N_C = N_0 * E^C$, which states that the amount of product after C cycles (N_C) is equal to the starting concentration (N_0) times the efficiency (E) to the power C; 100% efficiency would give an efficiency value of 2.

Efficiency values from fitted curves with R-squared values below 0.999 were considered as unreliable; Ct values and efficiencies from such reactions were removed from further calculations. Medians of Ct values and efficiencies were calculated and used in further calculations.

Public server setup

The web-based QuantPrime program runs on a Linux-based server, with two Intel 1.6 GHz QuadCore 64-bit processors and 4 GB of RAM, configured to run up to six design/testing threads in parallel, always leaving two virtual processors available for database and web handling. This was found to be the most efficient configuration for this single server; setting up the program and database in a clustered environment with specialized data and computation nodes should lead to synergistic speed improvements, as the amount of data transferred between database and executing threads are kept very low.

In silico benchmarking

For the random selection of transcripts from annotations, the built-in random function in MySQL was used to order all transcripts from the respective annotation having a transcript length of more than 300 bp, of which the top 5000 were selected.

The run times given are real time (not CPU time), meaning the difference of the time point when the experiment started and when it finished. The average time per transcript is the total time divided by the number of transcripts. Due to the parallel nature of the program, the typical time to design one specific primer pair for a transcript is longer.

Availability and requirements

Project name: QuantPrime

Project home page: <http://www.quantprime.de/>

Operating systems: Platform independent

Programming languages: Python and PHP (web interface)

Other requirements: Web browser (supporting JavaScript) for using the public server; for standalone use: BioPython 1.4 or higher, MySQL 5.0 or higher, Primer3 1.1.1 or higher, NCBI BLAST 2.2.13 or higher

Any restrictions to use by non-academics: License needed

Authors' contributions

SA designed and programmed QuantPrime, carried out most of the primer testing and drafted the manuscript. MK designed the graphical user interface and contributed to the design of the program, carried out the tests with barley and revised the manuscript. DMRP helped out to design the program, prepared sequence databases, installed and administrates the public server and revised the manuscript. BMR supervised the group, helped out with the design and testing of the program and helped drafting the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

List of customizable parameters in QuantPrime. A comprehensive list of all parameters that can be customized in QuantPrime, with parameter ranges and default values.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-465-S1.xls>]

Additional file 2

Examples of primer pairs with gel images. Examples of primer pairs for different species with images of agarose gel separations of their PCR amplification products.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-465-S2.pdf>]

Additional file 3

Comparison of QuantPrime with other primer design software. A comparison table including QuantPrime and other commonly used primer design software.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-465-S3.xls>]

Acknowledgements

SA is supported through the EU Marie Curie Research Training Network 'VaTEP – Vacuolar Transport Equipment for Growth Regulation of Plants' (MRTN-CT-2006-035833) which the authors greatly acknowledge. MK thanks the DAAD for a fellowship provided through the program 'Modern Applications of Biotechnology' (No. A/06/04209) and the Polish Ministry of Science and Higher Education for financial support (research grant 2 P04C 056 30). DMRP and BMR thank the Interdisciplinary Research Center 'Advanced Protein Technologies' of the University of Potsdam for financial support. BMR thanks the Fonds der Chemischen Industrie for financial support (No 0164389). DMRP acknowledges financial support by the Bundesministerium fuer Bildung und Forschung (BMBF) (GABI-FUTURE grant 0315046). BMR thanks the BMBF for funding of the systems biology research unit 'GoFORSYS – Potsdam-Golm BMBF Forschungseinrichtung zur Systembiologie. Photosynthesis and Growth; a Systems Biology Based Approach' (FKZ 0313924).

Thanks to Luiz Gustavo Guedes Correa (DAAD fellowship) for primer testing, to Raúl Trejos-Espinosa (GoFORSYS) for providing *Chlamydomonas reinhardtii* cDNA, to Agnieszka Janiak (University of Silesia, Katowice, Poland) for providing *Hordeum vulgare* RNA and to Anika Wiese and Anup Karwa (VaTEP members at ICG III, FZ-Juelich, Germany) for testing the program.

References

- Higuchi R, Fockler C, Dollinger G, Watson R: **Kinetic PCR analysis: real-time monitoring of DNA amplification reactions.** *Biotechnology (NY)* 1993, **11**:1026-1030.
- Czechowski T, Bari RP, Stitt M, Scheible W, Udvardi MK: **Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes.** *Plant J* 2004, **38**:366-379.
- Caldana C, Scheible W, Mueller-Roeber B, Ruzicic S: **A quantitative RT-PCR platform for high-throughput expression profiling of 2500 rice transcription factors.** *Plant Methods* 2007, **3**:7.
- Horak CE, Snyder M: **Global analysis of gene expression in yeast.** *Funct Integr Genomics* 2002, **2**:171-180.
- Gordon PMK, Sensen CW: **Osprey: a comprehensive tool employing novel methods for the design of oligonucleotides for DNA sequencing and microarrays.** *Nucl Acids Res* 2004, **32**:e133.
- Fredslund J, Lange M: **Primique: automatic design of specific PCR primers for each sequence in a family.** *BMC Bioinformatics* 2007, **8**:369.
- Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132**:365-386.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM: **Primer3Plus, an enhanced web interface to Primer3.** *Nucl Acids Res* 2007, **35**:W71-4.
- Wrobel G, Kokocinski F, Lichter P: **AutoPrime: selecting primers for expressed sequences.** *Genome Biology* 2004, **5**:P11.
- You FM, Huo N, Gu YQ, Luo M, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD: **BatchPrimer3: a high throughput web application for PCR and sequencing primer design.** *BMC Bioinformatics* 2008, **9**:253.
- Pattyn F, Robbrecht P, De Paepe A, Speleman F, Vandesompele J: **RTPrimerDB: the real-time PCR primer and probe database, major update 2006.** *Nucl Acids Res* 2006, **34**:D684-8.
- Wang X, Seed B: **A PCR primer bank for quantitative gene expression analysis.** *Nucl Acids Res* 2003, **31**:e154.
- Cui W, Taub DD, Gardner K: **qPrimerDepot: a primer database for quantitative real time PCR.** *Nucl Acids Res* 2007, **35**:D805-9.
- Han S, Kim D: **AtRTPrimer: database for Arabidopsis genome-wide homogeneous and specific RT-PCR primer-pairs.** *BMC Bioinformatics* 2006, **7**:179.
- Fredslund J: **DATFAP: a database of primers and homology alignments for transcription factors from 13 plant species.** *BMC Genomics* 2008, **9**:140.

16. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucl Acids Res* 1997, **25**:3389-3402.
17. SantaLucia JJ: **A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.** *Proc Natl Acad Sci USA* 1998, **95**:1460-1465.
18. Sambrook J, Sambrook J, Maniatis T: *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press; 2001.
19. May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupperecht J, Weiss J, Recuenco-Munoz L, Ebenhoh O, Weckwerth W, Walther D: **Metabolomics- and Proteomics-Assisted Genome Annotation and Analysis of the Draft Metabolic Network of *Chlamydomonas reinhardtii*.** *Genetics* 2008, **179**:157-166.
20. Kwasniewski M, Szarejko I: **Molecular Cloning and Characterization of beta-Expansin Gene Related to Root Hair Formation in Barley.** *Plant Physiol* 2006, **141**:1149-1158.
21. Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM: **Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data.** *Neurosci Lett* 2003, **339**:62-66.
22. Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP: **The TIGR Plant Transcript Assemblies database.** *Nucl Acids Res* 2007, **35**:D846-51.
23. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L: **Database resources of the National Center for Biotechnology.** *Nucl Acids Res* 2003, **31**:28-33.
24. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E: **The Arabidopsis Information Resource (TAIR): gene structure and function annotation.** *Nucl Acids Res* 2008, **36**:D1009-14.
25. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucl Acids Res* 2007, **35**:D61-5.
26. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L, Marshall WF, Qu L, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen C, Cognat V, Croft MT, Dent R, Dutcher S, Fernández E, Fukuzawa H, González-Ballester D, González-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanov M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral J, Riaño-Pachón DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen C, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martínez D, Ngau WCA, Otilar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR: **The *Chlamydomonas* genome reveals the evolution of key animal and plant functions.** *Science* 2007, **318**:245-250.
27. FluBase: **The FlyBase database of the *Drosophila* genome projects and community literature.** *Nucl Acids Res* 2003, **31**:172-175.
28. Yamasaki C, Murakami K, Fujii Y, Sato Y, Harada E, Takeda J, Taniya T, Sakate R, Kikugawa S, Shimada M, Tanino M, Koyanagi KO, Barrero RA, Gough C, Chun H, Habara T, Hanaoka H, Hayakawa Y, Hilton PB, Kaneko Y, Kanno M, Kawahara Y, Kawamura T, Matsuya A, Nagata N, Nishikata K, Noda AO, Nurimoto S, Saichi N, Sakai H, Sanbonmatsu R, Shiba R, Suzuki M, Takabayashi K, Takahashi A, Tamura T, Tanaka M, Tanaka S, Todokoro F, Yamaguchi K, Yamamoto N, Okido T, Mashima J, Hashizume A, Jin L, Lee K, Lin Y, Nozaki A, Sakai K, Tada M, Miyazaki S, Makino T, Ohyanagi H, Osato N, Tanaka N, Suzuki Y, Ikeo K, Saitou N, Sugawara H, O'Donovan C, Kulikova T, Whitfield E, Halligan B, Shimoyama M, Twigger S, Yura K, Kimura K, Yasuda T, Nishikawa T, Akiyama Y, Motono C, Mukai Y, Nagasaki H, Suwa M, Horton P, Kikuno R, Ohara O, Lancet D, Eveno E, Graudens E, Imbeaud S, Debily MA, Hayashizaki Y, Amid C, Han M, Osanger A, Endo T, Thomas MA, Hirakawa M, Makalowski W, Nakao M, Kim N, Yoo H, De Souza SJ, Bonaldo MDF, Niimura Y, Kuryshev V, Schupp I, Wiemann S, Bellgard M, Shionyu M, Jia L, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Zhang Q, Go M, Minoshima S, Ohtsubo M, Hanada K, Tonellato P, Isogai T, Zhang J, Lenhard B, Kim S, Chen Z, Hinz U, Estreicher A, Nakai K, Makalowska I, Hide W, Tiffin N, Wilming L, Chakraborty R, Soares MB, Chiusano ML, Suzuki Y, Auffray C, Yamaguchi-Kabata Y, Itoh T, Hishiki T, Fukuchi S, Nishikawa K, Sugano S, Nomura N, Tateno Y, Imanishi T, Gojbori T: **The H-Invitational Database (H-InvDB), a comprehensive annotation resource for human genes and transcripts.** *Nucl Acids Res* 2008, **36**:D793-9.
29. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR: **The TIGR Rice Genome Annotation Resource: improvements and new features.** *Nucl Acids Res* 2007, **35**:D883-7.
30. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perraud P, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu S, Stueber K, Theodoulou FL, Tu H, Peer Y Van de, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL: **The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants.** *Science* 2008, **319**:64-69.
31. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroove S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Lepél J, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Ruzumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky T, Terry A, Tsai C, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Peer Y Van de, Rokhsar D: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**:1596-1604.
32. Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, Dunn B, Dwight S, Riles L, Mortimer RK, Botstein D: **Genetic and physical maps of *Saccharomyces cerevisiae*.** *Nature* 1997, **387**:67-73.
33. Jaillon O, Aury J, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguency P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gasparo G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon A, Weissenbach J, Quétier F, Wincker P: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla.** *Nature* 2007, **449**:463-467.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

3.1 Summary

To gain deepened understanding of the mechanisms behind biomass accumulation, it is important to study plant growth behavior. Here we describe an automated growth phenotyping platform for *Arabidopsis thaliana* with an annotation and analysis pipeline that makes it straightforward for the user to add important experimental information (plant genotypes, treatment conditions and annotation of plant development stages for the captured images) to the automatically collected phenotypic information. The pipeline performs statistical analyses and rosette area data modeling using linear mixed-effects models and reports reproducible quantitative results for areas and relative growth rates (RGR), corrected for variations between individual plants within a genotype and known fixed effects such as photosynthetic photon fluence rate differences in the growth chamber. The technical variations within the system are very low; the technical coefficient of variation (CV) for rosette area is generally below 2 % while we observed a biological CV of 8-12 % for the rosette area within a genotype, meaning that weak phenotypes are detectable even without prior knowledge of the nature of the growth phenotype or at what development stage it may appear. With our system one can link quantitative and qualitative changes in growth behavior to specific plant developmental stages with a minimum of manual effort, making it possible to perform highly informative analyses in larger screens. To demonstrate the quantitative capabilities of the method, we present data measured on the growth-impaired starch excess mutant *sex4-3*, which shows retardation in seedling establishment and reduced RGR and area throughout development.

3.2 Introduction

In order to gain deeper understanding of the functional role of gene networks and the basis for biomass production, it is necessary to combine datasets of molecular characteristics (such as primary and secondary metabolites, proteins and messenger RNA concentrations) with data about micro- and macroscopic developmental processes. In current plant systems biology projects, great efforts are being undertaken to provide the scientific community with quantitative data for the major molecular species of many model plants such as *Arabidopsis thaliana*. However,

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

quantitative information on plant development characteristics (e.g. leaf area, biomass and developmental stage) is rare and usually presented for a smaller number of genotypes. Larger screens often include only qualitative information about plant phenotypes, e.g. (Kuromori *et al.*, 2006), which although helpful for detecting strong phenotypes are less informative when developmental processes as such are being studied.

Currently used non-invasive aerial tissue phenotyping methods for smaller plants like *Arabidopsis* depend on highly manual workflows that require extensive plant handling as well as manual measurements with human intervention. Such protocols are often characterized by low precision and possible bias in the results due to semi-invasive plant handling, small sample populations or subjective manual interpretation. A major bias is the possible impact of the early stages of development (germination and seedling establishment), which can be difficult to account for when considering phenotypes at later stages. Typically, repetitive daily inspections of each individual plant are necessary; otherwise misleading or exaggerated conclusions are easily made. Invasive methods (typically cutting off leaves followed by weighing and/or scanning) are due to their nature less suitable for longer time series or when the individual plant variability is high. Even though they can be more precise in the measurement itself, huge sample populations are typically necessary introducing more complexity in plant handling and challenging the feasibility of bigger screens.

Therefore, for more un-biased and effective screening of growth phenotypes, we suggest the use of non-invasive automated phenotyping. Non-invasive 2-D-image phenotyping in high throughput for smaller plants was first described in 1999 (Leister *et al.*, 1999), reporting the applicability of projected leaf area as a proxy for biomass. However, to date, only a few larger screens of *Arabidopsis* mutant collections using such a system have been reported, notably the one carried out by the same research group (Varotto *et al.*, 2000). Other recent screens of larger mutant populations were carried out with manual classification of phenotypes (Kuromori *et al.*, 2006).

Current medium- to high-throughput (hundreds to thousands of plants) methods using cameras on robotic arms include the GROWSCREEN/FLUORO setup (Walter *et al.*, 2007; Jansen *et al.*, 2009), employing chlorophyll measurement and automated leaf counting, and the PHENOPSIS system (Granier *et al.*, 2006), featuring automated soil water content control, hence allowing for soil water deficit response screens. The DISP system, recently adapted for *Arabidopsis* (Wiese *et al.*,

2007), provides high temporal and spatial resolution of growth patterns of single leaves, but is a low-throughput method.

Here we present our growth phenotyping system, based on the commercially available LemnaTec Scanalyzer HTS system which ships with a software for image capture and analysis as well as a database for storing and organizing the captured images and the analysis data. The main differences to the systems described above are the enhanced level of automation (barcode tracking of trays) and that image analysis is fully automated. In addition to software shipped with the system, we have developed new modules to simplify larger screens (pot position randomization, label printing) and provide a complete analysis pipeline for rapid developmental stage annotation, data quality control, as well as statistical analyses and data plotting. These additional modules can be applied to data originating from other similar systems with minor modifications.

To test the capabilities of the system, to get practical experience with plant handling and to tune analysis parameters we repeated several batches of measurements under normal growth conditions with the *Arabidopsis* wild-type and the starch degradation mutant *sex4-3* (Niittylä *et al.*, 2006). The *sex4-3* null mutant has a strongly reduced rate of starch degradation in the dark, greatly reducing its energy reserves and hence capacity to grow (Zeeman *et al.*, 1998; Niittylä *et al.*, 2006). We show that the combination of detailed straightforward developmental stage annotation with quantitative area data, as provided by the organized image capture and analysis system, can be a powerful tool for rapidly detecting relevant growth phenotypes in larger screens, even of weaker phenotypes that would not be detectable with the naked human eye.

3.3 Results

3.3.1 System and experiment description

The system consists of a plant imaging chamber holding four plant trays at a time (see **Figure 3.1**), with a digital camera and a barcode scanner placed on a robotic arm and connected to the controlling computer (similar to the GROWSCREEN/FLUORO and PHENOPSIS systems). The software delivered with the system allows for detailed configuration of the pot and barcode positions, camera settings (zoom, focus, aperture, shutter speed) and the image analysis (see Experimental procedures for details).

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

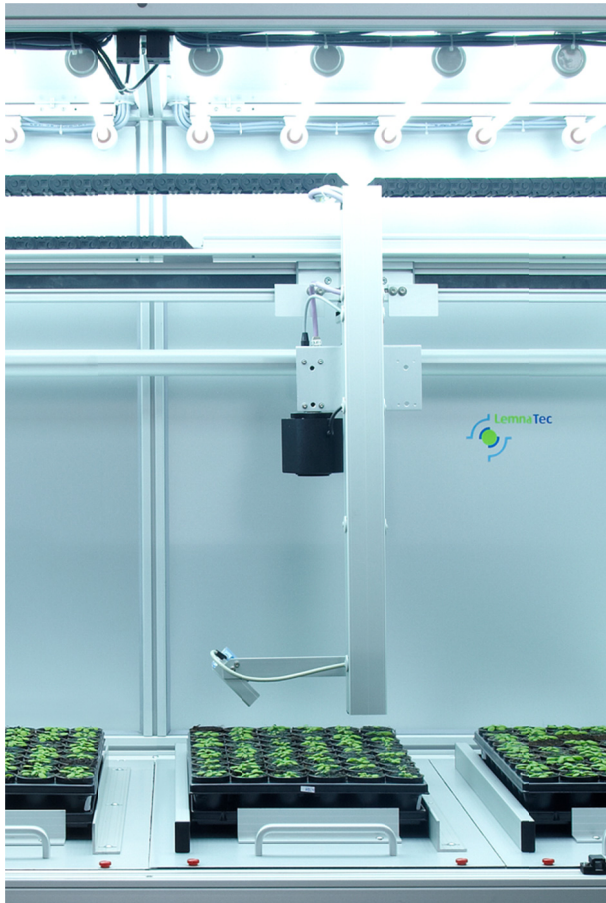


Figure 3.1: The imaging chamber. The robotic arm with the camera and barcode reader over a QuickPot 54R tray is shown. When the chamber is closed, the light is evenly distributed to provide optimal imaging conditions. The light spectrum and photon fluence rate are similar to those in the growth chamber.

We performed one measurement daily, three hours after the lights were switched on (± 30 min). The imaging system is activated and after a ten-minute lamp warm up, to avoid spectral variations in the illumination, it is ready for imaging. During this warm-up time, the plants to be measured are manually inspected by the researcher to remove doubles (when more than one seed germinated), and bigger patches of mosses, algae or other foreign items. The plants are then transferred from the growth chamber into the imaging cabinet and photographed by the camera. The barcode scanner identifies the trays and saves this information along with the images into a database. In parallel, a background process performs the image analysis, saving this information as well into the database. **Figure 3.2** shows an example of the image analysis of a plant.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

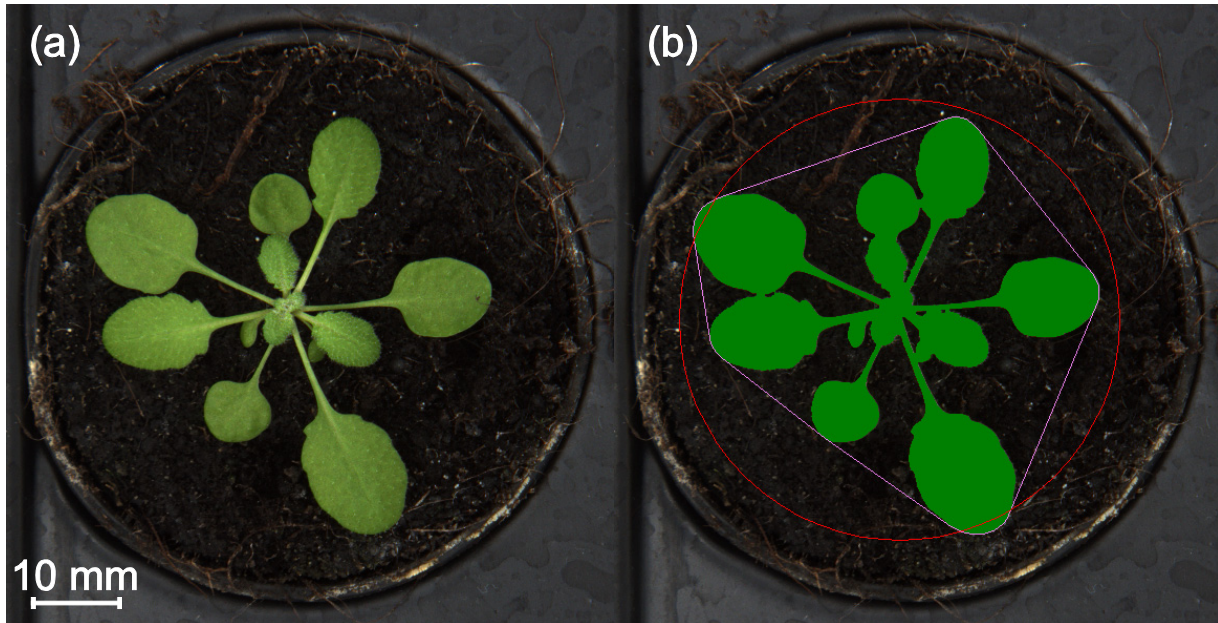


Figure 3.2: Example of a plant image before and after processing. Plant image before (a) and after (b) analysis using the LemnaGrid software delivered with the system. In (b) the green area is the detected leaf area and the purple line outlines the convex hull; the compactness is the total leaf area divided by the convex hull area.

Four trays with 15, 35 or 54 pots can be completely imaged in 7-14 minutes (**Table 3.1**). When including manual tray handling, sixteen (54- and 35-pot trays) to twenty-four (15-pot trays) trays can be manually inspected and imaged per hour, giving a throughput of 360, 560 or 864 plants per hour, making it possible to screen up to ~7000 plants concurrently when using several growth chambers with shifted day-night periods. Each tray spends only 20 minutes outside the growth cabinet; temperature and photosynthetic photon fluence rate within the imaging setup are close to the growth conditions minimizing the stress on the plants; only humidity is not controlled.

Table 3.1: Comparison of phenotyping capabilities and plant densities for different tray types.

Tray type	Mean imaging time per tray	Hourly throughput (plants hour ⁻¹) ^a	Daily throughput (plants day ⁻¹) ^a	Plant density (plants m ⁻²)
QuickPot 54R	2:30	864	6912	300
QuickPot 35R	3:00	560	4480	190
QuickPot 15RW	1:45	360	2880	83

^a Including manual handling time for one operator.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

Initially, we set up the system to work with 35-pot trays using 6-cm-diameter round pots (QuickPot 35R trays) and performed measurements on three batches of plants. To increase the daily throughput and minimize area usage in the growth chambers we then assessed two batches of plants with 54-pot trays (using 5-cm-diameter pots; QuickPot 54R trays) where the plant density and measurement throughput were increased by slightly more than 50 %. However, as the leaves between plants overlapped earlier (~7 days) the possible measurement periods were reduced. To solve this problem we reduced the number of plants per line to the half after ~30 days of measurement, placing them in a checkerboard fashion in the new trays. In this way the rosette areas of most plants could be measured until the 19-leaf stage was reached; after this we continued measurements to be able to easily determine the time of bolting and final leaf number, even though the rosette area would not be recorded. The quality control check for each image robustly detects overlapping of leaves from neighboring plants and invalidates the area from such a measurement.

We performed five experiments to assess the phenotyping pipeline capabilities. The first experiment included 175 plants from two genotypes, WT (90 plants) and *sex4-3* (85 plants), placed in five 35-pot trays in a checkerboard pattern. Then two experiments with 8 trays containing 14 genotypes with 20 plants from each (in 35-pot trays) followed, and subsequently two experiments with 24 genotypes with 18 plants from each (in 54-pot trays) were carried through. Each of the four last experiments contained the WT and *sex4-3* genotypes.

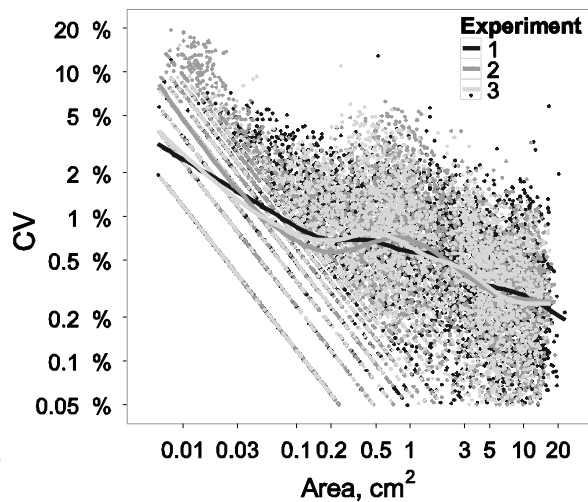
To assess the technical reproducibility of the image analysis we performed technical replicates of each image by rotating the trays and performing second measurements immediately afterwards, for the first three experiments. We placed different barcodes on each corner of the trays to be able to differentiate between their orientations. Assuming no growth between the replicated measurements (conducted within 15 minutes) the technical variance across different positions could be assessed. **Figure 3.3** is a log-log plot of the CV for the mean of each technical replicate and the mean of the rosette area. The CV is inversely correlated to the area; Spearman $\rho = -0.58$ ($p < 2 \times 10^{-16}$). At areas greater than 3 cm² (corresponding to a typical WT plant at the nine-leaf stage), the CV is typically below 1 % and only rises over 2 % when the area is lower than 0.2 cm² (a WT seedling at the two-leaf stage); for very small seedlings where only cotyledons are visible the CV is typically around 5 %. The results obtained from these experiments did not show significant differences. Since the

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

technical errors were low we skipped the duplicated measurements in the latter two experiments to reduce the time the plants spent outside the growth chamber.

Figure 3.3: Plot of CV of technical replicate measurements against mean rosette area.

The dependence of the coefficient of variance (CV) for technical replicate measurements on area is clearly illustrated in a log-log plot; the CV is inversely correlated with the rosette area (Spearman $\rho = -0.58$, $p < 2 \times 10^{-16}$). As a guide, 0.01, 0.2 and 3.0 cm² correspond to typical WT plants at cotyledon unfolding stage (BBCH 1.0), two-leaf stage (BBCH 1.02) and nine-leaf stage (BBCH 1.09), respectively. The lines



correspond to smoothed fits (using the LOESS algorithm) of the points from each of the three experiments. The diagonal organization of the dots at the lower left part of the graph is due to the discreteness of the data as areas and measurement errors approach unit sizes.

3.3.2 Data annotation

After finishing the measurements, the images were reviewed using our own annotation software. We developed this software as a web application (see **Figure 3.4a, b**) which shows the captured images and analysis results from the image capture database, and saves user-specified data along with the analysis results to be employed in the data pre-processing and analysis. The software is used to specify developmental stages according to the *Arabidopsis*-adapted BBCH scale (Boyes *et al.*, 2001) as well as to mark images where multiple plants were visible or other problems were apparent (such as infected plants/pots or damaged plants, which cannot be robustly detected automatically). After an introduction of about one hour, a researcher, lab technician or gardener can perform such annotation, typically needing two minutes to e.g. completely annotate 40 images collected for one plant over the course of an experiment, making it possible to fully annotate 240 plants (9600 images) in one working day of eight hours.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

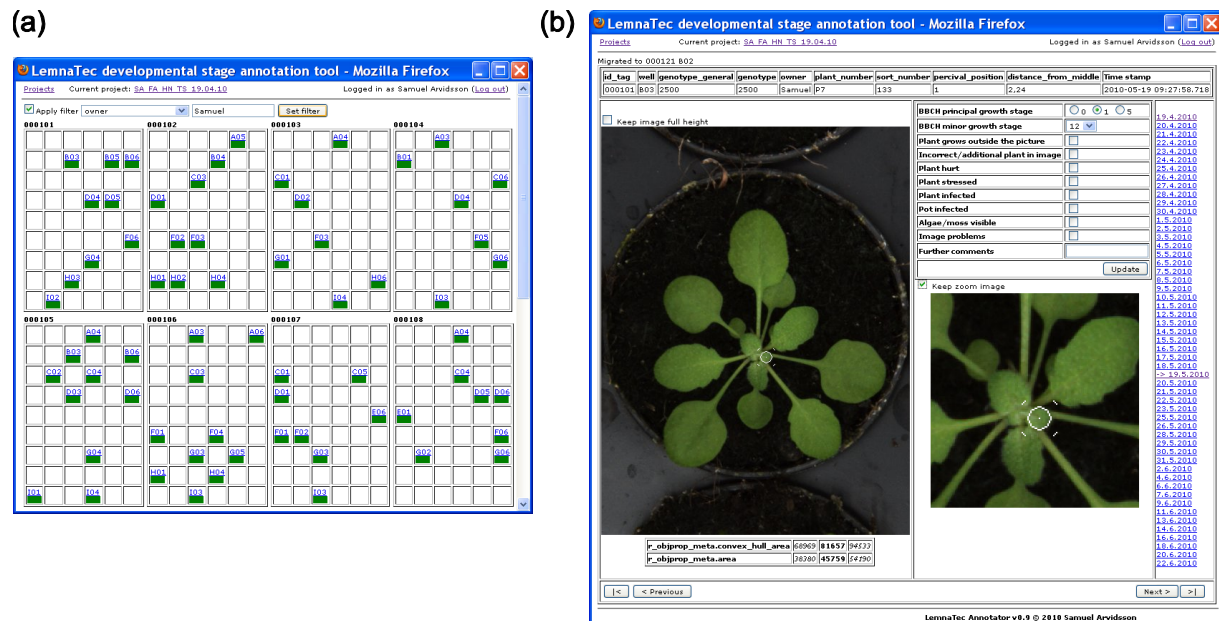


Figure 3.4: Screenshots of the annotation tool. (a) Overview of the trays (identified by different barcodes) as used in the current project; the user can filter the view to see particular plants, that e.g. belong to a certain experimenter or genotype. The color of the position indicates the progress in annotation; a fully green bar indicates that all images for that plant are annotated. (b) Annotation page for a plant. The user can enter specific quality control and development stage information for each image. When moving the mouse cursor over the image, it takes the shape of a 2-mm-diameter circle (white in the figure), which is used for counting leaf numbers. To visualize small objects (e.g. germinating seeds), a 2× zoom view of the center of the pot is available next to the main image.

3.3.3 Data analysis

Subsequently the data analysis was performed in four steps (*Steps 1 to 4* below) to aggregate the data, run statistical tests and draw plots for visual interpretation. Four types of automated analysis, carried out by running scripts in the R statistical environment (R Development Core Team, 2010), automatically produce text files with reports on significant findings, spreadsheet files with quantitative data, and PDF documents with data plots, providing the user with tools suitable for detailed investigation of putative phenotypes. The pre-processed data for all phenotypic parameters from the five experiments are included in the supplementary material (**Table S3.1**, **File S3.1**), as well as an R script for performing *Step 1* of the data analysis (**File S3.2**).

Step 1: The first analysis provides the user with an overview of possible effects. Plots of rosette area, compactness, relative growth rate (RGR), and leaf count are created with smoothed fits (LOESS – locally weighted scatterplot smoothing) of the data

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

points for each genotype, showing the mean trend and the standard error (SE) of the fit. To assess the influence a genotype might have on the phenotype, four ways to plot the data are used; the dependent variable is plotted as a function of days after sowing (DAS), days after germination (DAG), days after seedling establishment (DASE, defined as the day when the plant developed two primary leaves; BBCH 1.02), and the leaf count. The four plots for each phenotype-dependent variable are presented on the same page, so that the user can get an impression of which predictor variable might be most helpful to model the phenotype. A plot of the data collected from the five experiments for *sex4-3* and WT is shown in **Figure 3.5**.

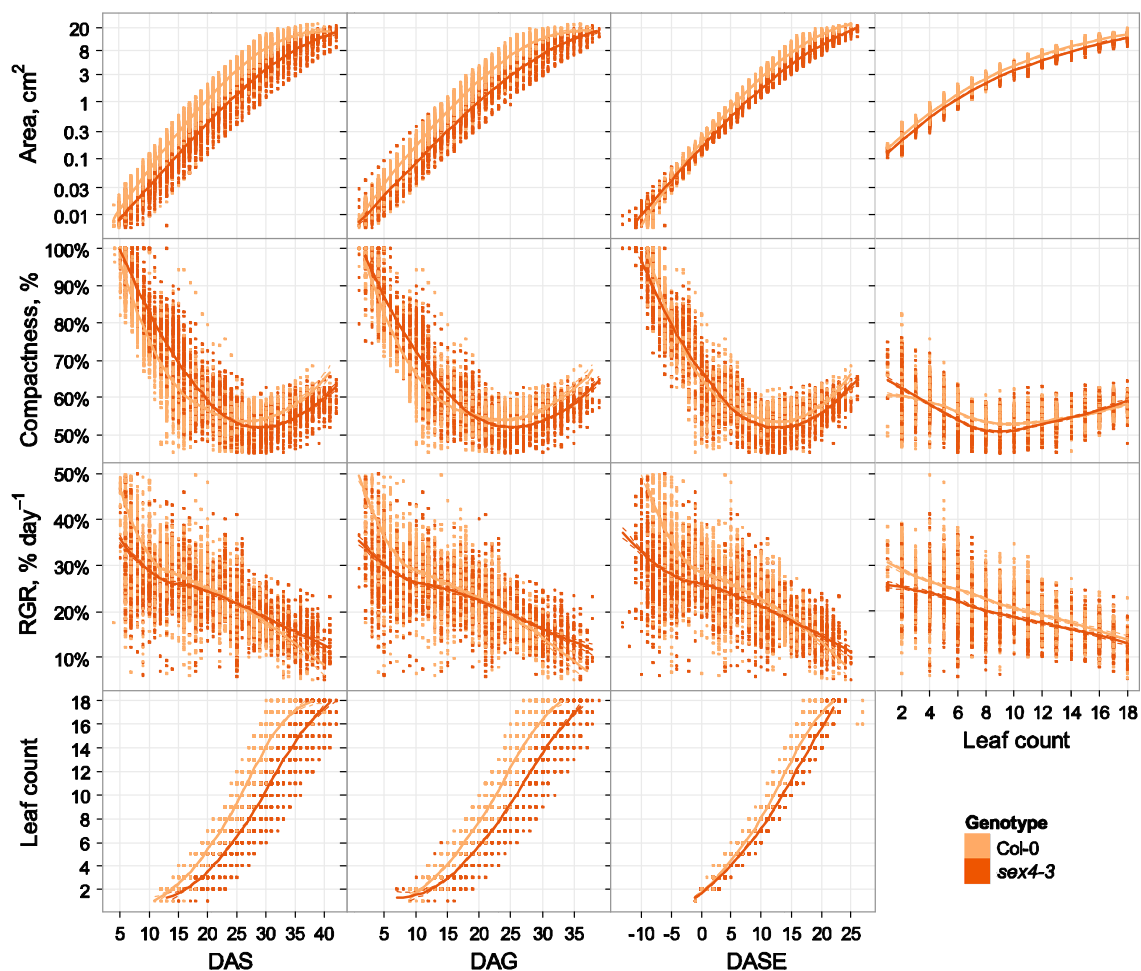


Figure 3.5: Smoothed fits for three phenotypic parameters using four different predictor variables. The plots indicate which phenotypes might be typical for the genotypes analyzed and the predictor variable(s) which model(s) the data optimally. Typically, the leaf count and DASE plots give lower variances, however it might be important to additionally check the DAS plots to verify whether punctual stresses occurred at a specific day.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

Step 2: In the second analysis step, the linear mixed-effects model given in (1) is fitted for rosette area, time, genotype and photosynthetic photon flux density (PPFD) data using the 'lme' function of the R nlme package (Pinheiro *et al.*, 2009). It is a standard mixed model with random intercepts that takes into account, that repeated measurements were done.

$$\begin{aligned} \log(\text{Area}_{ijk}) = & \mu + \alpha \text{DASE}_{ijk} + \beta \text{DASE}_{ijk}^2 + \text{gen}_i + \\ & + \text{gen}_i \times \text{DASE}_{ijk} + \delta \text{PPFD}_{ij} + \gamma \text{PPFD}_{ij} \text{DASE}_{ijk} + \zeta_{ij} + \varepsilon_{ijk} \end{aligned} \quad (1)$$

Here, $i = 1, \dots, g$ denotes the genotypes $1, \dots, g$, $j = 1, \dots, n_i$ the number of plants of genotype i , and $k = 1, \dots, m_{ij}$ the number of individual measurements for the j^{th} plant of the i^{th} genotype. ζ_{ij} and ε_{ijk} are normally distributed independent random variables, $\zeta_{ij} \sim \text{NIID}(0, \sigma_1^2)$ and $\varepsilon_{ijk} \sim \text{NIID}(0, \sigma_2^2)$. gen_i denotes the absolute effect of genotype i on the area and the interaction effect with time (intercept effect on the RGR, as shown below). PPFD_{ij} denotes difference to the mean photon flux recorded at the position of the plant in the growth chamber.

We developed this model based on the fact that the total leaf area, which can be considered as a proxy of shoot biomass (Walter *et al.*, 2007), generally follows an exponential function (Blackman, 1919). Since we only follow initial growth (until bolting at the longest), we chose not to use a more complex sigmoid function, which is commonly used to describe biomass accumulation in crop species (Poorter, 2002; Yin *et al.*, 2003) when the whole life cycle needs to be considered. The quadratic term accounts for the increasing leaf overlap and general growth deceleration (modeling a linear reduction of the RGR) and the ζ_{ij} term allows each individual plant its own intercept which is important, as seed weight and loading can cause individual variance in total absolute area. The factor δ models a response to the differences in photosynthetic photon fluence rates within the growth chamber, and was found to be significant in all experiments.

We found DASE to be a better predictor for modeling than DAS and DAG, which gave worse fits (according to estimated R^2 and Akaike's information criterion for the fitted models; this can as well be observed in **Figure 3.5**). Biologically this makes sense since seed loading and germination time, which can be expected to vary between individuals within a genotype, will have a strong impact on growth during the first two weeks of seedling establishment, after which genotypic and growth

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

conditions will take over as major effects. Genotypic influence on germination and seedling establishment is assessed separately in another step of the analysis (see *Step 3*).

In order to improve model predictions and reduce noise, we added detection of outlier plants and data points as a further feature. The algorithm considers data points with standardized residuals greater than 2 standard deviations as outliers, and plants having more than six such points are excluded from the model, which is then re-fitted with the existing data points. The user is warned and can manually inspect the growth curves for these plants; in some cases the areas measured on single days can be excluded for problematic plants (by setting quality control values in the annotation software). Typically, few plants (< 3 %) were considered as outliers, and they were almost exclusively found to have been damaged during handling or to suffer from fungi or insects, where the stress caused atypical growth patterns (not shown).

The fitting of this model allows the detection of whether a genotype has a significant effect or not on rosette area and RGR, and supplies the user with the estimated effects and *p*-values. Genotypes not showing any significant difference in comparison to WT are dropped from the dataset in order to improve the statistical power of the test.

Model (1) assumes a linear evolution of the RGR with a constant slope for all genotypes, as can be seen when studying the derivative of the right side of the formula:

$$RGR_{ijk} = \frac{d \log(\text{Area}_{ijk})}{dDASE_{ijk}} = \alpha + 2\beta DASE_{ijk} + gen_i + \gamma PPF_{ij} \quad (2)$$

This is an obvious simplification, since the RGR rather follows an inverse logarithmic decline over time (observed from experimental data; see **Figure 3.6**); however, to simplify the model this relationship can be simplified for segments with a linear function. To provide a better correction for this, a variant of the proposed model is fitted where an individual slope of the RGR is allowed for each genotype:

$$\begin{aligned} \log(\text{Area}_{ijk}) = & \mu + \alpha DASE_{ijk} + \beta DASE_{ijk}^2 + gen_i + gen_i \times DASE_{ijk} + \\ & + \lambda_i DASE_{ijk}^2 + \delta PPF_{ij} + \gamma PPF_{ij} DASE_{ijk} + \zeta_{ij} + \varepsilon_{ijk} \end{aligned} \quad (3)$$

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

This modification results in a better fit in many cases, especially when the RGR clearly changes faster over time in the mutant than in the WT; the RGR then follows the function

$$RGR_{ijk} = \alpha + 2\beta DASE_{ijk} + gen_i + 2\lambda_i DASE_{ijk} + \gamma PPF_{ij} \quad (4)$$

which allows an individual intercept and slope for each genotype. For genotypes showing major differences in RGR at different developmental stages, we recommend this updated formula, and in some cases local fits can be modeled to segments of the data for better quantification of the effects, if the RGR clearly evolves in a non-linear fashion.

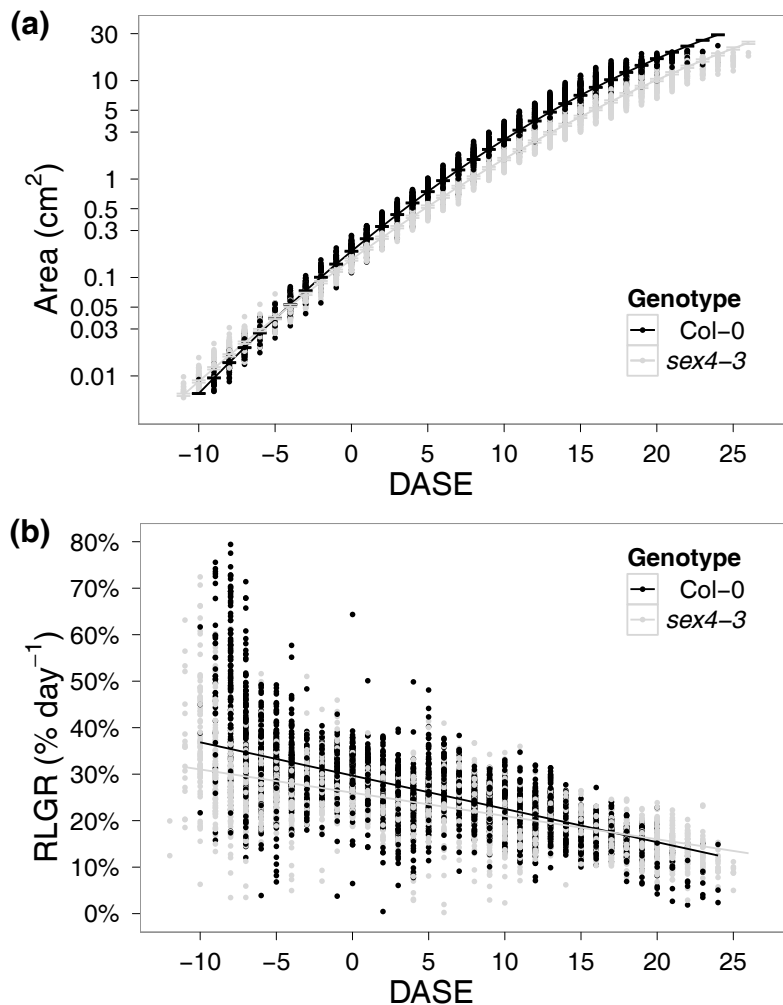


Figure 3.6: Linear mixed-effects model fit for rosette area and RGR of *sex4-3*. (a) The estimated model rosette area means are plotted as lines (error bars are SE), the real data are plotted as dots. (b) The estimated linear functions for RGR are plotted as lines, observed RGR data as dots.

The fitted mean RGR for each genotype over time is plotted together with a scatter plot of the observed values and provides a visual inspection of the fit (**Figure 3.6b**). Also, plots of the residuals with color marking for each genotype are provided for

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

verification that the models were well fitted without bias to any genotype (provided by **File S3.2**). To assess the reproducibility between experiments we compared the parameters fitted for the models (using formula (3)) for the WT and *sex4-3* plants in each of these experiments. The results are shown in **Table 3.2**.

Table 3.2: Comparison of observed phenotype values for *sex4-3* and WT between five experiments. The different phenotype values obtained with the mixed-effects models (3) for the five experiments ($n = 18-20$ plants per genotype) are shown. RGR difference was found to be the most robust phenotype value (CV of 9.5 %, compared to 32 % for rosette area). All values are calculated for 0 DASE.

Experiment	Relative difference in rosette area of <i>sex4-3</i> compared to WT (%)	RGR difference of <i>sex4-3</i> compared to WT ^a (% day ⁻¹)	RGR slope difference of <i>sex4-3</i> compared to WT ^b (% day ⁻²)
1	-15.9	-3.80	0.24
2	-26.5	-4.24	0.28
3	-16.7	-3.74	0.24
4	-13.5	-3.52	0.12
5	-22.4	-4.37	0.37
Median ± SD	-16.7 ± 5.28	-3.80 ± 0.36	0.24 ± 0.063

^a The median RGR for WT was 29.9 % day⁻¹.

^b The median RGR slope for WT was -0.71 % day⁻², thus the RGR for *sex4-3* decreased less over time in comparison to WT.

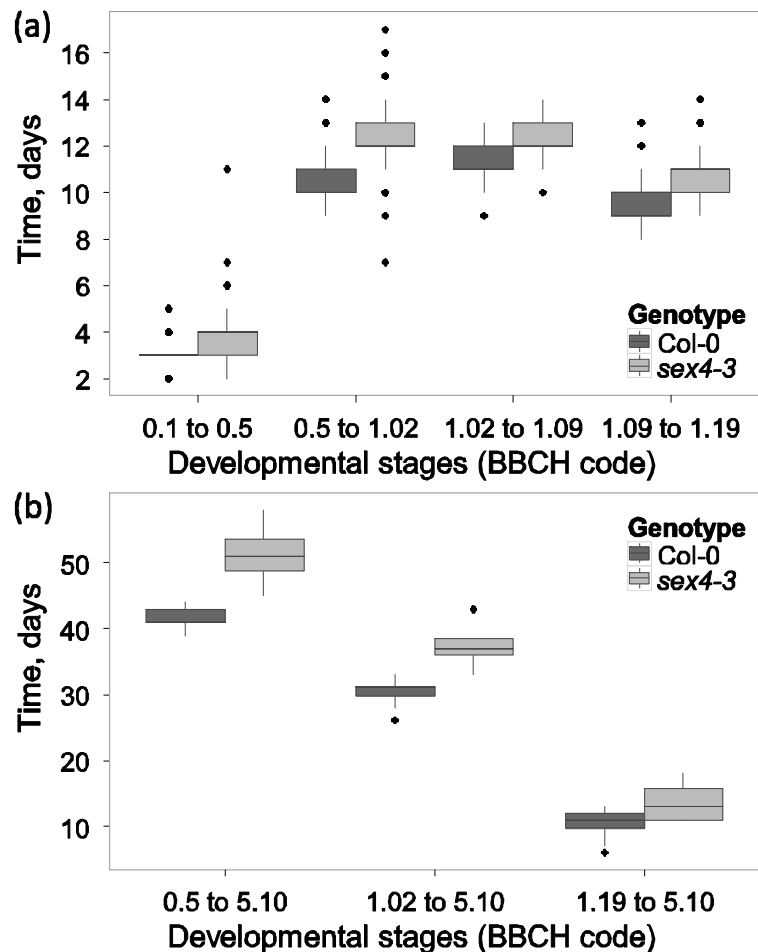
Step 3: This step analyzes the development timelines to detect significant genotype effects on the length of the different developmental stages. The following development times are assessed:

- Germination (BBCH 0.1 to 0.5)
- Seedling establishment (BBCH 0.5 to 1.02)
- Early rosette development (BBCH 1.02 to 1.09)
- Mid rosette development (BBCH 1.09 to 1.19)
- Late rosette development (BBCH 1.19 to 5.10)
- Bolting time (BBCH 0.5 to 5.10 and 1.02 to 5.10)

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

For each stage, a one-way ANOVA is carried out to detect whether the genotype of the mutant has a significant effect in comparison to the WT. The PPFD term is included in the ANOVA model and was found to be significant for all stages but seed germination. The data are then visualized in box plots (**Figure 3.7**), and a spreadsheet table is produced with mean effects of each genotype and statistics from the ANOVA. As can be seen in **Figure 3.7**, we separated the plot into two parts (early and late phenotypes). For bolting, we found it helpful to include comparisons based on germination, seedling establishment as well as late development, to give the user an indication as to where the main effects lie; if a plant was greatly delayed in seedling establishment or early rosette development, the time given by the BBCH 0.5 to 5.10 comparison could in some cases be misleading.

Figure 3.7: Development times box plots. The development times (in days) for important development steps (expressed as BBCH differences) are shown in the box plots. (a) Early development steps; 0.1 to 0.5 (germination), 0.5 to 1.02 (seedling establishment), 1.02 to 1.09 (early rosette development), 1.09 to 1.19 (late rosette development). (b) Late development steps: 0.5 to 5.10 (bolting time from germination), 1.02 to 5.10 (bolting time from seedling establishment), 1.19 to 5.10 (late development time).



A summary of the development times for sex4-3 and WT from the five experiments is presented in **Table 3.3**.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

Table 3.3: Development times for *sex4-3* and WT. Medians were calculated over the mean effect as determined by ANOVA in each of the five experiments ($n = 18-20$ plants per genotype in each experiment). The values in the difference column were found to be significant in at least three of the five experiments for all developmental stages.

Developmental stage	Development time (median \pm SD), days			
	BBCH	WT	<i>sex4-3</i>	<i>sex4-3</i> - WT
Germination	0.1 to 0.5	3.10 \pm 0.44	3.77 \pm 0.44	0.61 \pm 0.08
Seedling establishment	0.5 to 1.02	9.95 \pm 0.71	12.00 \pm 0.71	1.50 \pm 0.51
Early rosette	1.02 to 1.09	10.33 \pm 0.66	11.73 \pm 0.76	1.26 \pm 0.21
Mid rosette	1.09 to 1.19	9.73 \pm 0.62	10.93 \pm 0.62	0.69 \pm 0.51
Bolting time ^a (DAG ^b)	0.5 to 5.10	40.85	48.77	7.92
Bolting time ^a (DASE ^c)	1.02 to 5.10	30.78	37.00	6.22
Bolting time ^a (after appearance of leaf 19)	1.19 to 5.10	9.89	13.56	3.67

^a Bolting times were only measured in one experiment.

^b Days after germination.

^c Days after seedling establishment.

Step 4: Finally, the leaf development of each genotype is assessed at each point of the rosette development stages (BBCH 1.01 to 1.19). For this, one-way ANOVAs are performed for rosette area, compactness of rosette and relative growth rate (RGR) for the first day of each leaf stage. In this analysis the PPFD for each pot is accounted for, although we typically see less significant effects here compared to the time-based analyses in *Step 3*.

This analysis provides line plots of the main genotype effect of each phenotypic parameter versus leaf count, with information on significant differences in comparison to WT (**Figure 3.8**). The idea is to provide a time-independent view on rosette development to allow ‘fair’ comparisons of genotypes showing a delay of early development or retarded/accelerated leaf production, which cannot properly be accounted for in DAS, DAG or DASE-based analyses. The plots look similar to the LOESS-smoothed plots versus leaf count produced in *Step 2*; however, here we additionally provide the significance level of the ANOVA statistics for the difference to WT. A summary of the rosette areas recorded in the five experiments is given in

Table 3.4.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

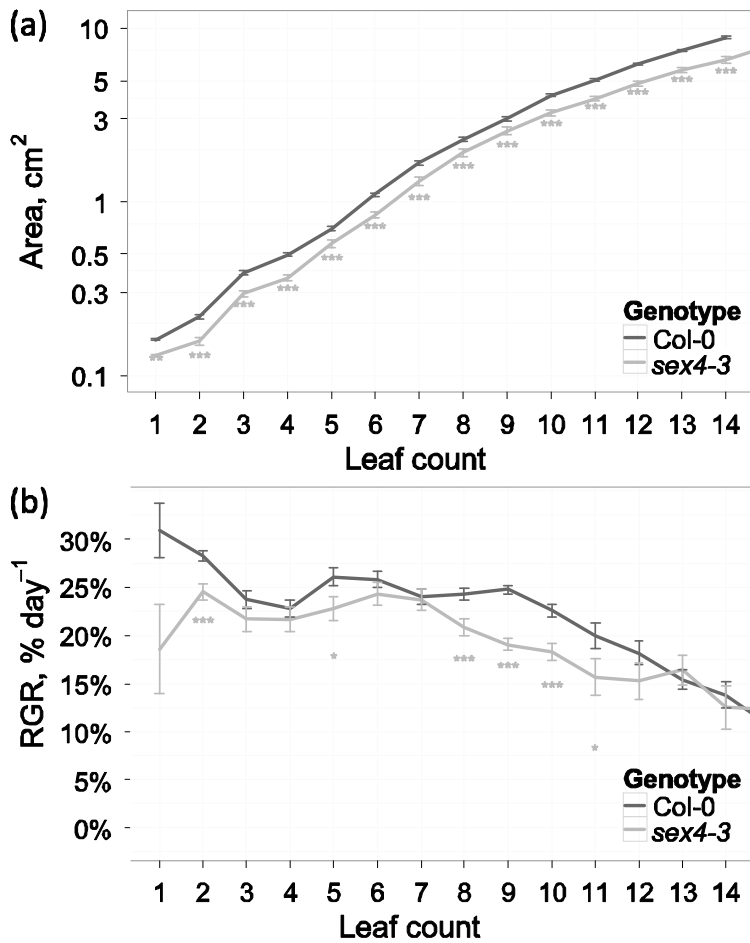


Figure 3.8: Development stage plots for phenotypes. Presented are results of *Step 4* of the statistical analysis which gives the mean effects for each genotype from a one-way ANOVA for each leaf stage, indicating values that are significantly different from WT (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). (a) Total rosette area. (b) Weighted mean of the RGR of three days (using the weights 1, 2 and 1) around the day when the leaf stage was reached.

Table 3.4: Rosette areas of *sex4-3* and WT at different growth stages. Medians and standard deviations were calculated over the mean effect as determined by one-way ANOVA in each of the five experiments ($n = 18-20$ plants per genotype in each experiment). The values in the area difference column were found to be significant ($p < 0.05$) in all five experiments for all development stages, except in one experiment where the area difference at the 1.18 stage was not significant.

BBCH	Rosette area WT (cm ²)	Rosette area <i>sex4-3</i> , (cm ²)	Rosette area difference of <i>sex4-3</i> compared to WT (%)
1.02	0.19 ± 0.02	0.16 ± 0.02	-14.9 ± 7.37
1.05	0.89 ± 0.09	0.66 ± 0.08	-17.1 ± 4.78
1.09	3.28 ± 0.20	2.57 ± 0.17	-18.6 ± 2.81
1.12	6.40 ± 0.29	5.33 ± 0.46	-19.7 ± 4.65
1.15	10.8 ± 0.57	8.83 ± 0.94	-17.5 ± 4.63
1.18	14.8 ± 0.66	12.5 ± 1.22	-15.3 ± 4.55

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

3.3.4 Comparison of experiments and tray types

In order to assess the robustness of the method, we performed several tests to compare the results obtained between the five experiments and two tray types used. As is shown in **Tables 3.2** and **3.4**, rosette area and RGR were reproducible across experiments, with the lowest CV for RGR. Indeed, when we compiled a complete dataset from the five experiments and added a term for the experiment (exp_l) as well as one for tray type ($tray_m$) to the existing model, resulting in (5), we could fit the complete data and test whether the experiment and tray type had significant effects:

$$\log(Area_{ijklm}) = \mu + \alpha DASE_{ijk} + \beta DASE_{ijk}^2 + gen_i + gen_i \times DASE_{ijk} + \lambda_i DASE_{ijk}^2 + \delta PPF_{ij} + \gamma PPF_{ij} DASE_{ijk} + exp_l + tray_m + \zeta_{ij} + \varepsilon_{ijk} \quad (5)$$

Since these factors are not orthogonal – the QuickPot 35R trays were used in the first three experiments, while the QuickPot 54R trays were used in the last two – we had to test them individually. After fitting the model we performed Tukey range tests comparing mean effects between each pair of experiments (testing whether the difference in means is zero), and found that the experiment means were significantly different ($p < 0.05$) in some cases; notably experiments two and three were significantly different from experiments one, four and five. However, we found no difference between the two tray types ($p > 0.53$). Thus, we concluded that it is necessary to include an experiment term in the model when combining data from several experiments.

3.3.5 Detection of weaker phenotypes

The *sex4-3* shows a strong growth reduction phenotype (a 12.7 % reduction in RGR and a 16.7 % smaller area at 0 DASE in comparison to WT) which could be easily detected by the naked eye already at 5 DASE where the total rosette area reduction in comparison to WT reaches 30 %. A screening system is only helpful when it can detect phenotypes not easily seen by the naked eye. We therefore included one mutant genotype of the *GRF9* (*GROWTH REGULATING FACTOR 9*) gene for which a tendency of slightly increased leaf area has been reported (Horiguchi, G. Kim, *et al.*, 2005), although in that study this effect was not found to be significant. When analyzing the area data of the *grf9* mutant in comparison to WT we could detect a significant increase of total rosette area (15.5 % increased area compared to WT at 0

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

DASE; $p = 0.00133$, $n = 20$; see **Figure 3.9a, b**), although the RGR did not significantly change ($p = 0.976$); thus the relative rosette area compared to WT stays constant throughout development.

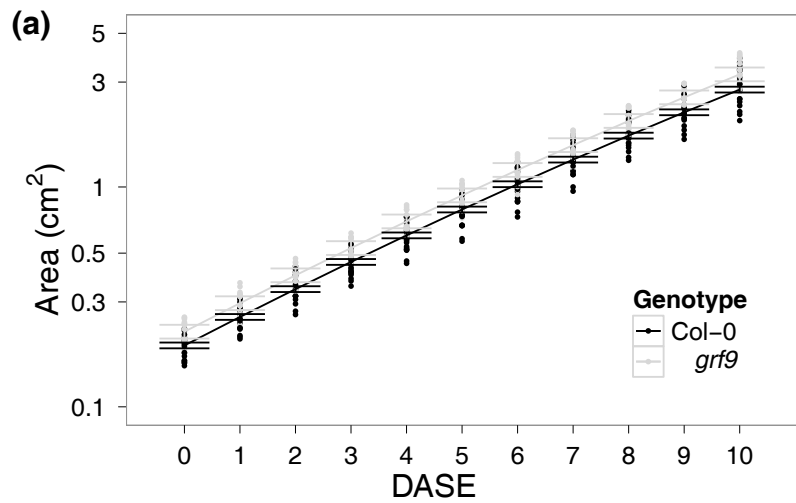
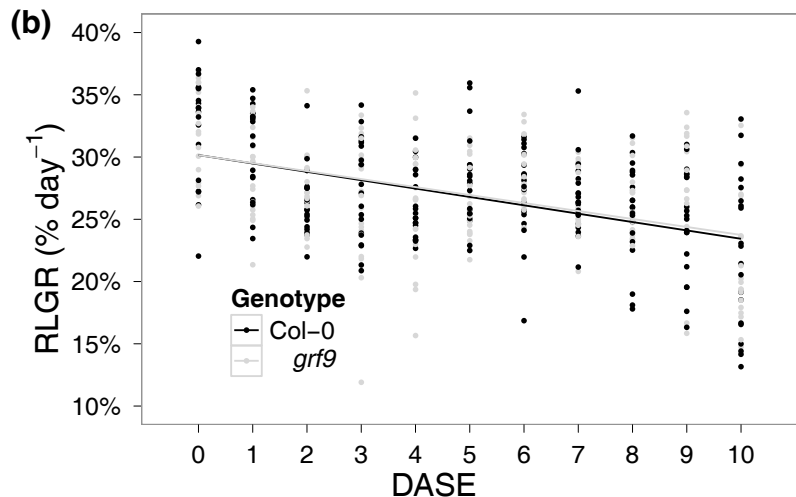


Figure 3.9: Linear mixed-effects model fit for rosette area and RGR of *grf9*. (a) The estimated model rosette area means are plotted as lines (error bars are SE), the real data are plotted as dots. (b) The estimated linear functions for RGR are plotted as lines, observed RGR data as dots.



3.4 Discussion

Although knowledge about the regulation of plant growth has strongly increased over the last decade, owing to the availability and analysis of numerous growth-affected mutants, many open questions remain. In particular, quantitative analysis of plant growth often lags behind the well-developed and sophisticated tools available for molecular and biochemical studies. Here we employed a commercially available plant growth analysis system (Scanalyzer HTS, LemnaTec) to develop an experimental pipeline for the rapid and robust analysis of plant growth parameters, using *Arabidopsis* as a model. Notably, we have established standard growth protocols, improved the plant tracking capabilities provided with the LemnaTec software, created an image annotation software for developmental stages, and designed a

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

data analysis pipeline performing quality control and analysis, including modeling and graphical representation of the data.

Based on our experimental findings we propose to use a linear mixed-effects model to fit the total rosette area data over time, by using the day after seedling establishment (DASE) as predictor variable. The proposed model fits well to the data from the five experiments performed, and provides quantitative information on genotype effects. When comparing the fitted coefficients between the five different experiments (see **Table 3.2**), we observed that the difference in absolute area is quite variable (CV = 32 %). However, the overall difference and slope of the RGR between *sex4-3* and WT is more stable between experiments (CV = 9.2 %), making RGR the parameter of choice for robust comparisons between genotypes, as found previously (Walter *et al.*, 2007), possibly even across experiments when proper controls are included.

By providing LOESS-smoothed plots (see **Figure 3.5**) of various aspects of the data we provide the user with a quick way to assess the data. In addition to typical statistical analysis, e.g. the approach proposed for classical growth analysis by (Hunt *et al.*, 2002), this helps the researcher in choosing a suitable predictor variable for the phenotype and thus avoid bias. Here we see that leaf count and DASE form the better basis for the analysis of the area phenotype of the *sex4-3* mutant, although DAG and DAS show greater differences between the mutant and WT means; the variance is also greater. Thus, DASE and leaf count can be used to model area, and will be more powerful than DAS and DAG. When studying the compactness and RGR plots, we again see better fits with DASE and leaf count, although these phenotypic variables follow more complex dynamics than area and thus are difficult to model. It can also be observed that the major difference in RGR for *sex4-3* in comparison to the wild-type occurs before the two-leaf stage and thus it is important to analyze the RGR using DASE as predictor variable, as the area information prior to the one-leaf stage cannot be assessed when using the leaf count as predictor variable.

Using the leaf count ANOVA tables and plots (see **Figure 3.8** and **Table 3.4**), quantification of rosette area, RGR and compactness can be carried out for specific growth stages. The strength of this analysis is that it compares plants at the same developmental stage, without considering the time needed to reach it. This can be an advantage when complex phenotypes occur that are difficult to describe on a DAS, DAG or DASE basis. The main disadvantage is that areas are associated with a rank

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

(the developmental stage or leaf count), and modeling over several stages may not always be meaningful. Currently, the analysis is performed at single stages, which makes it impossible to model and deduce individual plant variances from the genotype mean, as is possible in the DASE-based linear mixed-effects model approach (analysis *Step 1*).

A summary of the rosette areas recorded in the five experiments is found in **Table 3.4**. As can be seen the results concur with the modeled coefficients in *Step 1*; the area of the *sex4-3* mutant is generally 17 % lower than in the WT. It is also evident, that the CV of the differences within a genotype (determined to be 8-12 %) is clearly higher than the technical noise introduced by the method (typically below 2 %), which leads us to conclude that the method is highly suitable for detecting weak phenotypes, where the number of plants per genotype or treatment included will define the detection limit. We could confirm the ability to detect weak phenotypes with the *grf9* null mutant genotype, for which we record a significantly increased leaf area of 15.5 % throughout development (see **Figure 3.9a**); this difference is not easily detectable by the naked human eye.

As can be seen in **Table 3.3**, the *sex4-3* mutant grows slower than the WT during all development stages; in total it needed ~8 days more to progress from germination to bolting.

By using 54-pot trays, we could increase the plant density significantly over the 35-pot trays, without introducing any significant changes in growth phenotypes (as shown in the Results section); however, to follow the rosette area of all plants until bolting, it is necessary to increase the spacing between the pots. This was achieved by reducing the number of plants to half after ~30 days of measurement. Currently, we are refining experiments with 15-pot trays (using 8-cm pots) where it is possible to follow the area of plants from sowing to bolting without reducing the number of plants in the trays, therefore simplifying handling.

To our knowledge, no automated growth phenotyping system has been presented integrating manual annotation of images and automated modeling and data analysis. Also, the annotation and analysis methods we have developed are not limited to the imaging platform we are using; they could be adapted to different database structures to work with images and data outputs from other imaging platforms.

As major individual and genotypic variations occur in the germination and seedling establishment phases, which can introduce unwanted bias in phenotypic data of

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

rosette or inflorescence development, we suggest using DASE instead of the commonly used terms DAG or DAS as descriptor for the experimental time point. Of note, the term DAG is often incorrectly used when in fact plants at a later stage (BBCH 1.0 or 1.02) are compared. We recommend using the well-defined DASE term (as days after BBCH 1.02), unless seedling establishment is specifically studied, where usage of the equally well defined DAG term (days after BBCH 0.5) is sensible. The scientific community will gain more knowledge from the phenotype data produced when it is richly annotated with biologically relevant information and analyzed accordingly. Even though growth phenotypes could be assessed without adding manual information to the data, individual variances and delays in early growth stage will influence the data strongly (see **Figure 3.5**), and only very strong phenotypes can be detected with possibly biased conclusions. However, equipped with a properly set up imaging system and adequate analysis, such as the phenotyping pipeline presented here, even weak growth phenotypes can be assessed in larger screens, with a minimum of manual effort.

3.5 Materials and methods

3.5.1 Plant cultivation

Arabidopsis thaliana (ecotypes Col-0, the T-DNA insertion null mutants *sex4-3* (Niittylä *et al.*, 2006) and *grf9* (SALK_140746C obtained from NASC) were grown in growth cabinets with tightly controlled environmental conditions (Percival Scientific Inc., Perry, Iowa, <http://www.percival-scientific.com>) at a PPFD of $98 \pm 11 \mu\text{mol m}^{-2} \text{s}^{-1}$, 22 °C and 70 % humidity (day), 18 °C and 80 % humidity (night), at a 12 hour day/night cycle; except for the first ten nights in each growth cycle which were at 6 °C and 80 % humidity for stratification. Seeds were directly sown and grown in 5-cm round pots in 54-pot trays (QuickPot 54R, HerkuPlast-Kubern, Ering am Inn, Germany, <http://www.herkuplast.com>), 6-cm round pots in 35-pot trays (QuickPot 35R), or 8-cm pots in 15-pot trays (QuickPot 15RW). The substrate consisted of two layers, the lower being standard soil (Einheitserde Typ P, Einheitserde Gebr. Patzer, Sinntal-Jossa, Germany, <http://www.einheitserde.de>) mixed with vermiculite (9:1 soil:vermiculite), and the upper being standard soil without vermiculite (~1 cm) to ensure good foreground/background separation in the image analysis with smaller plants.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

3.5.2 *Development time definitions*

We define all development stages according to an Arabidopsis-adapted BBCH scale, similar to a previously proposed one (Boyes *et al.*, 2001), with the exception that we count leaves when they are longer than 2 mm (for more robust measures) as opposed to the suggested 1 mm. We use the following terms to describe time: DAS, Days After Sowing; DAG, Days After Germination (after BBCH stage 0.5; radicle emergence); and DASE, Days After Seedling Establishment (after BBCH stage 1.02; two primary leaves > 2 mm).

3.5.3 *Image capture*

The images were captured using an automated system with a robot arm holding a camera and barcode reader placed in a cabinet with optimal light control (Scanalyzer HTS, LemnaTec, Wuerselen, Germany, <http://www.lemnatec.com>). For each experiment, images were captured at a specific time window each day (three hours after onset of daylight, \pm 30 min). The image resolution was 81 px mm⁻² for the QuickPot 35R pot setup and 73.96 px mm⁻² for the QuickPot 54R setup.

3.5.4 *Image analysis*

The images were analyzed by the software provided with the image capturing system (LemnaGrid). The following steps are included in the image analysis grids (the specific thresholds and weights vary between the 54- and 35-pot setups due to different scale and slightly different exposure times):

1. A specific RGB channel weighting (4×Green - (3×Blue) - Red) is used to create a grayscale image which is especially bright for green-colored areas.
2. A binary mask is created by setting a threshold of intensity of 130 or higher.
3. Small holes (caused by particles or small, differently colored or darker parts of leaves) are filled (smaller than 13 pixels).
4. The binary mask is converted to objects.
5. To smoothen leaf edges and reconnect thin objects that became disconnected in the masking steps (typically thin petioles), the objects are grown and shrunk once with one pixel per operation.
6. Finally, all objects within the area of a given pot are composed into one; objects (leaves) that grew outside the pot area are included as long as they are connected to the rest of the plant.

3. A growth phenotyping pipeline for *Arabidopsis thaliana*

7. Phenotypic parameters are calculated for each plant (notably area, convex hull and compactness) and the data are saved into the database.

Manual image annotation: The images are then reviewed by the user for certain characteristics:

1. Image quality control is performed, e.g. to detect incorrect objects that cannot be handled by the automatic image analysis (foreign objects in the image, pieces of mosses, algae).
2. Additional information is annotated, notably the current plant developmental stage using a modified BBCH scale, where 2 mm is used as the threshold to detect new leaves (see above).

3.5.5 *Data pre-processing*

Using scripts developed for the R statistical language (R Development Core Team, 2010) quality control filters based on the manual image annotation are enforced, as well as a detection mechanism for plants growing out of the images – in these cases the plant area is invalidated, but the data on developmental stage is conserved for the time point. If technical replicates of images were captured on the same day, one value per phenotypic feature is extracted using different heuristic approaches depending on the phenotypic feature, such as to include only growing values for leaf and convex hull areas compared to the previous day (a shrinking plant is not expected) or increasing developmental stages (a plant will not lose a leaf) and then using the medians (as an outlier-robust alternative to the arithmetic mean) in cases where one single value cannot be favored. Finally, the data are integrated into a tabular form for each phenotypic feature (area, compactness, BBCH code) with the axes being the plants and the time points (days after sowing). These data can be exported in CSV format to spreadsheets, for manual inspection by the end user.

3.5.6 *Data analysis*

The data analysis steps are described in the Results section. For modeling, the 'lme' function from the R package nlme was used (Pineiro *et al.*, 2009). For ANOVA, the 'aov' function was used (Chambers *et al.*, 1992). For Tukey range tests the 'glht' function from the multcomp R package was employed (Hothorn *et al.*, 2008). All plots were created with the ggplot2 package (Wickham, 2009).

4. Identification of leaf growth-related tonoplast protein genes

4.1 Summary

The aim of this study was to identify tonoplast protein coding genes involved in leaf cell expansion, exploring the role of the tonoplast transporters and channels in driving controlled cell expansion through regulation of the osmotic pressure in the vacuole. Initially, a RT-qPCR platform for all known tonoplast protein coding genes (117 genes) was designed and supplemented with known growth-related genes; expansins, growth regulating factors (*GRFs*), cell cycle genes among others. Samples for transcription analysis were collected from leaves at different growth stages, cut into parts with higher and lower association with cell expansion. The measured transcript levels were evaluated using a template-based clustering method that ranked the genes according to their association with expanding leaf zones, and the 19 highest-scoring genes (including 11 tonoplast protein genes) were selected for further studies.

To study the role in leaf growth of these genes, an automated growth phenotyping pipeline using 2-D image analysis, developmental stage annotation and area modeling was employed, capable of quantifying total rosette area, relative growth rate (RGR) and development time phenotypes separately. Knockout mutant lines for the candidate growth-related genes were screened for phenotypes and effects were quantified. Interestingly, we identified strong area and RGR reductions as well as delayed development in knockout mutants for the tonoplast Na^+/H^+ -antiporter *NHX4* and for *EXPANSIN 6*, as well as slightly increased areas in mutants for *EXPANSIN 3*, *GROWTH-REGULATING FACTOR 9* and a subunit of the vacuolar H^+ -ATPase (*VHA-E3*), leading us to conclude that the tonoplast is worth taking a longer look at in studies of cell expansion regulation.

4.2 Introduction

In a very simplified view, plant leaf growth can be reduced to the sum of two processes, cell division and expansion (Beemster *et al.*, 2005). Both are partly independently and partly concurrently regulated, and both types of regulation must be taken into account when studying the growth of a determinate organ like the leaf. The phenomenon of compensation, causing greater cell sizes when the cell number is reduced (Tsukaya, 2008), can affect the impact on the growth behavior of the

4. Identification of leaf growth-related tonoplast protein genes

complete organ when studying loss-of-function mutants or overexpressors affecting or enhancing single genes. Although the mechanism behind compensation is not understood, it has been shown that the cell cycle regulation of endoreduplication (leading to polyteny) is crucial in this process (Donnelly *et al.*, 1999; Tsukaya, 2008), however the ploidy level does not directly dictate the final size of the leaf.

The driving force of cellular expansion is the turgor pressure, which can reach several bars (Hüsken *et al.*, 1978), caused by the high osmotic potential of the cell. The cell wall is relaxed ectopically, allowing slippage between the polymers (cell wall creep) through the action of expansins (weakening hydrogen bonds between polymers), pectin methylesterase, xyloglucan-endotransglycolase and -hydrolase, endo-(1,4)- β -d-glucanase and reactive oxygen species (modifying covalent links between polymers) (Cosgrove, 2005; Hamant and Traas, 2010). The internal osmotic potential causes the cell to swell and push apart the relaxed cell wall polymers, while newly synthesized polymers are deposited into the cell wall. These processes must be closely coordinated, and the osmotic pressure of the vacuole must be continuously adjusted to maintain turgor and allow uninterrupted cytosolic solute homeostasis even at high cell expansion rates. E.g., the regulation of aquaporins in expanding plant tissues has been shown to be important in several studies (Ludevid *et al.*, 1992; Chaumont *et al.*, 1998; Balk and de Boer, 1999). Cell wall relaxation alone does not seem to determine the limits for cell expansion, as increased expression of certain tonoplast aquaporins leads to reduced or increased cell sizes (Reisen *et al.*, 2003; Lin *et al.*, 2007). However, a recent report showed that the otherwise growth-enhancing overexpression of an endogenous aquaporin in protoplasts was blocked in cell wall presence (Okubo-Kurihara *et al.*, 2009).

To date, no comprehensive study has been conducted on the general role of tonoplast transporters and channels in growth; as most studies so far focused on specific aquaporins. To identify key tonoplast protein coding genes involved in growth, we have carried out a high-throughput expression study covering all tonoplast protein coding genes in *Arabidopsis thaliana*, followed by mutant line phenotyping, and describe the results here.

4.3 Results

4.3.1 *Arabidopsis thaliana* growth-related tonoplast RT-qPCR platform

Since there is no comprehensive list of tonoplast protein coding genes available, and

4. Identification of leaf growth-related tonoplast protein genes

no known protein motif for tonoplast localization that could be used for *in silico* predictions, we compiled such a list for *Arabidopsis thaliana* based on literature information. As a starting point, a list of identified genes in recent tonoplast proteomic studies (Szponarski *et al.*, 2004; Shimaoka *et al.*, 2004; C. Carter *et al.*, 2004; Endler *et al.*, 2006; Jaquinod *et al.*, 2007) was gathered, including all the detected transmembrane proteins from these studies. This list was then filtered, keeping the genes shown experimentally to have protein products localized to the tonoplast, for *Arabidopsis* or for orthologues in other species, or where the protein was detected in several proteomics studies and the predicted function were suggesting a tonoplast localization. We also added genes not showing up in these proteomic studies, but which were previously shown to have protein products localized to the tonoplast, many of which were found using SUBA; the 'Arabidopsis Subcellular Database' (Heazlewood *et al.*, 2007). In total, 117 genes encoding for tonoplast proteins were included.

Subsequently, genes known to be involved in different mechanisms of cellular growth and cell wall relaxation were included: all expansins (*EXPs*) known to be expressed in leaves (Sampedro and Cosgrove, 2005; Choi *et al.*, 2006), growth regulating factors (*GRFs*) (J.H. Kim *et al.*, 2003), *AINTEGUMENTA* (Mizukami and Fischer, 2000), *ARGOS* (Hu *et al.*, 2003), *ARGOS-like* (Hu *et al.*, 2006), *AGF1* (Matsushita *et al.*, 2007) and *AGL25/FLC* (Michaels and Amasino, 1999). Twelve cell cycle and cell cycle-related genes reported as putative indicators of cell division and cell endoreduplication were added to the list; *CDKA;1*, *CDKB1;1*, *CDKB1;2*, *CDKB2;1*, *CDKB2;2*, *CYCB1;1*, *CYCB1;2*, *CYCB1;3*, *KRP1*, *KRP2*, *KRP4* and *KRP5* (Donnelly *et al.*, 1999; Ormenese *et al.*, 2004; Boudolf *et al.*, 2004; Beemster *et al.*, 2005; del Pozo *et al.*, 2006). A summary of the gene families represented in the RT-qPCR platform is given in **Table 4.1**, a comprehensive list with references is available in the supplementary material (**Table S4.1**).

Furthermore, we included five reference genes reported to have stable expression over all development stages of *Arabidopsis thaliana* (Czechowski *et al.*, 2005), with different orders of magnitude of expression to reduce possible bias in RT-qPCR data normalization (Vandesompele *et al.*, 2002). Transcript-specific primers were designed for all the genes mentioned above and tested *in silico* for specificity against the whole *Arabidopsis* genome using QuantPrime (Arvidsson *et al.*, 2008). The complete list with primer sequences is found in the supplementary material (**Table**

4. Identification of leaf growth-related tonoplast protein genes

S4.2). Stringent quality control criteria were applied to only include primer pairs of high efficiency and specificity in the RT-qPCR platform (see materials and methods for details).

Table 4.1: A summary of the RT-qPCR platform. Here we list the categories and gene families represented in the growth-related tonoplast protein coding gene RT-qPCR platform. Gene families with at least three genes represented on the platform are listed.

Category	Gene family	Count
Cell cycle	Core cell cycle genes	12
Growth-related	Expansins	12
Growth-related	GRF transcription factors	9
Growth-related	Other or unknown	5
Tonoplast	Primary pumps (ATPases)	31
Tonoplast	Organic solute cotransporters	19
Tonoplast	Inorganic solute cotransporters	12
Tonoplast	Antiporters	10
Tonoplast	Aquaporins	10
Tonoplast	Ion channels	8
Tonoplast	ABC transporters	4
Tonoplast	Other or unknown	23
Reference	Other or unknown	5

4.3.2 Sampling of expansion-associated leaf regions and RT-qPCR data analysis

In order to identify candidate growth-related genes, we sampled and cut the blade of leaf #11 from wild type *Arabidopsis thaliana* Col-0 plants into three parts (equally wide; see **Figure 4.1** for a schematic representation) at three different leaf stages – 33 %, 50 % and 100 % expanded leaf; equivalent to 10, 15 and 30 mm length (excluding the petiole), under our growth conditions. The idea behind this sampling strategy was to compare the expression levels of our gene panel over rapidly expanding, moderately expanding and non-expanding mature leaf parts, as cell expansion has been shown to be concentrated to certain leaf regions (restricted in space and time) (Donnelly *et al.*, 1999; Beemster *et al.*, 2005; Wiese *et al.*, 2007). All sampling was carried through at 2 hours after lights-on, to include the effect of the morning growth peak which has been reported for young *Arabidopsis* leaves (Wiese *et al.*, 2007) and which could be reproduced by us under the sampling growth

4. Identification of leaf growth-related tonoplast protein genes

conditions using a similar technique (Berns *et al.*, 2007); see the figure and description in the supplementary material (**Figure S4.1**) for more details.

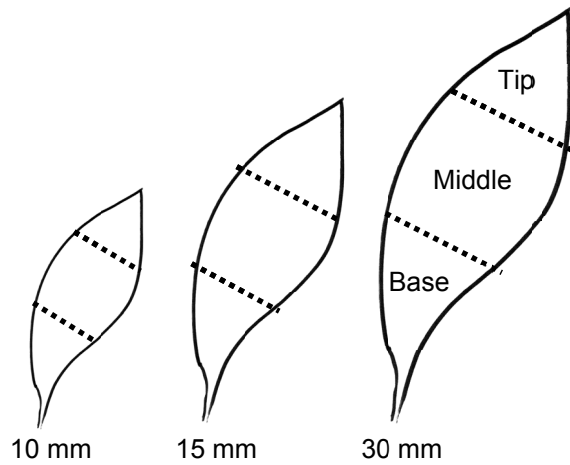


Figure 4.1: Schematic of leaf samples used for expression analysis. A schematic representation of the leaf parts sampled for expression analysis. We used the same leaf (number 11) at different stages of its development (10, 15 and 30 mm), and cut it in three equally long parts, 'base', 'middle' and 'tip', as shown in the figure.

Total RNA was extracted from the leaf samples, cDNA was synthesized and qPCR measurements using the above described platform were carried through. In order to rank genes according to their association with expanding regions, we constructed a scoring template for expression differences within and between leaves (see **Table 4.2**). Briefly, genes highly expressed in the base of young, rapidly expanding leaves get high scores, as well as extra points are given to genes which show a low and constant expression across fully expanded mature leaves.

Table 4.2: Scoring template for growth-association of genes. The expression values for each gene are verified against the rules which assign scores to expression differences between reference and sample cDNAs, to assign high scores to growth-associated genes.

Reference	Sample	Rule	Assigned score
30 mm base	10 mm base	$\Delta\Delta C_q > 2$	+5
30 mm base	10 mm base	$1 < \Delta\Delta C_q < 2$	+3
30 mm mid	10 mm middle	$\Delta\Delta C_q > 2$	+4
30 mm mid	10 mm middle	$1 < \Delta\Delta C_q < 2$	+2
10 mm tip	10 mm base	$\Delta\Delta C_q > 2$	+4
10 mm tip	10 mm base	$1 < \Delta\Delta C_q < 2$	+2
30 mm base	15 mm base	$\Delta\Delta C_q > 2$	+4
30 mm base	15 mm base	$1 < \Delta\Delta C_q < 2$	+2
15 mm tip	15 mm base	$\Delta\Delta C_q > 2$	+3
15 mm tip	15 mm base	$1 < \Delta\Delta C_q < 2$	+1
30 mm base	30 mm tip	$-1 < \Delta\Delta C_q < 1$	+1
30 mm base	30 mm middle	$-1 < \Delta\Delta C_q < 1$	+1

4. Identification of leaf growth-related tonoplast protein genes

The 29 highest scoring genes (of which 12 are tonoplast protein coding genes) are presented in **Table 4.3**; a complete score table is included in the supplementary materials (**Table S4.3**). The complete set of expression values, as ΔC_q values, is included in the supplementary materials in a list form (**Table S4.4**) as well as $20-\Delta C_q$ values in bar plots (**Figure S4.2**).

Table 4.3: The genes with the highest scores for growth-association.

Platform category	Gene locus	Gene name	Growth-association score
Tonoplast	AT1G16390	<i>OCT3</i>	20
Tonoplast	AT5G47450	<i>TIP2;3</i>	18
Tonoplast	AT3G16240	<i>TIP2;1</i>	17
Tonoplast	AT2G26690	<i>NTP2</i>	15
Tonoplast	AT3G06370	<i>NHX4</i>	14
Tonoplast	AT2G36830	<i>TIP1;1</i>	14
Tonoplast	AT3G51490	<i>TMT3</i>	12
Tonoplast	AT1G64200	<i>VHA-E3</i>	9
Tonoplast	AT2G48020	<i>ERD6-like</i>	9
Tonoplast	AT3G26520	<i>TIP1;2</i>	9
Tonoplast	AT5G13740	<i>ZIF1</i>	9
Tonoplast	AT5G62890	<i>NAT6</i>	9
Cell cycle	AT4G37490	<i>CYCB1;1</i>	18
Cell cycle	AT2G38620	<i>CDKB1;2</i>	16
Cell cycle	AT3G54180	<i>CDKB1;1</i>	16
Cell cycle	AT5G06150	<i>CYCB1;2</i>	16
Cell cycle	AT1G20930	<i>CDKB2;2</i>	15
Cell cycle	AT3G11520	<i>CYC2</i>	13
Growth-related	AT2G06200	<i>GRF6</i>	19
Growth-related	AT2G28950	<i>EXP6</i>	17
Growth-related	AT3G29030	<i>EXP5</i>	17
Growth-related	AT4G37750	<i>ANT</i>	16
Growth-related	AT1G26770	<i>EXP10</i>	15
Growth-related	AT2G20750	<i>EXPB1</i>	15
Growth-related	AT3G52910	<i>GRF4</i>	15
Growth-related	AT4G37740	<i>GRF2</i>	13
Growth-related	AT2G45480	<i>GRF9</i>	13
Growth-related	AT2G37640	<i>EXP3</i>	9
Growth-related	AT3G55500	<i>EXP16</i>	9

4. Identification of leaf growth-related tonoplast protein genes

The score threshold of 9 was selected based on the score distributions for the growth-related and cell cycle genes. We expected a larger fraction of these genes to have high scores in comparison to the tonoplast genes. This is indeed the case; while the tonoplast gene scores are lognormally distributed (with few high score outliers) with a median of 2 and a SD of 3.6, the growth-related and cell cycle gene scores show a double peaked distribution. If the split is made at a score of 9, one group with a median of score 2 (SD = 1.6, $n = 21$) and another at 15 (SD = 2.7, $n = 17$) appear. This led us to hypothesize that a similar split for the tonoplast gene group would give us a reasonable enrichment of growth association among the higher scoring genes. The high-scoring fraction of the tonoplast gene group (12 genes) was selected for further analysis, together with the high scoring cell cycle and growth-related genes (17 genes).

4.3.3 Selection of knockout mutant lines for the candidate genes

In order to study the function of the candidate genes, we obtained seeds for insertion knockout mutants of *Arabidopsis thaliana*, all in ecotype Col-0 background but from different collections, including T-DNA insertion lines of various origin and dSpm lines (A.F. Tissier *et al.*, 1999; Sessions *et al.*, 2002; Rosso *et al.*, 2003; Alonso *et al.*, 2003; Woody *et al.*, 2007). These were selected by the insertion point (determined with the SALK T-DNA Express website and the NASC AtEnsembl-viewer) to have insertions in coding regions; where this was not possible, we selected lines with insertions in the first introns or close promoters. We obtained one to six lines per gene, 65 lines in total (see supplementary material **Table S4.5** for a complete listing with estimated insertion points). After obtaining the seeds, we screened the progeny for homozygous insertions, checking genome insertions with PCR and verified the absence of transcript with RT-qPCR. Thus, we obtained the 24 lines listed in **Table 4.4** after two rounds of screening for each line; for the rest we failed to obtain homozygous progeny.

4. Identification of leaf growth-related tonoplast protein genes

Table 4.4: Knockout lines for which homozygous progeny was obtained and which were screened with growth phenotyping. The genes are listed according to their loci. The 'Line' column denotes the designation used in this study; the 'Line code' is the collection number as designated by the creator and identifies the germplasm at NASC.

Disrupted locus	Disrupted gene name	Genotype	Line code
AT1G16390	<i>OCT3</i>	<i>oct3</i>	SM_3_20320
AT1G64200	<i>VHA-E3</i>	<i>vha-e3</i>	GK-138C07
AT2G23150	<i>ATNRAMP3</i>	<i>nramp3</i>	SALK_023049
AT2G26690	<i>NTP2</i>	<i>ntp2</i>	WiscDsLox322_H05
AT2G28950	<i>EXP6</i>	<i>exp6</i>	GK-522C09
AT2G36830	<i>TIP1;1</i>	<i>tip1;1</i>	SM_3_32402
AT2G37640	<i>EXP3</i>	<i>exp3</i>	SALK_048023
AT2G45480	<i>ATGRF9</i>	<i>grf9</i>	SALK_140746
AT2G48020	<i>ERD6-like</i>	<i>erd6-like</i>	SALK_144885
AT3G06370	<i>NHX4</i>	<i>nhx4-1</i>	SALK_112901
AT3G06370	<i>NHX4</i>	<i>nhx4-2</i>	SAIL_87_A09
AT3G06370	<i>NHX4</i>	<i>nhx4-3</i>	GK-770A08
AT3G16240	<i>TIP2;1</i>	<i>tip2;1</i>	SM_3_39039
AT3G29030	<i>EXP5</i>	<i>exp5-1</i>	SALK_043239
AT3G29030	<i>EXP5</i>	<i>exp5-2</i>	WiscDsLox495_F06
AT3G52910	<i>ATGRF4</i>	<i>grf4</i>	SALK_077829
AT3G55500	<i>EXP16</i>	<i>exp16</i>	GK-863H08
AT4G37740	<i>ATGRF2</i>	<i>grf2</i>	SALK_003203
AT4G37750	<i>ANT</i>	<i>ant</i>	GK-874H08
AT5G13740	<i>ZIF1</i>	<i>zif1</i>	SALK_016418
AT5G47450	<i>TIP2;3</i>	<i>tip2;3-1</i>	SALK_127491
AT5G47450	<i>TIP2;3</i>	<i>tip2;3-2</i>	SALK_142179
AT5G62890	<i>NAT6-like</i>	<i>nat6-like-1</i>	GK-340A03
AT5G62890	<i>NAT6-like</i>	<i>nat6-like-2</i>	SALK_078079

4.3.4 Growth phenotype screening

In order to assess the actual importance in growth regulation and cellular expansion of the candidate growth-associated candidate genes, we carried through an extensive growth phenotyping study on the previously mentioned knockout lines using an automated image analysis phenotyping pipeline; for a full description of the phenotyping pipeline and data analysis see **Chapter 3**. Briefly, the rosette area is determined daily using an image capture and analysis platform, subsequently the

4. Identification of leaf growth-related tonoplast protein genes

images are annotated to describe growth stages using an adapted BBCH scale (Boyes *et al.*, 2001) and finally the data are analyzed, comparing and quantifying genotype effects on overall area and relative growth rate (RGR) as well as specific development times with linear mixed models (area and RGR) and ANOVA (development times). By including wild type Col-0 (WT) and the genotype *sex4-3* (Niittylä *et al.*, 2006) with a known, strongly growth-impaired phenotype in all batches we could compare data from several batches in one analysis (see **Chapter 3** for details). The results of the total rosette area modeling using linear mixed-effects models (formula (3) in **Chapter 3.3.3**) for the genotypes where significant phenotypes were observed are listed in **Table 4.5**; listing the quantified area and RGR effects at 0 days after seedling establishment (DASE, days from BBCH 1.02). Graphical representations of the predicted models and observed data are shown in **Figure 4.2** (the three *nhx4* lines) and **Figure 4.3** (*exp3*, *exp6*, *grf9* and *vha-e3*). Photos of representative plants at 14 DASE from the *exp3*, *exp6*, *grf9*, *vha-e3* and *nhx4* lines are shown in **Figure 4.4**. The results of the ANOVAs for development timelines are listed in **Table 4.6** and **4.7**, and the data are as well shown as box plots in **Figure 4.5**; but only for the *nhx4* lines as there are not enough data available for the other lines. In the supplementary material we included the complete pre-processed growth phenotyping dataset (**File S4.1**) and plots (**Figure S4.3**) with comparisons of four predictor variables (days after sowing – DAS, days after germination – DAG, DASE and leaf count).

Table 4.5: Growth effects for genotypes with significant effects when modeling area over time. Genotype effects as estimated with linear mixed effects model for 0 DASE, as area (here relative to WT Col-0) and RGR varies over time. Data are estimated group means \pm SE. * denotes $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ for the estimates, p reported by the ‘lme’ function call (Pinheiro *et al.*, 2009). n denotes the number of plants.

Genotype	Relative area (%)	RGR (% day ⁻¹)	n
Col-0	100 \pm 2.59	30.1 \pm 0.0464	71
<i>exp3</i>	112 \pm 4.03 **	30.2 \pm 0.0997	19
<i>exp6</i>	63.9 \pm 5.42 ***	28.9 \pm 0.11 ***	10
<i>grf9</i>	115 \pm 3.98 ***	30 \pm 0.0977	20
<i>nhx4-1</i>	72.4 \pm 7.40 ***	28.2 \pm 0.157 ***	5
<i>nhx4-2</i>	75.1 \pm 9.30 ***	29 \pm 0.207 ***	3
<i>nhx4-3</i>	85.7 \pm 3.72 ***	22.7 \pm 0.0692 ***	23
<i>vha-e3</i>	114 \pm 3.99 **	30.1 \pm 0.0983	20

4. Identification of leaf growth-related tonoplast protein genes

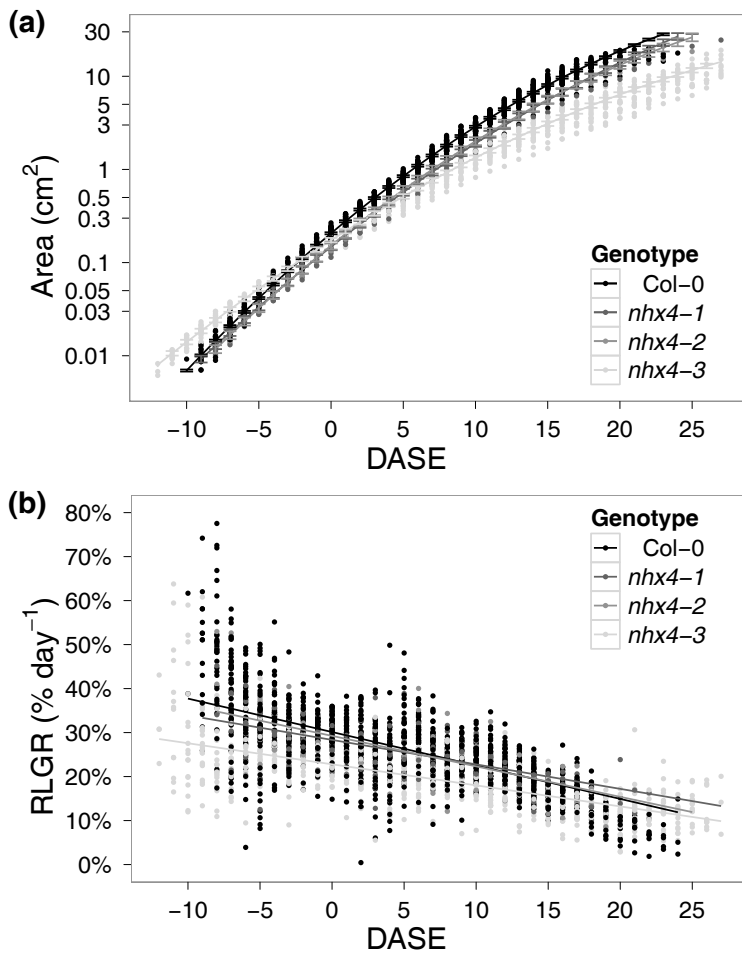


Figure 4.2: Area and RGR modeling for *nhx4* mutants.

Linear mixed-effects models of total rosette area and RGR, predicted using area data collected using automated image capture and analysis. (a) Area versus DASE (days after seedling establishment) as estimated with linear mixed models. (b) The relative growth rate (RGR) component of the model.

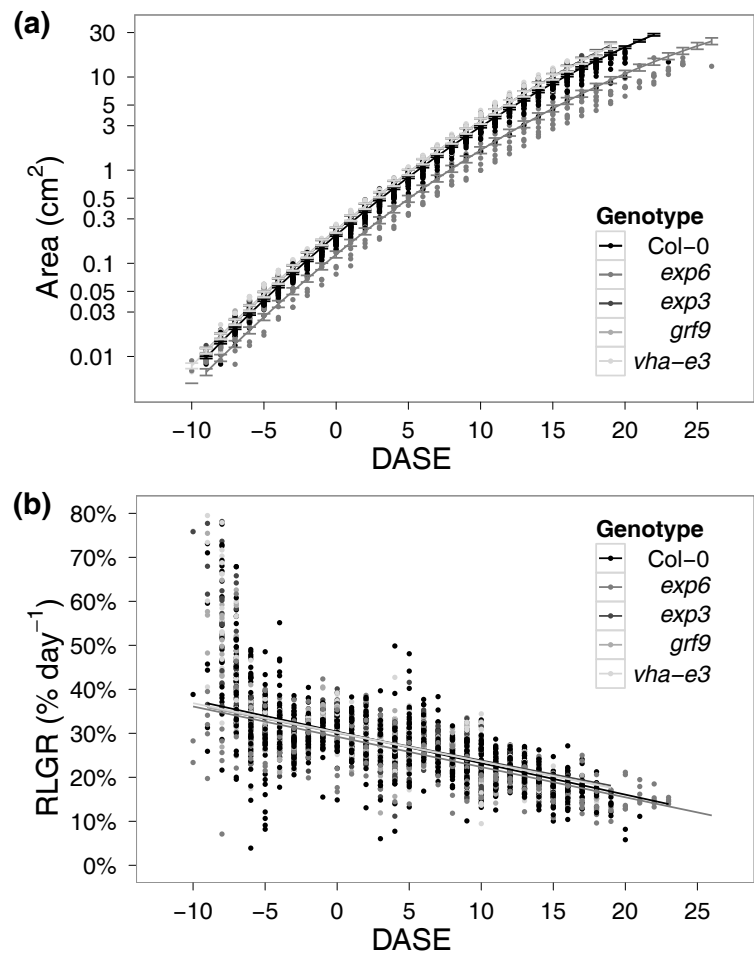


Figure 4.3: Area and RGR modeling for *exp3*, *exp6*, *grf9* and *vha-e3* mutants. As Figure 4.2. (a) Total rosette area. (b) Relative growth rate.

4. Identification of leaf growth-related tonoplast protein genes

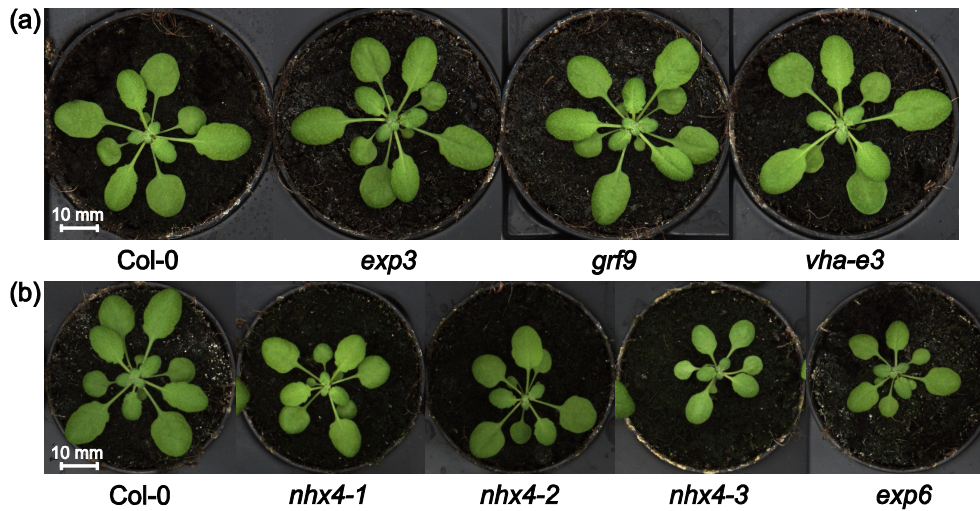


Figure 4.4: Photos of representative mutant plants at 14 DASE. The photos were selected from the growth image phenotyping database by choosing the plants with the median area from each genotype at 14 DASE. (a) Plants grown in 6 cm pots. (b) Plants grown in 5 cm pots. Here, the difficulty to assess small area difference by the naked human eye is illustrated; in the figure the *exp3*, *grf9* and *vha-e3* plants all have a significantly larger total rosette area than WT Col-0 (115 %, 113 % and 111 % of the WT Col-0 shown, respectively), however the two WT Col-0 plants shown have an insignificant area difference (less than 0.5 %). The significantly smaller *nhx4-1*, *nhx4-2*, *nhx4-3* and *exp6* plants can be easily qualitatively assessed, but for them the area differences are also much more prominent (64 %, 61 %, 36 % and 41 % of the WT Col-0 shown, respectively), however to measure the quantitative differences among the *nhx4* lines an automated aid is necessary.

Table 4.6: Significant plant development effects for genotypes showing significant effects – early developmental stages. Major development timelines for each genotype showing significant as estimated with one-way ANOVA. Data are estimated group means \pm SE. * denotes $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. *n* denotes the number of plants.

Genotype	Germination ^a		Seedling establishment ^b		Early rosette development ^c	
	(days)	<i>n</i>	(days)	<i>n</i>	(days)	<i>n</i>
Col-0	3.13 \pm 0.12	51	10.2 \pm 0.285	50	10.5 \pm 0.126	73
<i>exp3</i>	NA ^d	0	NA ^d	0	10.2 \pm 0.183 **	19
<i>exp6</i>	3.19 \pm 0.297	10	13.7 \pm 0.556 ***	10	11.5 \pm 0.275 ***	9
<i>grf9</i>	NA ^d	0	NA ^d	0	10.3 \pm 0.18 *	20
<i>nhx4-1</i>	3.2 \pm 0.403	5	11.8 \pm 0.725 *	5	11.1 \pm 0.349	5
<i>nhx4-2</i>	3.66 \pm 0.511	3	10.3 \pm 0.907	3	11.4 \pm 0.438	3
<i>nhx4-3</i>	3.33 \pm 0.213	24	13.1 \pm 0.38 ***	24	12.9 \pm 0.175 ***	25
<i>vha-e3</i>	NA ^d	0	NA ^d	0	10.4 \pm 0.18	20

^a BBCH 0.1 to 0.5, ^b BBCH 0.5 to 1.02, ^c BBCH 1.02 to 1.09, ^d NA: no available data.

4. Identification of leaf growth-related tonoplast protein genes

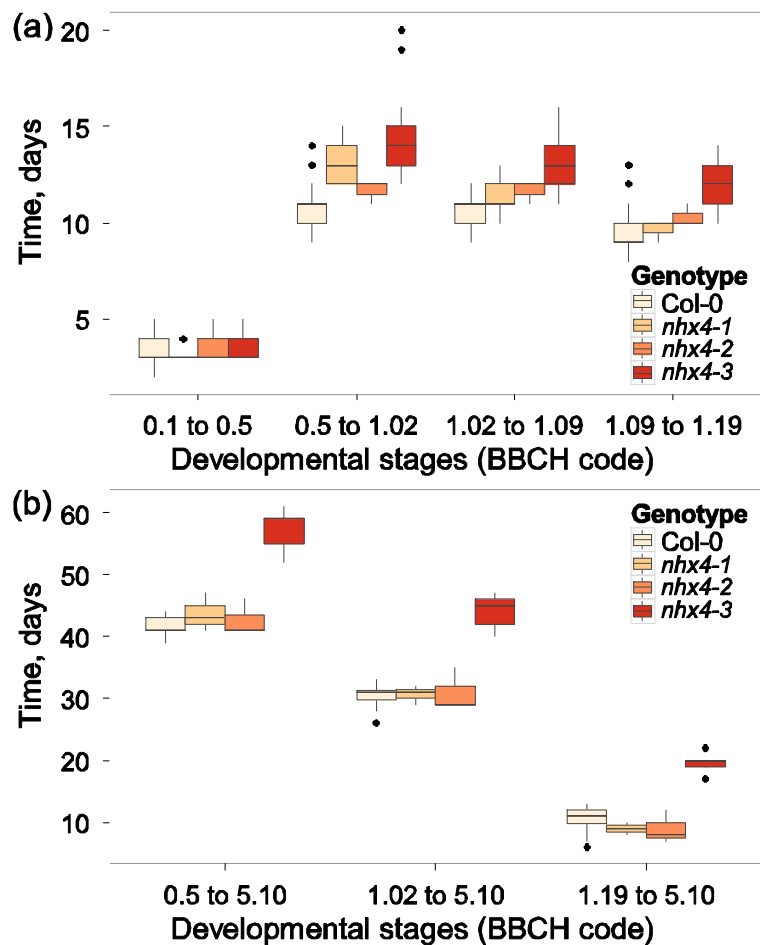
Table 4.7: Significant plant development effects for genotypes showing significant effects – late developmental stages. Major development timelines for each genotype showing significant as estimated with one-way ANOVA. Data are estimated group means \pm SE. * denotes $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. n denotes the number of plants.

Genotype	Mid rosette development ^a	n	Late rosette development ^b	n	Bolting time ^c	
	(days)		(days)		(days)	n
Col-0	9.88 \pm 0.189	42	10.4 \pm 0.462	19	42.3 \pm 0.635	16
<i>exp3</i>	9.89 \pm 0.669	11	NA ^d	0	NA ^d	0
<i>exp6</i>	10.6 \pm 0.45	9	NA ^d	0	NA ^d	0
<i>grf9</i>	10.1 \pm 0.668	11	NA ^d	0	NA ^d	0
<i>nhx4-1</i>	9.66 \pm 0.732	3	9 \pm 1.25	3	42.6 \pm 1.26	3
<i>nhx4-2</i>	10.3 \pm 0.732	3	9 \pm 1.25	3	42.3 \pm 1.26	3
<i>nhx4-3</i>	11.9 \pm 0.333 ***	20	19.6 \pm 1.01 ***	5	57.4 \pm 1.06 ***	5
<i>vha-e3</i>	10.1 \pm 0.66	12	NA ^d	0	NA ^d	0

^a BBCH 1.09 to 1.19, ^b BBCH 1.19 to 5.10, ^c BBCH 0.5 to 5.10, ^d NA: no available data.

Figure 4.5: Plant development comparison for *nhx4* mutants.

Box plots of the time (in days) needed for major developmental stages (in BBCH code) of the *nhx4* knockout mutants and WT Col-0. (a) Early plant development stages, BBCH 0.1 to 0.5 (germination), 0.5 to 1.02 (seedling establishment), 1.02 to 1.09 (early rosette development) and 1.09 to 1.19 (mid rosette development). (b) Bolting-related developmental stages, BBCH 0.5 to 5.10 (germination to bolting), 1.02 to 5.10 (seedling establishment to bolting) and 1.19 to 5.10 (late plant development to bolting).



4. Identification of leaf growth-related tonoplast protein genes

As can be expected, due to gene function redundancy, most lines showed no significant growth phenotype, however we could detect and quantify a strong reduction in total rosette area and RGR (see **Table 4.5**) as well as generally delayed development for *nhx4-3* and *exp6* (see **Table 4.6** and **4.7**) as well as slightly increased area for *exp3*, *grf9* and *vha-e3*; although any significant difference in RGR could not be detected for these lines ($p > 0.05$ for these genotype coefficients in the model).

NHX4 encodes for a tonoplast Na^+/H^+ antiporter and has been described in a recent study (H. Li *et al.*, 2009). No growth phenotype was observed by the authors when they studied knockout mutant plants (lines *nhx4-1* and *nhx4-2*) under standard growth conditions. We obtained the same lines and could, in an initial screen together with a population from a *nhx4-3* hemizygous parental line, see similar phenotypes as for the *nhx4-3* line in homozygous *nhx4-1* and *nhx4-2* plants, although with a less pronounced reduction on total rosette area and RGR (illustrated in **Figure 4.2** and **Figure 4.4**). For all the *nhx4* lines we observed Mendelian co-segregation of the phenotype with the insertion in *NHX4*.

4.3.5 Transcription analysis of rice orthologues of Arabidopsis candidate genes

In order to verify the cross-species importance of the putative growth-related genes we carried through expression analysis for rice orthologues of the putative growth-related genes, as predicted with InParanoid (Östlund *et al.*, 2010). Here we compared expanding regions of young leaves with non-expanding regions in the same as well as mature leaves, and could verify the growth-association for the putative rice orthologue of Arabidopsis *TIP1;1*; Os03g05290. For the putative rice orthologue of Arabidopsis *OCT3*, Os07g37510, we observed a basipetal gradient in all studied leaves, but no significant up-regulation in expanding parts of young leaves in comparison to the basal part of mature leaves. The rice orthologues of Arabidopsis *GRF8* and *GRF9* (Os07g37140 and Os07g28430, respectively) were more highly expressed in expanding regions in comparison to mature regions, which for *GRF9* interestingly coincides with the Arabidopsis gene pattern.

4.4 Discussion

To our knowledge, this is the first report on tonoplast transcriptomics, and one of few approaches in identifying cell expansion-related genes in Arabidopsis; other notable

4. Identification of leaf growth-related tonoplast protein genes

studies include those by Mockaitis and Estelle (2004) and Beemster *et al.* (2005). We propose the tonoplast RT-qPCR platform as a resource for the tonoplast/vacuolar research community which could be used in various experiments; the low technical errors imply that it is a precise and robust tool for tonoplast protein coding gene expression analysis with clear advantages over microarrays (higher precision, higher specific coverage) and high-throughput sequencing methods (lower detection limit, higher specific coverage).

We assessed our gene growth-association ranking (which uses template based scoring for the expression data) by observing the relatively large fraction of cell cycle genes and previously reported growth-related genes with high scores; 45 % of these genes have scores above the selected threshold of 9. Thus we can assume a high enrichment in growth-association among the twelve tonoplast protein genes above this threshold (see **Table 4.3**). So far we have mainly assessed the growth-related patterns of the tonoplast protein gene expression dataset; however we supply the data in supplemental material **Table S4.5** so that researchers with other specific questions could analyze it in other ways.

We include data from four phenotyping batches of 24 knockout mutant lines (18-20 plants per line) analyzed with our growth phenotyping pipeline. We saw an absence of significant growth-related phenotypes for most (17) of these lines, which is expected considering typical gene function redundancy in Arabidopsis (AGI, 2000; Kuromori *et al.*, 2009).

NHX4 has been suggested to transport Na⁺ from the vacuolar lumen into the cytosol, and gene disruption has been shown to confer an increased salt tolerance (H. Li *et al.*, 2009). Since we observed a strong growth inhibition phenotype of the *nhx4-3* line in the first batch of phenotyping with this line we included it in two more batches, where we could confirm the phenotype and obtain similar quantitative values; 14 % smaller total rosette area and 24 % lower RGR in comparison to WT at 0 DASE. These results are in contrast with a previous study (H. Li *et al.*, 2009), where no growth phenotype was observed in two other insertion lines (*nhx4-1* and *nhx4-2*). To confirm our findings, we obtained seeds for *nhx4-1* and *nhx4-2*, and included them in the last measurement batch. For plants showing low or undetectable levels of mRNA (as we used segregating populations of *nhx4-1* and *nhx4-2*), we could observe growth reduction phenotypes as in the *nhx4-3* line, and WT-like behavior for other plants (see **Table 4.5** and **Table 4.6** for details). Although the growth inhibition

4. Identification of leaf growth-related tonoplast protein genes

observed in *nhx4-3* plants is strong, it may be less apparent under different growth conditions; and the subject of the mentioned study was not growth but salt tolerance (H. Li *et al.*, 2009). It is unlikely that a secondary T-DNA insertion or another unrelated genotypic difference in the *nhx4-3* line is causing the stronger phenotype, as we observed that the phenotype co-segregates with the homozygous insertion; wild-type progeny of a hemizygous parent were undistinguishable from the Col-0 WT line. However, it will be necessary to compare these three lines further on the molecular level to see whether truncated transcripts or other secondary effects of gene disruption could explain the observed differences. For a more complete understanding of the role of NHX4, detailed growth phenotyping of the *nhx4* lines under salt stress would be helpful to establish at which stage the effect is most prominent. Then, further experimentation could help in finding out if specific and limited physiological processes are involved in the observed salt tolerance phenotype, or whether it is a general effect. If the suggested directionality of NHX4 is correct and a general effect is assumed, then the most feasible hypothesis we can suggest is that the reduced growth seen in the *nhx4* lines is caused by a reduced capability of maintaining Na⁺ homeostasis in the cytosol. However, if NHX4 transports Na⁺ from the cytosol into the vacuole, a more specific hypothesis can be formulated where the reduced growth in the mutants would be explained by generally increased levels of Na⁺ in the cytosol, leading to cell and organism stress. These possible hypotheses do not yet include any feedback regulation or compensation by alternative transporters, however.

In a recent study of a *vha-E3* mutant line (Dettmer *et al.*, 2010), although not the same line we included in our screen, there was no report on a growth phenotype. Contrastingly, we see a weak but significant phenotype of increased area (14 % increase in comparison to WT at 0 DASE, $p = 0.0010$) which however is not easily detected by the naked eye and therefore possibly overlooked; this is clearly illustrated in **Figure 4.4** (photos of plant comparisons at 14 DASE).

We also observe a reduced area (36 % decrease, $p < 0.001$) and RGR (4.1 % decrease, $p < 0.001$) in the *exp6* mutant line, which could be expected for a knockout line of an expansin. This gene has not been characterized yet, but it has been reported to together with *EXP4* account for 70 % of total expansin transcripts in wood-forming *Arabidopsis* hypocotyls (Gray-Mitsumune *et al.*, 2004). However, we surprisingly see a weak increased area in the *exp3* null mutant (13 % increase, $p =$

4. Identification of leaf growth-related tonoplast protein genes

0.0035). To our knowledge no functional characterization has been performed on a knockout line for this gene, although there has been an indication of increased leaf size in a *CaMV35S::EXP3* line (Kwon *et al.*, 2008). We also see an increased area for the *grf9* line (15 % increase, $p < 0.001$), confirming the tendency reported previously for another knockout mutant line for this gene (Horiguchi, G. Kim, *et al.*, 2005).

Taken together, our results lead us to conclude that the tonoplast is worth a longer look at when studying cell expansion regulation. It is very likely that a systematic survey of all tonoplast associated protein knockout lines using high throughput phenotyping will yield novel cell expansion candidates. This would provide the scientific community with additional data for fruitful hypothesis generation, and help in solving more parts of the complex puzzle of cell and organ growth regulation.

4.5 Materials and methods

4.5.1 General

Standard molecular techniques were performed as described (Sambrook and Russell, 2001). Oligonucleotides were obtained from MWG (Ebersberg, Germany). DNA sequencing was performed by MWG. Unless otherwise indicated, other chemicals were purchased from Roche (Mannheim, Germany), Merck (Darmstadt, Germany), or Sigma (Deisenhofen, Germany). RT-qPCR primers were designed with QuantPrime (Arvidsson *et al.*, 2008).

Data analysis was performed with R (R Development Core Team, 2010) unless otherwise indicated, and the package ggplot2 (Wickham, 2009) was used for generating plots.

4.5.2 Plant cultivation

Arabidopsis thaliana (L.) Heynh. cv Col-0 plants were used in all experiments unless otherwise mentioned.

For the transcriptomics experiments, the plants were grown in growth chambers with an 8-h day length, illuminated by fluorescent light at $120 \mu\text{mol m}^{-2} \text{s}^{-1}$ (50 % intensity during the first and last 30 minutes of the light period), a day/night temperature of 20/16°C and relative humidity of 60/75 %.

4. Identification of leaf growth-related tonoplast protein genes

For genotyping and seed propagation, plants were grown in greenhouse cabins with a 16-h day length, illuminated by sunlight and fluorescent light at $200\text{--}350 \mu\text{mol m}^{-2} \text{s}^{-1}$, day/night temperatures were kept at $22/18 \text{ }^\circ\text{C}$ and relative humidity at 70 %.

For growth phenotyping, plants were grown in growth cabinets with tightly controlled environmental conditions (Percival Scientific Inc., Perry, Iowa, <http://www.percival-scientific.com>) at a PAR of $98 \pm 11 \mu\text{mol m}^{-2} \text{s}^{-1}$, $22 \text{ }^\circ\text{C}$ and 70 % humidity (day), $18 \text{ }^\circ\text{C}$ and 80 % humidity (night), at a 12 hour day/night cycle; except the first ten nights in each growth cycle which were at $6 \text{ }^\circ\text{C}$ and 80 % humidity for stratification. Seeds were directly sown and grown in 5 cm round pots in 54-pot trays (QuickPot 54R, HerkuPlast-Kubern GmbH, Ering am Inn, Germany, <http://www.herkuplast.com>), 6 cm round pots in 35-pot trays (QuickPot 35R) or 8 cm pots in 15-pot trays (QuickPot 15RW). The substrate consisted of two layers, the lower being standard soil (Enheitserde typ P, Einheitserde- und Humuswerke Gebr. Patzer GmbH & Co. KG, Sinnatal-Jossa, Germany, <http://www.einheitserde.de>) mixed with vermiculite (9:1 soil:vermiculite) and the upper being standard soil without vermiculite (circa 1 cm) to ensure good foreground/background separation in the image analysis with smaller plants.

4.5.3 Sampling of leaf material

For the main expression experiment, sampling was performed at 1 to 1.5 hours after lights-on. Leaves of the three developmental stages ($10 \pm 1 \text{ mm}$, $15 \pm 1 \text{ mm}$ and $30 \pm 3 \text{ mm}$) were cut into three pieces (identically long) with a razor blade, called base, mid and tip portion, snap-frozen in liquid nitrogen and stored at $-70 \text{ }^\circ\text{C}$ before homogenization and RNA extraction. 15-25 leaves were sampled and pooled for each of three biological replicates, in order to even out expression differences between individual plants.

For the expression determination of the knockout mutant plants grown in the phenotyping experiments, young leaves were collected after the end of each phenotyping batch.

4.5.4 RNA extraction, quality control and cDNA synthesis

For the main expression experiment, total RNA was isolated, after grinding of plant material in liquid nitrogen, with Trizol reagent (Invitrogen, Karlsruhe, Germany) following the manufacturers specifications. For the expression determination of

4. Identification of leaf growth-related tonoplast protein genes

knockout plants grown in the phenotyping experiments, plant material was ground with metal beads under air cooling (using liquid nitrogen) using a specialized bead mill (Precellys 24TM, Bertin Technologies, Montigny-le-Bretonneux, France), and total RNA was isolated with the NucleoSpin 96-well RNA Tissue Core kit (Macherey-Nagel, Düren, Germany) automated on a purpose-configured pipetting robot (Tecan Freedom Evo with a MCA96 pipetting head) using a vacuum elution protocol.

RNA quality was determined spectrometrically ($A_{260}/A_{280} > 1.8$) using a NanoDrop ND-1000 spectrometer (NanoDrop, Detroit, USA) and by visual inspection of separated bands on agarose gels.

After isolation, genomic DNA was digested using Turbo DNA-free recombinant DNase I (Applied Biosystems Applera, Darmstadt, Germany) following the manufacturers specifications. The level of remaining genomic DNA contamination was measured by diluting the samples to the same concentration as the final cDNA samples ($10 \text{ ng } \mu\text{l}^{-1}$) and performing real-time PCR using primers for a genomic sequence (*UBQ10*; Fw 5'-GGC CTT GTA TAA TCC CTG ATG AAT AAG-3', Rev 5'-AAA GAG ATA ACA GGA ACG GAA ACA TAG T-3'). Samples with consistent quantification cycle (C_q) values below 35 were re-treated with DNase or new RNA extractions were performed.

Two μg of total RNA was used in 20 μl reactions for cDNA synthesis, using RevertAid R-minus cDNA synthesis kit (Fermentas, St. Leon-Rot, Germany), following the manufacturers specifications and using oligo-d(T)₁₈ for priming. In order to reduce bias caused by efficiency differences between reverse transcription reactions, we pooled the cDNA from five reactions before proceeding with the subsequent steps. The cDNA was finally diluted 1:10 in order to reduce the effect of RNA isolation and cDNA synthesis buffer on the subsequent qPCR reactions.

4.5.5 Quantitative real-time PCR, quality control and analysis

All primers used for RT-qPCR are found in the supplementary materials; the primers for the tonoplast protein coding genes and growth-related genes were designed and checked for specificity with QuantPrime (Arvidsson *et al.*, 2008), the primers for reference genes were previously described (Czechowski *et al.*, 2005) as well as those for the cell cycle related genes (Skirycz *et al.*, 2008).

qPCR was carried out in technical triplicates or quadruplicates using 0.5 or 1 μl of diluted cDNA in 5- or 10- μl reactions, 2 or 4 μl of 500 nM primer pairs and 2.5 or 5 μl

4. Identification of leaf growth-related tonoplast protein genes

of 2x Power SYBR Green PCR Master Mix (Applied Biosystems). The following PCR protocol was used on Applied Biosystems 7300 (96-well plates) and 7900HT (384-well plates) real-time PCR systems: 10 min at 95 °C, 15 sec at 95 °C, and 1 min at 60 °C repeated in 50 cycles, followed by melting curve analysis. When testing primer pairs, the PCR products were then separated on a 2 % agarose gel and visualized with ethidium bromide, using 50 bp DNA ladder (Invitrogen) for size determination.

All primer pairs were thoroughly tested on whole-plant cDNA to ensure adequate amplification efficiency and specificity. Also, genomic DNA was used as a template in a separate run to verify to which extent the primers would amplify genomic DNA. Only primer pairs which consistently showed a high efficiency (E over 1.8) and clean amplification, one clear DNA band at the correct size on an agarose gel and no obvious primer-dimers were accepted. New primer pairs were designed and tested until all transcripts could be satisfactory quantified. Where possible, primer pairs not amplifying genomic DNA were chosen.

C_q values for each reaction were calculated using Applied Biosystems SDS software, with baseline set to cycle 3–15 and threshold to 0.2 Rn, recorded from the SYBR Green I dye signal normalized against the ROX dye signal. For each primer pair and cDNA combination we performed 4 (all genes) to 8 (reference genes) technical replications of the qPCR reaction.

Amplification efficiencies were calculated using the LinRegPCR tool (Ramakers *et al.*, 2003), using the best-fit method for 4 to 6 points. This tool uses linear regression on log-values of normalized fluorescence data from individual reactions to calculate E in the equation for PCR kinetics, $N_C = N_0 \times E^C$, which states that the amount of product after C cycles (N_C) is equal to the starting concentration (N_0) times the efficiency (E) to the power C; 100% efficiency would give an efficiency value of 2.

Efficiency values from fitted curves with R^2 values below 0.999 were considered as unreliable; C_q values and efficiencies from such reactions were removed from further calculations. Medians of technically repeated C_q values and efficiencies were calculated and used in further analysis.

The C_q values were normalized using the geometric average of the three most stable reference genes (out of five originally included and tested) as described (Vandesompele *et al.*, 2002) for that cDNA, which were *UBQ10* (AT4G05320), *PP2A-A3* (AT1G13320) and *TIP4;1-like* (AT4G34270) in this study; for these transcripts we used previously described primers (Czechowski *et al.*, 2005); giving ΔC_q values. After

4. Identification of leaf growth-related tonoplast protein genes

normalization, medians of ΔC_q values over biologically replicated cDNAs were calculated (resulting in sample ΔC_q values), and relative expression was determined between samples (resulting in inter-sample $\Delta\Delta C_q$ values).

For template based clustering the scoring table (**Table 4.2**), as described in the results section, was applied on the inter-sample $\Delta\Delta C_q$ values.

4.5.6 Plant genotyping

In order to confirm genome insertions and positions on the chromosome, PCR primers were designed on genomic DNA flanking the estimated insertion point (as given by SALK T-DNA Express, <http://signal.salk.edu/cqi-bin/tdnaexpress> or NASC AtEnsembl, <http://atensembl.arabidopsis.info>) using Primer3 (Rozen and Skaletsky, 2000); primers for all insertion lines are found in the supplementary materials. To confirm the existence of insertions, PCR primers for the different insertion DNA fragments were obtained; for SALK lines (Alonso *et al.*, 2003), we used the LBa1 (5'-TGG TTC ACG TAG TGG GCC ATC G-3') and LBb1 (5'-GCG TGG ACC GCT TGC TGC AAC T-3') primers, for SAIL lines (Sessions *et al.*, 2002) the LB1 (5'-GCC TTT TCA GAA ATG GAT AAA TAG CCT TGC TTC C-3') and LB3 (5'-TAG CAT CTG AAT TTC ATA ACC AAT CTC GAT ACA C-3') primers, for JIC dSpm lines (A.F. Tissier *et al.*, 1999) the dSpm1 primer (5'-CTT ATT TCA GTA AGA GTG TGG GGT TTT GG-3'), for Wisconsin *DsLox* T-DNA lines (Woody *et al.*, 2007) the p745 primer (5'-AAC GTC CGC AAT GTG TTA TTA AGT TGT C-3') and for GABI-Kat lines (Rosso *et al.*, 2003) the o8409 (5'-ATA TTG ACC ATC ATA CTC ATT GC-3') and o3144 (5'-GTG GAT TGA TGT GAT ATC TCC-3') primers were used.

4.5.7 Growth phenotyping

The growth phenotyping protocol and associated data analysis was previously described in detail (see **Chapter 3**). Briefly, images are captured daily and the total rosette leaf area is determined by image analysis software. The growth stage of each plant each day is determined manually from the images and saved (according to an adapted BBCH scale (Boyes *et al.*, 2001)) along with the area data. Next, the complete dataset is evaluated by linear mixed models (modeling area over time), giving an overall quantification of area and relative growth rate (RGR), and ANOVAs for development time, area, compactness and RGR between and at specific growth stages. Also, smoothed fits (loess) of area, compactness, RGR and leaf count as

4. Identification of leaf growth-related tonoplast protein genes

functions of days after sowing (DAS), days after germination (DAG), days after seedling establishment (DASE) and leaf count are generated for an overview of possible phenotypes and suitable predictor variables (see supplementary material figure **S9**). Finally, box plots of development timelines (germination, seedling establishment, early rosette development, late rosette development and bolting time) are presented to provide a simple detection of delayed or accelerated development.

5. General discussion and outlook

5.1 Overview with summary

The aim of this Ph.D. thesis was to identify Arabidopsis tonoplast protein coding genes involved in growth. The specific objectives were to establish a RT-qPCR platform for tonoplast-related transcripts, identify putative growth related tonoplast protein genes using it and finally perform functional characterization including growth phenotyping for these genes. To this end, a growth phenotyping platform was to be developed, which should be as generic as possible for use by other researchers in related projects.

These objectives were met as presented in the previous chapters; two methods were developed, for RT-qPCR primer design and growth phenotyping, respectively. The functional characterization of the putative growth-related tonoplast genes was initiated, but nevertheless leaves several questions open for future research.

In the following sections the main findings and outlook on future work are summarized and discussed.

5.2 QuantPrime – a tool for improved RT-qPCR platform design

In high-throughput expression analyses using RT-qPCR, primer design is a time-consuming part of the experimental design phase, and a critical point in order to obtain high quality results (Udvardi *et al.*, 2008). QuantPrime (see **Chapter 2**) was developed to simplify and speed up this task, making design of high quality primers possible for any molecular biologist with a minimum of experience with qPCR. The main features are batch primer design and *in silico* specificity testing, automated exon-border design to avoid amplification of genomic DNA; all packaged in a simple and efficient user interface. These features are still not matched by any freely available software that we know of. Some of the features (notably automatic exon-junction primer design) which were novel in QuantPrime later appeared in the NCBI tool Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Primer-BLAST is the only software (at the time of writing) that provides a similar richness of features as QuantPrime, however it lacks a batch mode and only provides extended features for RefSeq records, whereas QuantPrime supports many additional genome annotations. Further comparisons with other available software are discussed in more detail in **Chapter 2**.

5. General discussion and outlook

We tested the usefulness of the software by designing primer pairs for *Arabidopsis*, *Chlamydomonas reinhardtii* and barley, and we got high quality primers for more than 95 % of the detectable transcripts (see **Chapter 2**). Since publishing QuantPrime, we have created several platforms for various groups of genes for several species, most notably *Arabidopsis*, rice, *Physcomitrella patens* and *Selaginella moellendorffii*, with high primer specificity and quality rates (over 90 %). In a local collaboration a tomato transcription factor RT-qPCR platform was designed covering 1088 genes. In this platform we observed high quality results for over 95 % of the primer pairs (for detectable transcripts), even though an incomplete expressed sequence tag-based transcript annotation was used at the time of design.

At the time of writing (Oct. 2010), there are 1180 registered QuantPrime users coming from more than 60 countries (1.8 new user registrations day⁻¹, on average). There has been an average of 25 site visits day⁻¹, 140 new target transcripts day⁻¹ (for which 3 500 primer pairs day⁻¹ were tested for specificity), when calculating averages over the whole active period. In total, 2.3 million primers for 91 380 transcripts have been designed. We are currently recording 32 visits day⁻¹ and 2.2 new user registrations day⁻¹ (averages for the last six months), tendency rising. At the time of publication there were 336 annotations for 295 species in the database; currently there are 368 annotations for 299 species. We continuously receive requests to add more species and annotations.

We have received many requests for feature additions to QuantPrime. The most notable suggestions include:

- the possibility of designing primers for sequences provided by the user.
- testing specificity of primers supplied by the user.
- more flexibility in selecting databases for specificity tests – only cDNA or genomic DNA or both and/or the possibility for the user to concurrently check designed primer pairs for specificity in several species (where mixtures or contamination is an issue like in studies of symbiotic or parasitic organisms).
- comprehensive and explicated implementation of parameter relaxation, to aid the user when it is not possible to find a suitable primer pair due to limiting parameters.

We are planning to assess which of these features would be beneficial for a maximum of users, and implement them accordingly.

On the technical side we see the trend in conforming free bioinformatics software to open standards to improve the usability and interoperability; i.e. adopting to the rules of the semantic web (Antezana *et al.*, 2009). Thus, we are considering how parts of or the complete QuantPrime user interface could be made available over SOAP (Simple Object Access Protocol) (<http://www.w3.org/TR/soap12-part1/>) and REST (Representational State Transfer) (Fielding, 2000) APIs (application programming interfaces). In order to increase interoperability with qPCR analysis software, as well as to simplify MIQE (Minimum Information for Publication of Quantitative real-time PCR Experiments) (Bustin *et al.*, 2009) standard compliance for QuantPrime users, we are planning to add primer export possibilities in the standardized RDML format (Lefever *et al.*, 2009).

Last but not least, there are continuously new genomes sequenced, and meta-data for existing genome annotations improve all the time. So far we have implemented new annotations into the public QuantPrime server in a case-by-case manner (typically by user request), which has guaranteed a high quality of the service, due to manual curation and careful testing. Also, we can ensure high computing performance and availability of the service since sequence data and all the related metadata are stored and indexed locally. However, as sequence data increase exponentially and meta-data grow more and more complex, we might have to adopt a different strategy, e.g. set up automated data updates from major genome centers.

5.3 Plant growth phenotyping

The system for growth phenotyping was developed for mid- to high-throughput screens (hundreds to thousands of Arabidopsis plants), for which it should be able to detect minor growth-related phenotypes, otherwise not easily detectable by the naked human eye. Additionally, it should provide quantitative data for the observed phenotypes and relate them to specific developmental stages. The idea behind this is to provide a standardized screening setup that can be used for informative screens by various local researchers with little experience in growth phenotyping, providing as high data quality and informative basic analyses as possible. We believe that our growth phenotyping pipeline fulfills these purposes well, with certain limitations. Most notably, noise levels in the growth data could be reduced further by providing a

higher control of environmental factors. For example, additional automation for plant handling (conveyor track system for the trays) and watering (automated weighting and irrigation) would simplify and standardize the daily work with the system to a higher degree.

However, we have shown that the system can give data of high quality when strict growth and handling protocols are being followed. We established good practices for phenotyping *Arabidopsis thaliana* ecotype Col-0 plants under optimal growth conditions (see details in **Chapter 3**). In order to study genotype or treatment effects under stress conditions, or when working with other *Arabidopsis* ecotypes with different day length optima, rosette compactness/leaf overlap, leaf angles or higher anthocyanin levels, it will be necessary to recalibrate these protocols. Then it is possible to get data which are informative for the researcher (i.e. the experimental conditions are of physiological significance), of high technical quality (i.e. the image analysis is robust enough for possible color variations) and biologically reproducible (i.e. plant handling is standardized enough).

As shown in **Chapter 3**, we included the *sex4-3* mutant (Niittylä *et al.*, 2006) as a control in five plant batches spread over one calendar year, used two different pot types and performed technical replicates. Thus we were able to assess the biological and technical reproducibility of the system, as well as show the informative properties of the system under standard conditions. In **Chapter 4** we present the results of the screen on putatively growth-related genes, again demonstrating the kind of informative data that can be produced with the system. Additionally, we included genotypes from other local projects in these batches (in total we screened 68 genotypes; 44 plant lines additional to those presented in **Chapters 3** and **4**), getting reproducible data between batches (data to be published elsewhere).

The most important improvement in the data analysis is the 'normalization' of the data by using days after seedling establishment (DASE) or leaf count as bases for comparisons. Such comparisons have much lower variation between individual plants within genotypes than the DAS (days after sowing) or DAG (days after germination) measures, making it possible to detect weak phenotypes otherwise not detectable in a direct comparison by the naked human eye (see **Figure 4.4**). We also found it necessary to include environmental variables in the analysis. This is illustrated in a recent cross-lab comparison of typical growth phenotypic variables (sixth leaf area, total rosette area, epidermal cell density), where significant variation between labs

was reported, although effort was made to standardize growth conditions and measurement procedures (Massonnet *et al.*, 2010). Our results suggest that even the small variations (CV = 11 %) in PAR, measured in our growth chamber used in the phenotyping screens, can be modeled and has a clearly significant effect on total rosette area and RGR.

Other major improvements of the phenotyping system consist in the annotation and analysis modules, which are tightly integrated with the data produced from the image analysis. Thus, an inexperienced user can quickly learn how to specify development stages for the plants using the web-based interface. The automated analysis gives a set of plots for qualitative analysis and spreadsheet tables with quantitative data on the genotypes included in the measurement. Although these modules were developed to function with the data structure of our image capture and analysis setup, the data preprocessing and annotation module can be easily modified for data produced with other systems, e.g. the PHENOPSIS (Granier *et al.*, 2006) or GROWSCREEN/FLUORO (Walter *et al.*, 2007; Jansen *et al.*, 2009) systems.

For future projects, we are currently establishing screening protocols for plants grown in 8-cm pots (in 15-pot trays) in order to allow area analysis over the whole plant life time (no overlap between plants). This would also allow the usage of development stages that are based on percentages of the final rosette size (e.g. the BBCH 3 range), however also reduce the throughput due to the lower plant density. Also, we are planning to establish screening protocols for other commonly used ecotypes, especially C24, Landsberg *erecta* and Wassilewskija, to provide a good basis for phenotyping of lines with such genotypic background. We are working on adding more options to the annotation module, so that further phenotypes of interest can be easily annotated and tracked. Most notably, there is strong interest in the possibility of tracking the size of specific leaves, which certainly broadens the scope of analyses possible. However, it is not trivial to accomplish robust automated image analysis for the plant segmentation necessary in such analysis.

The advances in non-destructive growth analyses have been significant in the last decade (Leister *et al.*, 1999; Granier *et al.*, 2006; Walter *et al.*, 2007; Rajendran *et al.*, 2009; Jansen *et al.*, 2009), which has as well been described in a recent review (Berger *et al.*, 2010), and more and more quantitative data are published. In order to increase the informative impact in the general scientific community, we would however suggest to accompany publications containing results from such systems

with complete original data (individual plant identifiers, recorded areas, RGR, compactness/fill factor, stockiness etc.) in addition to detailed description of all environmental factors relevant for the experiment, using well-defined terminologies, as has been recently pointed out (Massonnet *et al.*, 2010; Berger *et al.*, 2010). This would greatly improve the possibility for meta-analyses to unravel complex genotype × environmental interactions, and thus provide valuable information for a broad scientific community.

5.4 Identification of growth-related tonoplast protein genes in *Arabidopsis thaliana*

As described in detail in **Chapter 4**, we used the RT-qPCR primer design and specificity evaluation tool (QuantPrime) discussed above, to create an expression analysis platform for tonoplast protein coding genes and genes known to be involved in leaf growth regulation. It covers 117 transcripts for tonoplast protein coding genes and 26 other growth-related genes, and due to the high quality of the platform (stringent specificity and efficiency controls were enforced for the primers) it can be used as a resource for various studies on tonoplast biology; where the commercially available microarrays offer lower quality and specificity for many of these transcripts. In studying expression patterns of growing *Arabidopsis* leaves using this platform, we could associate 12 tonoplast protein coding genes with cell expansion. The highest scoring genes include a putative organic cation transporter (*OCT3*) recently reported to be stress-regulated but without known specificity (Küfner and Koch, 2008), a putative Na⁺/H⁺ antiporter (*NHX4*) recently characterized and suggested to transport Na⁺ from the vacuole into the cytosol (H. Li *et al.*, 2009) as well as a putative nitrate transporter (*NTP2*) which has not been functionally characterized so far. Among the previously known growth-regulating genes, several expansins turned up with high scores (*EXP3*, -5, -6, -10, -16 and -B1), which we consider to be a good indication that the scoring method is valid, taken together with the high scores of genes known to be associated with active cell division, such as *CYCB1;2* and *CDKB1;1* (Donnelly *et al.*, 1999; Boudolf *et al.*, 2004). The role of the growth-regulating factors is more complicated, since both positive and negative regulators of cell division and expansion have been described in for this family (J.H. Kim *et al.*, 2003; Horiguchi, G. Kim, *et al.*, 2005; Horiguchi, Ferjani, *et al.*, 2005; Byung Ha Lee *et al.*, 2009). Among these, the members *GRF2*, -4, -6 and -9 have high scores.

The growth phenotyping pipeline described above was then used to screen knockout mutants for the highest-scoring tonoplast protein coding and other growth-related genes. In this screen we could identify a strong growth inhibition by the disruption of the *NHX4* and the *EXP6* genes (on RGR as well as area) and slight growth increase in lines with disrupted *GRF9*, *EXP3* or *VHA-E3* genes. Previously, no detailed growth phenotyping has been carried out on these genotypes, however two other knockout mutants for the *NHX4* gene (*nhx4-1* and *nhx4-2*) were characterized in another study (H. Li *et al.*, 2009) where they were reported to show no growth phenotype under normal nutrition – however, they did so when grown in hydroponics. We acquired seeds for *nhx4-1* and *nhx4-2* lines and in an initial phenotyping experiment we could see a significant growth inhibition in homozygous knockout plants, however not as prominent as in the *nhx4-3* line which we initially studied. Repetitions of growth phenotyping on these lines are planned to confirm these findings.

The only working hypothesis regarding the action of NHX4 on growth that we can assume is forcibly generic due to the limited functional characterization of this gene. We would suggest, in case that it is transporting Na⁺ from the vacuole into the cytosol as previously suggested (H. Li *et al.*, 2009), that the cytosolic homeostasis of Na⁺ is transiently or constitutively disturbed at critical development phases, raising the stress level and inhibiting normal growth. However, such a generic hypothesis is difficult to prove in a complex multicellular organism considering feedback loop regulation. Also interorgan transport and compartmentalization makes it difficult to detect increased ion accumulation unless high temporal and spatial resolution is used for sampling. Hence, it would be of importance to first confirm the direction and ion transport specificity of NHX4, e.g. using a patch-clamp technique on either *NHX4*-overexpressing Arabidopsis vacuoles or on *Xenopus* oocytes injected with *NHX4* mRNA. Subsequently, constitutive and inducible overexpression plant lines for *NHX4* would be helpful tools for further functional characterization of this gene.

6. References

- AGI** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408, 796-815.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C. and Ecker, J.R.** (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, 301, 653-657.
- Antezana, E., Kuiper, M. and Mironov, V.** (2009) Biological knowledge management: The emerging role of the Semantic Web technologies. *Brief. Bioinform.*, 10, 392-407.
- Arvidsson, S., Kwaśniewski, M., Riaño-Pachón, D.M. and Mueller-Roeber, B.** (2008) QuantPrime – a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics*, 9, 465.
- Avraham, S., Tung, C., Ilic, K., Jaiswal, P., Kellogg, E.A., McCouch, S., Pujar, A., Reiser, L., Rhee, S.Y., Sachs, M.M., Schaeffer, M., Stein, L., Stevens, P., Vincent, L., Zapata, F. and Ware, D.** (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucl. Acids Res.*, 36, D449-454.
- Balk, P.A. and de Boer, A.D.** (1999) Rapid stalk elongation in tulip (*Tulipa gesneriana* L. cv. Apeldoorn) and the combined action of cold-induced invertase and the water-channel protein γ -TIP. *Planta*, 209, 346-354.
- Beebo, A., Thomas, D., Der, C., Sanchez, L., Leborgne-Castel, N., Marty, F., Schoefs, B. and Bouhidel, K.** (2009) Life with and without AtTIP1;1, an Arabidopsis aquaporin preferentially localized in the apposing tonoplasts of adjacent vacuoles. *Plant Mol. Biol.*, 70, 193-209.
- Beemster, G.T.S., Fiorani, F. and Inzé, D.** (2003) Cell cycle: the key to plant growth control? *Trends Plant Sci.*, 8, 154-158.
- Beemster, G.T., De Veylder, L., Vercruyse, S., West, G., Rombaut, D., Van Hummelen, P., Galichet, A., Gruissem, W., Inzé, D. and Vuylsteke, M.** (2005)

6. References

Genome-wide analysis of gene expression profiles associated with cell cycle transitions in growing organs of *Arabidopsis*. *Plant Physiol.*, 138, 734-743.

Berger, B., Parent, B. and Tester, M. (2010) High-throughput shoot imaging to study drought responses. *J. Exp. Bot.*

Berns, M., Matsubara, S., Wiese, A., Höcker, U., Gilmer, F., Schurr, U. and Walter, A. (2007) Diel leaf growth pattern in *cry1cry2* mutants of *Arabidopsis thaliana* under different light conditions. *Comp. Biochem. Phys. A*, 146, S233.

Blackman, V.H. (1919) The compound interest law and plant growth. *Ann. Bot.*, 33, 353-360.

Boudolf, V., Vlieghe, K., Beemster, G.T., Magyar, Z., Acosta, J.A.T., Maes, S., Van Der Schueren, E., Inze, D. and De Veylder, L. (2004) The plant-specific cyclin-dependent kinase CDKB1;1 and transcription factor E2Fa-DPa control the balance of mitotically dividing and endoreduplicating cells in *Arabidopsis*. *Plant Cell*, 16, 2683-2692.

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

Brumfield, R.T. (1942) Cell growth and division in living root meristems. *Am. J. Bot.*, 29, 533-543.

Bustin, S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.*, 25, 169-193.

Bustin, S.A. (2010) Why the need for qPCR publication guidelines? – The case for MIQE. *Methods*, 50, 217-226.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. and Wittwer, C.T. (2009) The MIQE Guidelines: Minimum Information for publication of Quantitative real-time PCR Experiments. *Clin. Chem.*, 55, 611-622.

Bustin, S.A., Benes, V., Nolan, T. and Pfaffl, M.W. (2005) Quantitative real-time RT-PCR - a perspective. *J. Mol. Endocrinol.*, 34, 597-601.

Carter, C., Pan, S., Zouhar, J., Avila, E.L., Girke, T. and Raikhel, N.V. (2004) The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *Plant Cell*, 16, 3285-3303.

Chambers, Freeny and Heiberger (1992) Chapter 5: Analysis of variance: Designed

6. References

experiments. In: *Statistical models in S*. Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books and Software, 145-190.

Chaumont, F., Barrieu, F., Herman, E.M. and Chrispeels, M.J. (1998) Characterization of a maize tonoplast aquaporin expressed in zones of cell division and elongation. *Plant Physiol.*, 117, 1143-1152.

Choi, D., Cho, H. and Lee, Y. (2006) Expansins: expanding importance in plant growth and development. *Physiol. Plant.*, 126, 511-518.

Coleman, J., Blake-Kalff, M. and Davies, E. (1997) Detoxification of xenobiotics by plants: chemical modification and vacuolar compartmentation. *Trends Plant Sci.*, 2, 144-151.

Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.*, 6, 850-861.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. and Scheible, W. (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.*, 139, 5-17.

De, D.N. (2000) *Plant cell vacuoles: an introduction*. Collingwood, Australia: CSIRO Publishing.

Dettmer, J., Liu, T. and Schumacher, K. (2010) Functional analysis of Arabidopsis V-ATPase subunit VHA-E isoforms. *Eur. J. Cell Biol.*, 89, 152-156.

Donnelly, P.M., Bonetta, D., Tsukaya, H., Dengler, R.E. and Dengler, N.G. (1999) Cell cycling and cell enlargement in developing leaves of Arabidopsis. *Dev. Biol.*, 215, 407-419.

Endler, A., Meyer, S., Schelbert, S., Schneider, T., Weschke, W., Peters, S.W., Keller, F., Baginsky, S., Martinoia, E. and Schmidt, U.G. (2006) Identification of a vacuolar sucrose transporter in barley and Arabidopsis mesophyll cells by a tonoplast proteomic approach. *Plant Physiol.*, 141, 196-207.

Fielding, R.T. (2000) Architectural styles and the design of network-based software architectures.

Frigerio, L., Hinz, G. and Robinson, D.G. (2008) Multiple vacuoles in plant cells: rule or exception? *Traffic*, 9, 1564-1570.

Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S.J., Dauzat, M., Hamard, P., Thioux, J., Rolland, G., Bouchier-Combaud, S., Lebaudy, A., Muller, B., Simonneau, T. and Tardieu, F. (2006) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis*

6. References

thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.*, 169, 623-635.

Gray-Mitsumune, M., Mellerowicz, E.J., Abe, H., Schrader, J., Winzell, A., Sterky, F., Blomqvist, K., McQueen-Mason, S., Teeri, T.T. and Sundberg, B. (2004) Expansins abundant in secondary xylem belong to subgroup A of the α -Expansin gene family. *Plant Physiol.*, 135, 1552-1564.

Hamant, O. and Traas, J. (2010) The mechanics behind plant development. *New Phytol.*, 185, 369-385.

Harashima, H. and Schnittger, A. (2010) The integration of cell division, growth and differentiation. *Curr. Opin. Plant Biol.*, 13, 66-74.

Heazlewood, J.L., Verboom, R.E., Tonti-Filippini, J., Small, I. and Millar, A.H. (2007) SUBA: the Arabidopsis Subcellular Database. *Nucl. Acids Res.*, 35, D213-218.

Herman, E.M., Li, X., Su, R.T., Larsen, P., Hsu, H. and Sze, H. (1994) Vacuolar-type H⁺-ATPases are associated with the endoplasmic reticulum and provacuoles of root tip cells. *Plant Physiol.*, 106, 1313-1324.

Higuchi, R., Fockler, C., Dollinger, G. and Watson, R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Nat. Biotech.*, 11, 1026-1030.

Horiguchi, G., Ferjani, A., Fujikura, U. and Tsukaya, H. (2005) Coordination of cell proliferation and cell expansion in the control of leaf size in *Arabidopsis thaliana*. *J. Plant Res.*, 119, 37-42.

Horiguchi, G., Kim, G. and Tsukaya, H. (2005) The transcription factor *AtGRF5* and the transcription coactivator *AN3* regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J.*, 43, 68-78.

Hothorn, T., Bretz, F. and Westfall, P. (2008) Simultaneous inference in general parametric models. *Biometrical J.*, 50, 346-363.

Hu, Y., Poh, H.M. and Chua, N. (2006) The Arabidopsis *ARGOS-LIKE* gene regulates cell expansion during organ growth. *Plant J.*, 47, 1-9.

Hu, Y., Xie, Q. and Chua, N. (2003) The Arabidopsis auxin-inducible gene *ARGOS* controls lateral organ size. *Plant Cell*, 15, 1951-1961.

Hunt, R., Causton, D.R., Shipley, B. and Askew, A.P. (2002) A modern tool for classical plant growth analysis. *Ann. Bot.*, 90, 485-488.

Hüsken, D., Steudle, E. and Zimmermann, U. (1978) Pressure probe technique for

6. References

measuring water relations of cells in higher plants. *Plant Physiol.*, 61, 158-163.

Jansen, M., Gilmer, F., Biskup, B., Nagel, K.A., Rascher, U., Fischbach, A., Briem, S., Dreissen, G., Tittmann, S., Braun, S., De Jaeger, I., Metzloff, M., Schurr, U., Scharr, H. and Walter, A. (2009) Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Funct. Plant Biol.*, 36, 902-914.

Jaquinod, M., Villiers, F., Kieffer-Jaquinod, S., Hugouvieux, V., Bruley, C., Garin, J. and Bourguignon, J. (2007) A proteomics dissection of *Arabidopsis thaliana* vacuoles isolated from cell culture. *Mol. Cell. Proteomics*, 6, 394-412.

Kim, J.H., Choi, D. and Kende, H. (2003) The *AtGRF* family of putative transcription factors is involved in leaf and cotyledon growth in *Arabidopsis*. *Plant J.*, 36, 94-104.

Küfner, I. and Koch, W. (2008) Stress regulated members of the plant organic cation transporter family are localized to the vacuolar membrane. *BMC Res. Notes*, 1, 43.

Kuromori, T., Takahashi, S., Kondou, Y., Shinozaki, K. and Matsui, M. (2009) Phenome analysis in plant species using loss-of-function and gain-of-function mutants. *Plant Cell Physiol.*, 50, 1215-1231.

Kuromori, T., Wada, T., Kamiya, A., Yuguchi, M., Yokouchi, T., Imura, Y., Takabe, H., Sakurai, T., Akiyama, K., Hirayama, T., Okada, K. and Shinozaki, K. (2006) A trial of phenome analysis using 4000 *Ds*-insertional mutants in gene-coding regions of *Arabidopsis*. *Plant J.*, 47, 640-651.

Kwon, Y.R., Lee, H.J., Kim, K., Hong, S., Lee, S. and Lee, H. (2008) Ectopic expression of *Expansin3* or *Expansinβ1* causes enhanced hormone and salt stress sensitivity in *Arabidopsis*. *Biotechnol. Lett.*, 30, 1281-1288.

Lancashire, P.D., Bleiholder, H., Van Den Boom, T., Langeluddeke, P., Stauss, R., Weber, E. and Witzemberger, A. (1991) A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.*, 119, 561-601.

Lee, B.H., Ko, J., Lee, S., Lee, Y., Pak, J. and Kim, J.H. (2009) The *Arabidopsis GRF-INTERACTING FACTOR* gene family performs an overlapping function in determining organ size as well as multiple developmental properties. *Plant Physiol.*, 151, 655-668.

Lefever, S., Hellemans, J., Pattyn, F., Przybylski, D.R., Taylor, C., Geurts, R., Untergasser, A., Vandesompele, J. and on behalf of the RDML consortium

6. References

- (2009) RDML: structured language and reporting guidelines for real-time quantitative PCR data. *Nucl. Acids Res.*, 37, 2065-2069.
- Leister, D., Varotto, C., Pesaresi, P., Niwergall, A. and Salamini, F.** (1999) Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis. *Plant Physiol. Bioch.*, 37, 671-678.
- Li, H., Liu, H., Gao, X. and Zhang, H.** (2009) Knock-out of *Arabidopsis AtNHX4* gene enhances tolerance to salt stress. *Biochem. Bioph. Res. Co.*, 382, 637-641.
- Lin, W., Peng, Y., Li, G., Arora, R., Tang, Z., Su, W. and Cai, W.** (2007) Isolation and functional characterization of *PgTIP1*, a hormone-autotrophic cells-specific tonoplast aquaporin in ginseng. *J. Exp. Bot.*, 58, 947-956.
- Livak, K.J. and Schmittgen, T.D.** (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 25, 402-408.
- Ludevid, D., Hofte, H., Himmelblau, E. and Chrispeels, M.J.** (1992) The expression pattern of the tonoplast intrinsic protein γ -TIP in *Arabidopsis thaliana* is correlated with cell enlargement. *Plant Physiol.*, 100, 1633-1639.
- Ma, S., Quist, T.M., Ulanov, A., Joly, R. and Bohnert, H.J.** (2004) Loss of *TIP1;1* aquaporin in *Arabidopsis* leads to cell and plant death. *Plant J.*, 40, 845-859.
- Mahner, M. and Kary, M.** (1997) What exactly are genomes, genotypes and phenotypes? And what about phenomes? *J. Theor. Biol.*, 186, 55-63.
- Mar, J., Kimura, Y., Schroder, K., Irvine, K., Hayashizaki, Y., Suzuki, H., Hume, D. and Quackenbush, J.** (2009) Data-driven normalization strategies for high-throughput quantitative RT-PCR. *BMC Bioinformatics*, 10, 110.
- Marty, F.** (1999) Plant vacuoles. *Plant Cell*, 11, 587-600.
- Massonnet, C., Vile, D., Fabre, J., Hannah, M.A., Caldana, C., Lisec, J., Beemster, G.T., Meyer, R.C., Messerli, G., Gronlund, J.T., Perkovic, J., Wigmore, E., May, S., Bevan, M.W., Meyer, C., Rubio-Diaz, S., Weigel, D., Micol, J.L., Buchanan-Wollaston, V., Fiorani, F., Walsh, S., Rinn, B., Gruitsem, W., Hilson, P., Hennig, L., Willmitzer, L. and Granier, C.** (2010) Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three *Arabidopsis* accessions cultivated in ten laboratories. *Plant Physiol.*, 152, 2142-2157.
- Matsushita, A., Furumoto, T., Ishida, S. and Takahashi, Y.** (2007) AGF1, an AT-Hook protein, is necessary for the negative feedback of *AtGA3ox1* encoding GA 3-Oxidase. *Plant Physiol.*, 143, 1152-1162.
- Michaels, S.D. and Amasino, R.M.** (1999) *FLOWERING LOCUS C* encodes a novel

6. References

- MADS domain protein that acts as a repressor of flowering. *Plant Cell*, 11, 949-956.
- Mizukami, Y. and Fischer, R.L.** (2000) Plant organ size control: *AINTEGUMENTA* regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. USA*, 97, 942-947.
- Mockaitis, K. and Estelle, M.** (2004) Integrating transcriptional controls for plant cell expansion. *Genome Biol.*, 5, 245-245.
- Mullis, K.B., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H.** (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.*, 51 Pt 1, 263-273.
- Niittylä, T., Comparot-Moss, S., Lue, W., Messerli, G., Trevisan, M., Seymour, M.D.J., Gatehouse, J.A., Villadsen, D., Smith, S.M., Chen, J., Zeeman, S.C. and Smith, A.M.** (2006) Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. *J. Biol. Chem.*, 281, 11815-11818.
- Nolan, T., Hands, R.E. and Bustin, S.A.** (2006) Quantification of mRNA using real-time RT-PCR. *Nat. Protocols*, 1, 1559-1582.
- Okubo-Kurihara, E., Sano, T., Higaki, T., Kutsuna, N. and Hasezawa, S.** (2009) Acceleration of vacuolar regeneration and cell growth by overexpression of an aquaporin *NtTIP1;1* in tobacco BY-2 cells. *Plant Cell Physiol.*, 50, 151-160.
- Ormenese, S., de Almeida Engler, J., De Groot, R., De Veylder, L., Inzé, D. and Jacquard, A.** (2004) Analysis of the spatial expression pattern of seven kip related proteins (KRPs) in the shoot apex of *Arabidopsis thaliana*. *Ann. Bot.*, 93, 575-580.
- Östlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D.N., Roopra, S., Frings, O. and Sonnhammer, E.L.L.** (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucl. Acids Res.*, 38, D196-D203.
- Pfaffl, M.W.** (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucl. Acids Res.*, 29, e45.
- Pfaffl, M.W.** (2010) The ongoing evolution of qPCR. *Methods*, 50, 215-216.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Development Core Team** (2009) *nlme: Linear and nonlinear mixed effects models*. <http://lib.stat.cmu.edu/R/CRAN/web/packages/nlme/index.html>.
- Poorter, H.** (2002) *Plant growth and carbon economy*. In: *Encyclopedia of life sciences*. Chichester, UK: John Wiley & Sons Ltd.
- del Pozo, J.C., Diaz-Trivino, S., Cisneros, N. and Gutierrez, C.** (2006) The balance between cell division and endoreplication depends on E2FC-DPB,

6. References

transcription factors regulated by the Ubiquitin-SCFSKP2A pathway in Arabidopsis. *Plant Cell*, 18, 2224-2235.

R Development Core Team (2010) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rajendran, K., Tester, M. and Roy, S.J. (2009) Quantifying the three main components of salinity tolerance in cereals. *Plant Cell Environ.*, 32, 237-249.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L. and Moorman, A.F.M. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.*, 339, 62-66.

Reisen, D., Leborgne-Castel, N., Özalp, C., Chaumont, F. and Marty, F. (2003) Expression of a cauliflower tonoplast aquaporin tagged with GFP in tobacco suspension cells correlates with an increase in cell size. *Plant Mol. Biol.*, 52, 387-400.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K. and Weisshaar, B. (2003) An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.*, 53, 247-259.

Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, 132, 365-386.

Saiki, R., Scharf, S., Faloona, F., Mullis, K.B., Horn, G., Erlich, H. and Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230, 1350-1354.

Sambrook, J. and Russell, D.W. (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sampedro, J. and Cosgrove, D. (2005) The expansin superfamily. *Genome Biol.*, 6, 242.

Sang, F. and Ren, J. (2006) Capillary electrophoresis of double-stranded DNA fragments using a new fluorescence intercalating dye EvaGreen. *J. Sep. Sci.*, 29, 1275-1280.

Schüssler, M.D., Alexandersson, E., Bienert, G.P., Kichey, T., Laursen, K.H., Johanson, U., Kjellbom, P., Schjoerring, J.K. and Jahn, T.P. (2008) The effects of the loss of *TIP1;1* and *TIP1;2* aquaporins in *Arabidopsis thaliana*. *Plant J.*, 56, 756-767.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J.,

6. References

- Law, M. and Goff, S.A.** (2002) A high-throughput Arabidopsis reverse genetics system. *Plant Cell*, 14, 2985-2994.
- Shimaoka, T., Ohnishi, M., Sazuka, T., Mitsunashi, N., Hara-Nishimura, I., Shimazaki, K., Maeshima, M., Yokota, A., Tomizawa, K. and Mimura, T.** (2004) Isolation of intact vacuoles and proteomic analysis of tonoplast from suspension-cultured cells of *Arabidopsis thaliana*. *Plant Cell Physiol.*, 45, 672-683.
- Skirycz, A., Radziejowski, A., Busch, W., Hannah, M.A., Czeszejko, J., Kwaśniewski, M., Zanor, M., Lohmann, J.U., De Veylder, L., Witt, I. and Mueller-Roeber, B.** (2008) The DOF transcription factor *OBP1* is involved in cell cycle regulation in *Arabidopsis thaliana*. *Plant J.*, 56, 779-792.
- Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P. and Huala, E.** (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucl. Acids Res.*, 36, D1009-1014.
- Szponarski, W., Sommerer, N., Boyer, J., Rossignol, M. and Gibrat, R.** (2004) Large-scale characterization of integral proteins from Arabidopsis vacuolar membrane by two-dimensional liquid chromatography. *Proteomics*, 4, 397-406.
- Taiz, L.** (1992) The plant vacuole. *J. Exp. Biol.*, 172, 113-122.
- Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G. and Jones, J.D.G.** (1999) Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. *Plant Cell*, 11, 1841-1852.
- Tsukaya, H.** (2002) Interpretation of mutants in leaf morphology: genetic evidence for a compensatory system in leaf morphogenesis that provides a new link between cell and organismal theories. *Int. Rev. Cytol*, 217, 1-39.
- Tsukaya, H.** (2008) Controlling size in multicellular organs: focus on the leaf. *PLoS Biol.*, 6, e174.
- Udvardi, M.K., Czechowski, T. and Scheible, W.** (2008) Eleven golden rules of quantitative RT-PCR. *Plant Cell*, 20, 1736-1737.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F.** (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, 3, research0034.
- Varotto, C., Pesaresi, P., Maiwald, D., Kurth, J., Salamini, F. and Leister, D.**

6. References

- (2000) Identification of photosynthetic mutants of *Arabidopsis* by automatic screening for altered effective quantum yield of photosystem 2. *Photosynthetica*, 38, 497-504.
- Walter, A., Scharr, H., Gilmer, F., Zierer, R., Nagel, K.A., Ernst, M., Wiese, A., Virnich, O., Christ, M.M., Uhlig, B., Jünger, S. and Schurr, U.** (2007) Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytol.*, 174, 447-455.
- Wickham, H.** (2009) *ggplot2: elegant graphics for data analysis*. New York, USA: Springer.
- Wiese, A., Christ, M.M., Virnich, O., Schurr, U. and Walter, A.** (2007) Spatio-temporal leaf growth patterns of *Arabidopsis thaliana* and evidence for sugar control of the diel leaf growth cycle. *New Phytol.*, 174, 752-761.
- Wittwer, C.T., Ririe, K.M., Andrew, R.V., David, D.A., Gundry, R.A. and Balis, U.J.** (1997) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *BioTechniques*, 22, 176-181.
- Woody, S., Austin-Phillips, S., Amasino, R. and Krysan, P.** (2007) The WiscDsLox T-DNA collection: an *Arabidopsis* community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J. Plant Res.*, 120, 157-165.
- Wudick, M.M., Luu, D. and Maurel, C.** (2009) A look inside: localization patterns and functions of intracellular plant aquaporins. *New Phytol.*, 184, 289-302.
- Yin, X., Goudriaan, J., Lantinga, E.A., Vos, J. and Spiertz, H.J.** (2003) A flexible sigmoid function of determinate growth. *Ann. Bot.*, 91, 361-371.
- Zeeman, S.C., Northrop, F., Smith, A.M. and Rees, T.A.** (1998) A starch-accumulating mutant of *Arabidopsis thaliana* deficient in a chloroplastic starch-hydrolysing enzyme. *Plant J.*, 15, 357-365.

Supplementary material

All the supplemental material for **Chapters 2 to 4** are included on the CD-ROM attached to this thesis booklet and listed below. The files which could be formatted for printing are included in the following pages.

Table S2.1: List of customizable parameters in QuantPrime.

A comprehensive list of all parameters that can be customized in QuantPrime, with parameter ranges and default values.

File S2.1: Examples of primer pairs with gel images.

Examples of primer pairs for different species with images of agarose gel separations of their PCR amplification products.

Table S2.2: Comparison of QuantPrime with other primer design software.

A comparison table including QuantPrime and other commonly used primer design software.

Table S3.1: Pre-processed phenotype data (CD-ROM only).

This Excel file contains the pre-processed phenotype data for all WT, *sex4-3* and *grf9* plants from the five experiments. There are two sheets; the first one contains 'day-based' data, with phenotype values retrieved from the primary data for each day. The second sheet contains 'developmental stage-based' data, i.e. phenotype values retrieved from the primary data for each developmental stage. The values were collected from either the first day of each developmental stage, or a mean of the values was calculated for the days around that first day, see the 'Phenotype' column in the sheet.

File S3.1: Pre-processed phenotype data (CD-ROM only).

The file contains an R data frame with the complete dataset (like **Table S3.1**). Using the file requires a recent version of R (available from <http://www.r-project.org>).

File S3.2: Script for linear mixed-effects model analysis (CD-ROM only).

This file is an R script for analyzing total rosette area data with the described linear mixed-effects model. Running the script requires a recent version of R (available from

<http://www.r-project.org>) and the additional libraries 'nlme', 'multcomp' and 'ggplot2' (freely available from CRAN – The Comprehensive R Archive Network, <http://cran.r-project.org>). The file can also be viewed with a standard text editor.

Table S4.1: Genes included in the expression study (CD-ROM only).

This table contains a list of all the growth-related, cell cycle and tonoplast protein coding genes included in the expression study, with references to proteomics, localization, function characterization, regulation and expression studies for many of the tonoplast protein coding genes.

Table S4.2: Primers used for the RT-qPCR platform.

This table contains a list of all RT-qPCR primers used in the expression study.

Figure S4.1: Diel elongation profiles.

A figure with diel elongation profiles for *Arabidopsis thaliana* leaf #11 under sampling growth conditions and a short description of the method used.

Table S4.3: Complete scoring table.

This table contains the scores for all genes in the expression study (i.e. an extensive version of **Table 4.3**), applied according to the scoring template (**Table 4.2**).

Table S4.4: Gene expression table.

This table contains the gene expression values obtained in the expression study, as biological medians of reference gene normalized C_q values (i.e. ΔC_q values). The reference genes used for normalization are marked in bold; the two reference genes not used for normalization due to their higher variance are grayed out. Note that the lower the ΔC_q value, the higher the expression and vice versa, as well as that ΔC_q values are on a log₂-scale and they are only directly comparable within the same amplicon and gene across samples, not across amplicons or genes.

Figure S4.2: Expression value plots.

This figure displays the expression values for all genes and samples with bar plots, with the expression values transformed to $20-\Delta C_q$ for more intuitive visual

interpretation (higher values means higher expression and vice versa), with the SD of the biological replicates as error bars.

Table S4.5: Knockout line reference table.

This table contains genotype references (collection/ordering line IDs), loci, insertion positions and insertion verification primer data for the gene insertion disruption mutants obtained for the candidate growth-related genes.

File S4.1: Preprocessed phenotype data (CD-ROM only).

This file contains one Excel sheet and a R data file, both containing the pre-processed data obtained from the image analysis and annotation of the growth phenotyping experiments, used in statistical analysis and graph plotting.

Figure S4.3: Additional phenotype plots.

This figure contains the comparison plots of total rosette area, compactness, relative growth rate (RGR) and leaf count plotted against days after sowing (DAS), days after germination (DAG), days after seedling establishment (DASE) and leaf count; one page for the three *nhx4* mutants and one page for the *exp3*, *exp6*, *grf9* and *vha-e3* mutants. The plots offer the possibility to assess the suitability of the predictor variables (DAS, DAG, DASE or leaf count) to be used in comparison or modeling of the quantified dependent variables (total rosette area, compactness, RGR and leaf count).

Table S2.1 (1/2)

List of customizable parameters in QuantPrime 1/2
 The first three categories (primer design, Primer3 and specificity testing parameters) can be configured individually for each project. The last category of parameters are global for all projects on the server, thus only configurable for the standalone version of QuantPrime.

Primer design parameters	Value type	Range of meaningful values	Standard value real time PCR	Standard value end point PCR
Limit G/C content at 3' end of the primers	Yes/No	Yes/No	Yes	Yes
Limit G/C content, number of 3' bases to check	bp	3 - 6	5	5
Limit G/C content, maximum number of G/C bases	bp	2 - 4	3	3
Specificity prefiltering	Yes/No	testing)	Yes	Yes
Specificity prefiltering BLAST expectation value	Float	1e-5 - 10 Filtered out (= primer pair may not match any splice variant)	1	1
Specificity prefiltering splice variant options		Required (= primer pair needs to match all)	Neither filtered out nor required	Neither filtered out nor required
Primer3 parameters				
PRIMER_MAX_POLY_X	bp	See Primer3 documentation	3	3
PRIMER_GC_CLAMP	bp	See Primer3 documentation	1	1
PRIMER_TM_SANTALUCIA	0/1	See Primer3 documentation	1	1
PRIMER_SALT_CORRECTIONS	0/1	See Primer3 documentation	1	1
PRIMER_DNTP_CONC	uM	See Primer3 documentation	0.4	0.4
PRIMER_DIVALENT_CONC	mM	See Primer3 documentation	1.5	1.5
PRIMER_DNA_CONC	nM	See Primer3 documentation	50	50
PRIMER_SALT_CONC	mM	See Primer3 documentation	50	50
PRIMER_MIN_SIZE	bp	See Primer3 documentation	20	20
PRIMER_OPT_SIZE	bp	See Primer3 documentation	22	22
PRIMER_MAX_SIZE	bp	See Primer3 documentation	24	24
PRIMER_PRODUCT_SIZE_RANGE	bp ranges	See Primer3 documentation	60-80 80-150	150-1500
PRIMER_PRODUCT_MIN_TM	°C	See Primer3 documentation	75	75
PRIMER_PRODUCT_MAX_TM	°C	See Primer3 documentation	95	95
PRIMER_MIN_GC	%	See Primer3 documentation	45	45
PRIMER_OPT_GC_PERCENT	%	See Primer3 documentation	50	50
PRIMER_MAX_GC	%	See Primer3 documentation	55	55
PRIMER_MIN_TM	°C	See Primer3 documentation	61	61
PRIMER_OPT_TM	°C	See Primer3 documentation	64	64
PRIMER_MAX_TM	°C	See Primer3 documentation	67	67
PRIMER_MAX_DIFF_TM	°C	See Primer3 documentation	2	2
+ other available parameters in Primer3		See Primer3 documentation	not used	not used

Table S2.1 (2/2)

List of customizable parameters in QuantPrime 2/2
 The first three categories (primer design, Primer3 and specificity testing parameters) can be configured individually for each project. The last category of parameters are global for all projects on the server, thus only configurable for the standalone version of QuantPrime.

Specificity testing parameters	Value type	Range of meaningful values	Standard value real time PCR	Standard value end point PCR
BLAST maximum expectation value (all tests)	Float	50 (loose, fast) – 1000 (strict, slow)	200	200
Unspecific hit should be an amplicon (cDNA, gDNA)	Yes/No	Yes (No: meaninglessly strict checking)	Yes	Yes
Maximum unspecific amplicon length (cDNA, gDNA)	bp	1000 – 5000 (depending on extension time and polymerase in PCR reaction)	1500	3500
Minimum 3' match length (cDNA)	bp	1 (very strict) – 5 (loose)	2	2
Minimum 3' match length (gDNA)	bp	1 (very strict) – 5 (loose)	2	2
Consider splice variant amplicons as unspecific (cDNA)	Yes/No	Yes/No	No	No
Single primer unspecific binding testing, minimum 3' match length	bp	0 (very strict) – 4 (loose)	1	1
Single primer unspecific binding testing, minimum match percentage	%	65 (very strict) – 95 (loose)	75	75
Global parameters (configurable for standalone version)	Value type	Range of meaningful values	Public server	
Size of 3' region to favor when designing primer pairs	bp	100 – 1500 (depending on average transcript length, RNA secondary structure and RT used for cDNA synthesis)	1000	
Maximum cycles for automatic primer design	#	1 (low sequence homology between transcripts, slow design or testing) – 100 (high sequence homology between transcripts, fast design and testing)	10	
Primer pairs to design per cycle	#	10 (low sequence homology between transcripts, slow design) – 100 (high sequence homology between transcripts, fast design and testing)	50	
Minimum number of OK primer pairs per transcript	#	1 (faster, less choice) – 20 (slower, more choice)	10	

Examples of primer pairs

In the figures PCR amplification products separated on agarose gels (1 x TBE, 2 % agarose, 35 minutes separation at 100 Volt) are seen. The first and last lane of each gel contain DNA size markers of 200, 150, 100 and 50 bp (top to bottom).

Arabidopsis thaliana (A and B)

No	Transcript	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'	Amplicon size
1	AT2G38620.1	AGAGCTCCTGAAGTTTGGCTTGGT	TTGCCCTCCTAATCATCTCGGCA	96
2	AT3G54180.1	TGGTTCGGAGGCAAGCTCTTTT	TCAGTTGGTGTTCCTAGCAACCTG	85
3	AT4G37490.1	ACCAGACTCTCAAGCATCACAC	GCCAACAGCTTTCACAGTCCAT	68
4	AT1G09960.1	AGCATCGTGTCTGCTGTGT	TGGACGCTGTAAGTTATCGCC	72
5	AT1G19450.1	GGAGTTGGCGTTGTTCAAGTAGT	AAGCCGACGACCTGCTTTATCC	72
6	AT1G53210.1	CCATCATCTTCGCATCCCGCAA	TGTCACCTCCACCACATAGCTCG	74
7	AT2G40610.1	AGCCAGCTTTTCTCAGATCGCT	TGGTACTCTTCGGAAGAGACAGG	68
8	AT3G03220.1	TGCCGGTGCAGTATCGAAGGAT	TCGACTGTAACCCGCATGCTCC	61
9	AT3G15370.1	TGATAGCCCTGCTTCACTTGGAGG	AATCCGGCGTGGTACGGATTG	60
10	AT3G29030.1	ACTTTTGTCTCCTCTGTTGGTGC	CACCTTCTTGCACCGAACCCCTT	138
11	AT3G55500.1	TGCCCGTTTTCTCAAGATCGC	ACTCTTCTACATGCCACCCTG	80
12	AT4G38210.1	GATAGCCACAGCCACGCTTTCT	ATACCCACAAGCGCCTCCAGTA	73

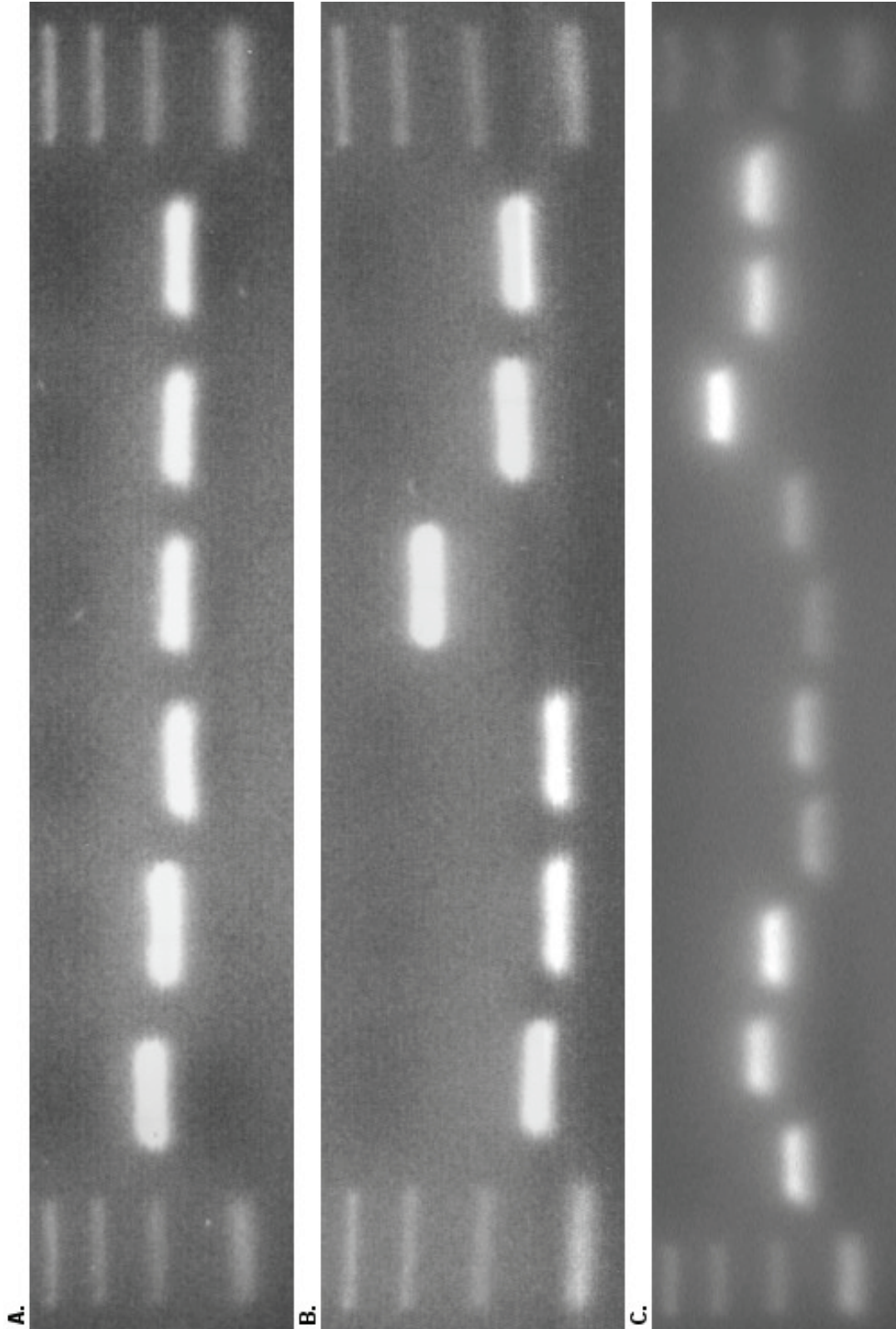
Chlamydomonas reinhardtii (C)

No	Transcript	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'	Amplicon size
1	BIS_CDK1	AGAGGTTAGCCAGGTGCTCTAGC	GTTGCGAATGTGGATGGGCTTG	79
2	BIS_DP	GTTTCTGCCCTTCATCCTGGTG	TCTGGAACGGCGAGTAGAAGTC	116
3	Chlre2_kg.scaffold_21000008	AAATCGTGCCTCGGAAATCCAC	CATGAGACCCGTCAAGTCGAAGTC	95
4	Chlre2_kg.scaffold_4000140	AACCACCTCCCGGCTTGTAAAG	TCGTACACCCAGTCGGGTAACCTG	63
5	Chlre2_kg.scaffold_50000017	CCCAAGCAGATGATGGAGTACCTG	AACTCGCTGCTGAACACCTTCTC	71
6	Chlre2_kg.scaffold_71000018	TCAAATCCGGCTCAACGTACC	GCGTCCAGCTGAGACTTATCATCG	64
7	estExt_fgenes2_kg.C_160082	GTGCAACCAACAACGCACGAAG	GTTGCTCGACCAATCCAATGCC	80
8	estExt_fgenes2_kg.C_560016	ATGTACTACCTGCCACCAAGG	GCTCATGGAGGGCATCTTCATCTG	141
9	estExt_fgenes2_kg.C_90002	CGCCTTCTATCAGAGAACGAGTGG	CCGCCTGAACAGCATGATATGTGG	108
10	estExt_fgenes2_pg.C_10108	TGCCCTACGTCAATCAAGGC	CAGGTAGTGTGCACGTCACATG	110

Examples of primer pairs

<i>Hordeum vulgare</i> (D)		Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'	Amplicon size
No	Transcript			
1	Bmic7	GCATTTGTGCATGTGTGCATCTC	TGGCCTTATTACAAAGCAGCAGGAG	65
2	Bmic8	TTTGTGTACGCGCGCCTGATAG	GTCAGTCGGATGCACAAAGGAC	66
3	Bmic9	CATCAGATTGAGCATCCAGTCCAG	TCGTGCGCTCCATTATGCTCAC	67
4	Bmic10	TCAGCTACGCCAGCTATGAAGC	AGCACCTCTTTGATCAGGTCACTG	76
5	Bmic13	AGTGCAAGCACTTCCCTCTCTGAC	GCGAAATGGTTCGACCCACCAAC	65
6	Bmic14	TGCAGTAAGCGGAGGAGATACG	ACAGGAAGCACCAATTTGTCCTC	65
7	Bmic15	ACCGTTTCATGGTCTCTTTGATGG	AGGAAGACCCGGAAGGCACATTG	72
8	Bmic16	AACCAAGTTGTCAGGGCCGATTC	TGCACAAACAGTTCCTGAAAGCC	68
9	Bmic19	ATCCAAGTGTCTCCACCGTCAAC	ATGCGCATGCTGTTTTCGCAGAC	76
10	Bmic21	CCCGTACATACAAAGTGCAAGCC	GATCGGCAAGAGAGGACCAAC	74
11	Bmic22	ACTTCCACGCTACACCATCTC	ATGCTGGAGTAGCCGAAACATGG	142
12	Bmic23	GCTTTGCAAATGCAGGGAAGC	ACTCCAACGCTTGGACCTTCTG	76

Examples of primer pairs



Examples of primer pairs

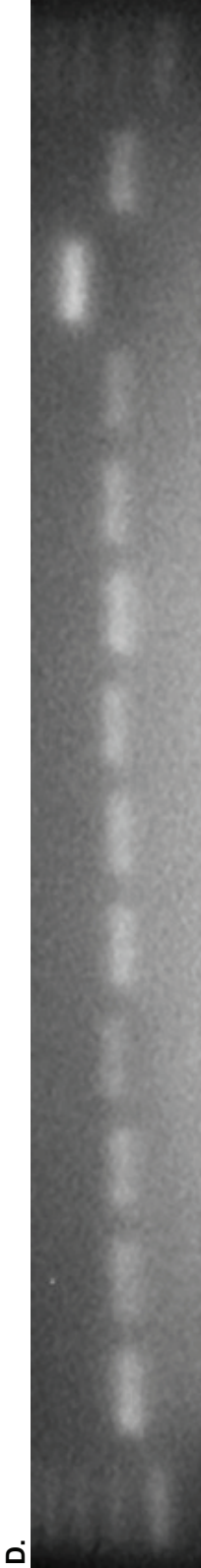


Table S4.2 (1/4)

Group	Gene locus	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
Cell cycle	AT1G20930	TGCCGATATTCTCTGTGCTG	ATCAGAAGCTTTCGTCAAGC
Cell cycle	AT1G76540	TGTTCTTGCCAGTGCTACGG	CACGTCGTCAGGTTAATGGA
Cell cycle	AT2G23430	CGCCGATTCAAATCCGATG	GTATCGACGGGGTACGAAG
Cell cycle	AT2G32710	AGCTTCAACAGGACCACAAG	AAGCTTTGTAGACGATCCCG
Cell cycle	AT2G38620	AGAGCTCCTGAAGTTTTGCTTGGT	TTGCCTCCTAATCATCTCGGCA
Cell cycle	AT3G11520	TCTTCAACACACCACGTAGC	TCTCAGCTCATGGATTGCTC
Cell cycle	AT3G24810	ACAGGAGCATGATAAGTGATTC	TTGTTGCTGTTCTGCGCTAG
Cell cycle	AT3G48750	TGGATGCATCTTTGCCGAGA	CTAGGTCCGTTGGTTTCCAT
Cell cycle	AT3G50630	AGTGAGGAATCGATGAACATG	AGCTTCCTTACCCGTCTCAT
Cell cycle	AT3G54180	TGGTTCGGAGGCAAGCTCTTTT	TCAGTTGGTGTTCCTAGCAACCTG
Cell cycle	AT4G37490	CCTCCATTCACTCTCAACAG	CCTCGCAGCTGTGGAATATG
Cell cycle	AT5G06150	CAGCTGTTTACACGGCAAGATGC	TAGCCGGTGTGGAAGTCAATG
Growth-related	AT1G26770	CTCAGTACAGAGCTGGAATCGTCC	CTCTTCTCTGCAAGAACCCCTTC
Growth-related	AT1G65680	TTGGTCTATCGCCGGATCTACG	TCCATAACCACAAGCTCCACCG
Growth-related	AT1G69530	GCCCGTATTTCAACGCATCGCT	TCTTCTCACGCACGGCACTCTT
Growth-related	AT2G06200	GGCTCATGTAGAGGCATCAACAA	CCAAGGATGAAGCAATGTCTGA
Growth-related	AT2G20750	GCCATAACGGTGTATCCGCAA	CCTCTGTATTTGCATGCCGTTCC
Growth-related	AT2G22840	CCTTGCCCTCCTAATTCTTTTGA	CATGTTACCACCGGAAAAGCC
Growth-related	AT2G28950	GGAAACTGCTCACGCCACTTTCT	ACCATAACCACAAGCTCCTCCC
Growth-related	AT2G36400	CCTCATTACCAACCTGCTTGGTAT	ACCATCCGTTCTCCTGCATCT
Growth-related	AT2G37640	TCTCGCCATGCCTATGTTCTCTCA	ACAAGGTACCCTGCGATAGGAGA
Growth-related	AT2G40610	AGCCAGCTTTTCTTCAGATCGCT	TGGTACTCTTCGGAAAGAGACAGG
Growth-related	AT2G44080	GCAGAACAGTCCAAGGAGGCTA	GACCAACAAGCACAACCATTGA
Growth-related	AT2G45480	AAAAGAGGCGAGTATGTGTGTTGG	CGAAGGTTTGTCTCATCACCTT
Growth-related	AT3G03220	TGCCGGTGCAGTATCGAAGGAT	TCGACTGTAAACCGCATGCTCC
Growth-related	AT3G13960	GACTCCTTCTCCACCAATCCCT	CTGGTTTTCTCCCAAATCCCAT
Growth-related	AT3G15370	TGATAGCCCTGCTTCACTTGGAGG	AATCCGGCGTGGTACGGATTGT
Growth-related	AT3G29030	ACTTTTGTCTCCTGGTGGTGC	CACTTCTCTTGCACCGAACCCCTT
Growth-related	AT3G52910	TCCTCATCACCAACCTTCTTGG	TCTCTTACACCTCCCTGGCTCA
Growth-related	AT3G55500	TGCCCGTTTTCTCAAGATCGC	ACTCTTCTTACATGCCACCCTG
Growth-related	AT3G59900	CGCCTCCGTTTATGCTGCTAT	CATGAAGCAAGAACGACGAGTA
Growth-related	AT4G24150	CCCTGTGCTTTCTACTCTTCCG	AAACCACCTCAGTCTCTGTGG
Growth-related	AT4G35390	TGTATCTAGCCGGAGGTCAAGGAC	ACCGGTCCCGAAGCAATTAACG
Growth-related	AT4G37740	GCCGTTTCTTACCAGATTCTT	CCAGAACCATATCCTCCTGAGCT
Growth-related	AT4G37750	GCATATGATCTTGCTGCACTCAAG	CCGCAGAGAAATTGGTGTGAGTAG
Growth-related	AT4G38210	GATAGCCACAGCCACGCTTTCT	ATACCCACAAGCGCCTCCAGTA
Growth-related	AT5G10140	AGCCACCTTAAATCGGCGGTTG	ACAAACGCTCGCCCTTATCAGC
Growth-related	AT5G53660	GATTCATCCGCATCCTACTCCATA	CATTGCTTGTCTCTCAAGCTCCTT
Reference	AT1G13320	TTCGTGCAGTATCGCTTCTCG	CCGCAGGTAAGAGTTTGAACAT
Reference	AT1G58050	CCATTTACTTTTTGGCGGCT	TCAATGGTAACTGATCCACTCTGATG
Reference	AT3G53090	AACTTTGGCTCTACGACCATTCC	TCAACAGCCAAATTCGTGTCC
Reference	AT4G05320	GGCCTTGATAATCCCTGATGAATAAG	AAAGAGATAACAGGAACGGAAACATAGT
Reference	AT4G34270	GGCACCAACTGTTCTTCGTGA	AAGTCAACTGGATACCCTTTTCGC
Tonoplast	AT1G06470	GCTGTCACCATAGTGGTTGCAGT	CACACCTTTCAGCCACGTAAATTC
Tonoplast	AT1G09960	AGCATCGTGTGTTGCTGCTGTGT	TGGGACGCTGTAAGTTATCGCC
Tonoplast	AT1G11260	GTCCGGTCGTATCTTGCTTGGTT	GAGAGGTACAGTGGCAGACCT
Tonoplast	AT1G12840	GCTTAGCTGTTCTGCTGACCTCTCA	CGAGATGTTCCGACTCGACAAT
Tonoplast	AT1G15690	TTGGAGTTGAGACCCCTCTGGT	TGCTGATATGGCGATCTGAACA
Tonoplast	AT1G16390	GTCTCATTCGCTTGGTTCTTCG	GTTGTGAGTCGGTGAAGACGGT

Table S4.2 (2/4)

Group	Gene locus	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
Tonoplast	AT1G17810	TGTTTCTGCAGCCATCAATGTC	CCTCCGATTAGAGCAGCAAAAGT
Tonoplast	AT1G19450	GGAGTTGGCGTTGTTTCAGGTAGT	AAGCCGACGACCTGCTTTATCC
Tonoplast	AT1G19910	TCGGCTACAGTCTTGTAAATGTTT	CACCCATGTTGCTTGTTTTCA
Tonoplast	AT1G20260	GGTACTGAAAACGCCTGTGTCA	CTTTCCCGAGCCGTTAAATATG
Tonoplast	AT1G20840	GTTAAGCGTGCCTTGGTTGTTG	CCATTGATACCTGAAAACCTGCTGC
Tonoplast	AT1G30400	AAAGCACAACCTCTGTCAAGGTT	TCACGAGGACAGGAATGCTGT
Tonoplast	AT1G31480	CTCATCAGCTCAGCACTCAACG	TTTAGACCCTTTCTCCACTGGC
Tonoplast	AT1G32410	TGGTGGCCAATGCTAAGTGTTA	CACCAAAGAAGAGCAAAGGCA
Tonoplast	AT1G53210	CCATCATCTTCGCATCCCGCAA	TGTCACTCCACCACATAGCTCG
Tonoplast	AT1G54370	CAGGAAGCAGGTTCAAGATGAAG	ACGCGGTGAATGATGTAGTGG
Tonoplast	AT1G58030	TGGCAGTTATCTCGGCTTCAA	CCGAATGGAAAGAACCCTGTT
Tonoplast	AT1G62200	TCTTACTGGGGAAGATACTGGACC	GAGAGTGTCAACAAAGCCATTCCA
Tonoplast	AT1G64200	AAACACGCTTGATGCGAGGTT	AGCGACTTTCGGATCTCTGGA
Tonoplast	AT1G64720	TCCCCTTCCATCCTCAAACT	TTCTCTTTGGAAGAAAGGCTGG
Tonoplast	AT1G73190	TTGTATTGGGAACATGCGGCTCAT	AACAGCAGCAAACAGAGCAAACG
Tonoplast	AT1G75220	AGGTCGTCGGCTTCTGCTACTA	GCAGCTGCAACAATTACAAGGCT
Tonoplast	AT1G75630	TTCTCTTTTTCAGTTCTTTTTCGTG	CCGAGTCATATAATTTTCAACCAAG
Tonoplast	AT1G76030	AATGAAGAGTGCTATCGGCGA	GCATACAGCTGGTTCGACACA
Tonoplast	AT1G77140	GGTTACGTCTGTGCGTGATTACA	CCGTTTCGGAGTCGAGTATGA
Tonoplast	AT1G78900	CCACCATCCAAGTTTACGAGGA	GCTTGTGTGTTTCAAGAACGG
Tonoplast	AT1G79610	AGAGGTTGTAGGTGATAGCCACGA	GACTGTTACCACCTCAAATCCAT
Tonoplast	AT1G80310	ATCATCCCTCTCCAGTTGTACG	TGAAGGCGAACTGAAGACCTTG
Tonoplast	AT2G02020	ACCGTAAATCGAATCTCAAGGTCC	GGGTTAGTGTTCATCTCCTTCG
Tonoplast	AT2G02040	ACGTTACTGGACCATCGCTTG	GAGTTAACGCAGACATCCCGA
Tonoplast	AT2G05170	TCCGGTTTCCAAGCTCACTCT	GGTGACGAGAAAGTTCTTTGCT
Tonoplast	AT2G16510	TCCAGTACTTTCCTTATCGATCAA	AAATATCGGCTCCCGATACTACAT
Tonoplast	AT2G21410	GAGATGCCCTCCGACTTTTTTCC	CTGATATTTGGCCACACCGTATG
Tonoplast	AT2G23150	TGGACCAAACAAGACAGTGTAAGC	GAGCCTTCTTCAACAAAACCAGAG
Tonoplast	AT2G25610	TGGTGAGAATCTCCCCGTACAC	CCGGTAAATGTAATTTCCCAGG
Tonoplast	AT2G25810	CCTTTCTCTTTGTCTTCGCTGG	TTTCCGACTAAACTGTGAGTGGC
Tonoplast	AT2G26690	TGCCGTAGACTACAAAGACGAC	TCTATCCCAAGAATGAGAGCGG
Tonoplast	AT2G26975	ACAGGCCTTGCTTATCTCGTGA	CGACAATGAAAACCTCCACCGTT
Tonoplast	AT2G28520	CCCACATTTGCTAAGACCCAGAT	TGAACTGCTGTGAAAGTTGCTC
Tonoplast	AT2G29410	TTTGTGTGTCTCTTTGTTGGCG	CGAGTTTCGCGGATTAGCTTC
Tonoplast	AT2G32830	GCCGCAAGAAAGTTTACGGTAT	TGATAGACCAGACCCGAGAGAAC
Tonoplast	AT2G34660	TCCAAGTCCAAACTGTTTCGT	ACATATTC AACGCTCCCAGGA
Tonoplast	AT2G36830	GGCTTGTTTACACCGTCTACGCT	GCAATTGTTCCAAGACTCCCG
Tonoplast	AT2G38020	AGGCGAAGTGCTAAGGCAGAT	AACAGCCTCCGACAGTGATGA
Tonoplast	AT2G38170	ACTTCCCCAGAAACAAAATGCC	GCAGTGACGACATTGTTTCATCG
Tonoplast	AT2G38940	TCAAAGGAATTCATGAGTCGCC	GGAACCATGTGATGTAGTGCC
Tonoplast	AT2G41560	GGACTTGTCTCCTCACGAAATGA	GAGACCGAGCCATTACCTGAATT
Tonoplast	AT2G43330	TTCACACAGGTTCCAGGAACATG	TGAGGATGAACTGAATCACAGCG
Tonoplast	AT2G46800	CCACTCATCACCATCATCACGA	TGCTTGCTCTCCATGACCAT
Tonoplast	AT2G47600	ACATCACCTGCAGTAACTCGGTG	TGTGTTTATCAGCCACGGAACTC
Tonoplast	AT2G48020	AGCAGGTTTTCCACAAGACTTG	GCGCAGTGATTACCACCTGAAG
Tonoplast	AT3G01390	CTGCAAGGACCGCAAAAATG	GCAATCTCTTTCTCTGCCTCTTCC
Tonoplast	AT3G03090	AAAGGAAAACCACAGCCAGAG	AGCTGGAAAAGAAAACGGAGGA
Tonoplast	AT3G03720	TCCTATCTCTCTGCATTTGGA	AGTGGTTAAGATGTACGCTGCC
Tonoplast	AT3G05030	CCGACTCTGTATGCACACTACAGG	CAACGCCCTCTCCAAATACAAG

Table S4.2 (3/4)

Group	Gene locus	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
Tonoplast	AT3G06370	TTACGTTGGAATGGACGCTCTC	CCAATCGACTGACCAGGACTGT
Tonoplast	AT3G08560	GCGAATCGACTACTCCACACAAC	CATGGCAGTGACAACGTCACTCT
Tonoplast	AT3G12520	TGTTGGCTCTTATGGTCGGAAT	CGAATAAGCCATCCAAGCCTC
Tonoplast	AT3G13320	TTGAGACGGCGATGCTGTT	ATTTCGATGACCCTTCCTGGAG
Tonoplast	AT3G16240	CTACGCAAAGCTGACGTGG	AACCATGACAAACCGCGATG
Tonoplast	AT3G26520	GATCACTGACAATGGAGCAACCA	AAAGCATGAGCTAAGGCAGCG
Tonoplast	AT3G26590	CTTCTCATTTCGGCATCATGCTT	GACAATTTTCTGCTCCAAACG
Tonoplast	AT3G28710	GCTCCTCATGTACAATCGTTGGT	TCTGGTAAACACAGAGCGACAGA
Tonoplast	AT3G28715	CGGAGAATAAGAGGAACCCATATG	GACCGCAAGATTGTACAGCAGA
Tonoplast	AT3G30390	AGGTGATGTGTTGGCTGAAAAG	TGACCAAACCATCCTTCGAGA
Tonoplast	AT3G42050	GTGGAATGGCTTTGTGCTCAGT	TGCAATTGGAACACCACGAGTA
Tonoplast	AT3G47440	GATTGCCTCTGTTATGGCTTGC	AAATCGGTACGTGCTGTTCCA
Tonoplast	AT3G51490	CGCTTCTCATAAGCGCCTTAA	CTCATGGAGACAAGAATGCAGG
Tonoplast	AT3G51860	AGGATTCTGGAGCGGATTTGC	TCCTCGATCGTGTCCACAACA
Tonoplast	AT3G53720	TTACGACCGGCACCGTCTAAAG	CAATGCCCTTCGTCTAATTCTT
Tonoplast	AT3G54860	CATTCTAAACCGCATGCAAGTG	TTCAGGCCTTCCCACATCAT
Tonoplast	AT3G58730	GGAGTTTGGCTGAGGAAATGGTTC	CGTTTATCGAAATCCCTCTCTGC
Tonoplast	AT3G58810	GGCATCAAGAAGACATGTGGAGA	TTGAACTGGTCTTGGCATCTGA
Tonoplast	AT3G62700	ATCGGAGAACCGGAATTAAC	TCGCGCAAGTTGTATCCTCTG
Tonoplast	AT4G01470	CAGCTTTCGCGGAGTTTTTCT	CATAAGCCATTCCAGAGCCTTG
Tonoplast	AT4G01840	TTGGTTACGGTGACATCGCTC	CGAATCCGAACAAGACGAAAAC
Tonoplast	AT4G02620	ATGATCGCCGATGAGGACACT	TGTCAACATTACCAACTCCAGCC
Tonoplast	AT4G03560	GACCATTTTCTGCGTGCTGTG	AAGCCCTCCAAAGACCTGTACG
Tonoplast	AT4G11150	TCCTCATGGTCTCCACTGCTCT	TGTTCTCGCACACGATTTTCC
Tonoplast	AT4G17340	GTTACCAATGGCGAGAGCGTAC	TCCTTCAATAGCTCCTAAACCGG
Tonoplast	AT4G18160	GTCTGTCTCTCAGTTTTTTCGCTGC	TTCTCCATCTCCTTTCAGTTTG
Tonoplast	AT4G23710	AGAGGTTGCTGAGCACAAAACC	CACTTGTGCTTTCGAGTTTCTT
Tonoplast	AT4G25950	TAAGTGTGAACAAGAAGCCGG	TCTTGCTAGTTTTTGGGTCCTC
Tonoplast	AT4G26710	CCCCTGCTCTTTCTGTATCTT	CGTTGTTCCACACCACGATAAG
Tonoplast	AT4G28770	GAATTGTTGTGGTAGCCATCCAG	CATCAAGCTCTGATTGCCGAG
Tonoplast	AT4G30120	ACAGAGAGCTGGATGCTTAACAGA	AGCAACCCCATATCGACAACCA
Tonoplast	AT4G32530	GCGGCTATAGAAGCTCCTCGTAT	GCTTCGCAAAAGATGACTACTGA
Tonoplast	AT4G34720	CGGAAACAAAACCGCCTATCT	TCTACATCGGAGAGTAACCACCG
Tonoplast	AT4G35300	CTGTGCTTGTGCTATTGCTGC	CCTGCAATAGTTGCGTTATCCC
Tonoplast	AT4G38510	CAGGCTGGTTTGGTTAAGCGT	CATCCTCCTGATGCTCGAGAAG
Tonoplast	AT4G38920	TGCGCTCGTCTTCTCATGTATG	TAGATGCCACTCCCACACCACT
Tonoplast	AT4G39080	TTTCAGTGAGGACCTTGAAGAC	GGATAATCGACCAGAAACCTCG
Tonoplast	AT5G01490	ACCTTTGGCCGAACGTATCAG	AATCCTCCAAGTGTGGACCG
Tonoplast	AT5G03570	TCTCTACTCTTTCAGGCACCGT	CCTTCCGACATAACAACAACCC
Tonoplast	AT5G13550	GTATAATGCTTGTCCCCAGGC	GACCGTAAATGGTGGAAAGGC
Tonoplast	AT5G13740	AGCTACCGCAATGTCTAGCTGTT	GCTAAAAAACCAAGAGCAGG
Tonoplast	AT5G14120	ATTTGGAACCTTCTTGGACGGAT	AGTCCCTGACAACGAGCTCG
Tonoplast	AT5G17010	TCTCAGTTGGACTGGCAATGC	AGCTGTCCACGTATCGGACTTG
Tonoplast	AT5G20650	AACCGCCGGATACGCTGTTTTT	TGGATCATCCGTGGCGGTATCA
Tonoplast	AT5G26340	TCGCCGTGCTTTACAGATTT	AGGCGCGTAGAACATAATCGC
Tonoplast	AT5G27150	TGACCGAGAGGTTGCCCTTAT	ACCGCTCAAGTCGAAAAGCTC
Tonoplast	AT5G39040	CGAATGGAAGCGAGAATGGTT	CAACATTTGCCGCCTCAACT
Tonoplast	AT5G39510	CACGAAACGCTTCATGGAGTAGA	CCTCCTCGTCATGTCTGTCAAGA
Tonoplast	AT5G40890	AAGCTTTGGTTCTGTTCCGGTC	GGTGACATCCCTGAAGCTTGAAT

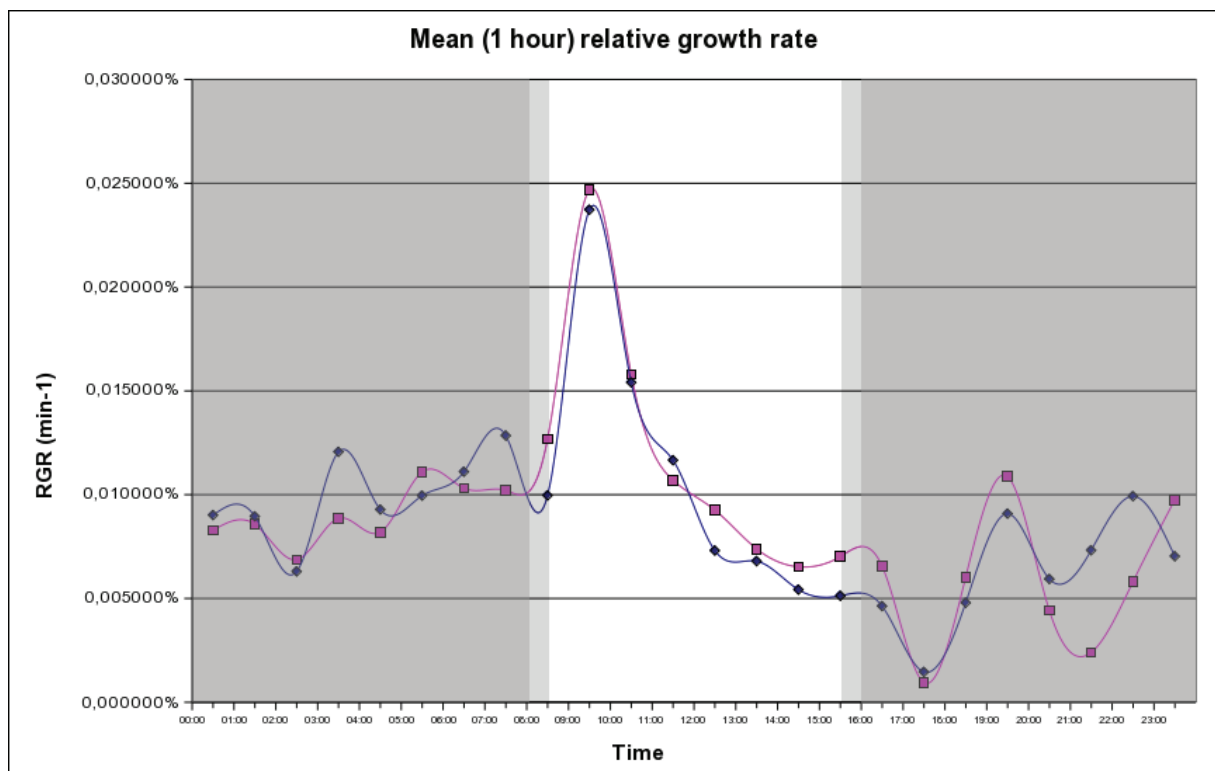
Table S4.2 (4/4)

Group	Gene locus	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
Tonoplast	AT5G43340	CGGACTCCACTTACTCGGAACA	GCTGTAGAAAGCGATGTCGAGG
Tonoplast	AT5G45370	CCGTGAAAGAACGATCAGACCT	TGGTTCCCAAATATCCCTGCT
Tonoplast	AT5G46360	AAGTGCTTGGCGAGAGCATATC	TCTGCTAAACTCAAGCGACCATC
Tonoplast	AT5G46370	CTCTCAGTTCTTAGATGCGG	GTCAAACCTGGAAACCGATCG
Tonoplast	AT5G46860	CAAAGCGGTGTCAATCCAAGTA	ACAGCTTGAAAGTCCCTTGCA
Tonoplast	AT5G47450	TGCTTGATGGTTGGTTGATGTT	TTGCAAGAAAGTGGACACCAAA
Tonoplast	AT5G47560	CCGTGGAACACTACAACATCCAT	CACGCAGAACACCAACGTTATG
Tonoplast	AT5G48410	CCAACGGTTTCGGCTTTATGT	CGCTCGTCCTTAGCTTTGAGATT
Tonoplast	AT5G55290	AACAACCCTAATCTTCGTCGTCG	CCCACATCATCCAACAGCAGA
Tonoplast	AT5G55630	CAGCTGATCTCGATGAAGATG	GCTCGAACTCATCCATTATCC
Tonoplast	AT5G62890	TCAAGATACGCAAGTGCAACGA	ATCAGAAATCGCAACTCCCTGC
Tonoplast	AT5G64870	CAAGATCGATGTAGCAGAGGCAA	TCCAGTCCTTTCTTTGGCACC
Tonoplast	AT5G67330	AGCTTTGGAAGAATCCGAATCG	CCAAAATCTAAGAAACGCCGG

Figure S4.1

Diel leaf elongation profiles

In order to determine the time of the day that would be suitable for harvesting leaf material in respect to capture optimal growth indicators, a diel elongation profile was determined with the use of Rotary Resistance Transducers (RRTs) (Berns *et al.*, 2007) borrowed from ICG-III Phytosphere, Forschungszentrum Jülich, Jülich, Germany. In the figure below there are plots of two typical curves of relative growth rates over a 24 hour period, as recorded from wild type Col-0 plants under the same conditions as used for sampling of leaf material. The relative growth rate shown is calculated using the mean absolute growth rate (in mm per minute) over each hour, then divided by the size of the leaf at the beginning of that hour and presented as % of the leaf size per minute. When comparing these results with the patterns reported previously with another comparable method (Wiese *et al.*, 2007), we see no obvious differences. No effects can be seen of the 30 min 50 % light on/off at the beginning or end of each day cycle, which however could be due to the limited time resolution; when taking means of shorter periods, the noise increases dramatically limiting the reproducibility of the results obtained.



Berns M, Matsubara S, Wiese A, Höcker U, Gilmer F, Schurr U, Walter A. 2007. Diel leaf growth pattern in *cry1cry2* mutants of *Arabidopsis thaliana* under different light conditions. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology* **146**: S233.

Wiese A, Christ MM, Virnich O, Schurr U, Walter A. 2007. Spatio-temporal leaf growth patterns of *Arabidopsis thaliana* and evidence for sugar control of the diel leaf growth cycle. *New Phytol.* **174**: 752-761.

Table S4.3 (1/2)

Transcript identifier	Group reference	Score	Transcript identifier	Group reference	Score
AT4G37490.1	Cell cycle	18	AT5G62890.1	Tonoplast	9
AT2G38620.1	Cell cycle	16	AT2G23150.1	Tonoplast	7
AT3G54180.1	Cell cycle	16	AT4G34720.1	Tonoplast	7
AT5G06150.1	Cell cycle	16	AT1G11260.1	Tonoplast	6
AT1G20930.1	Cell cycle	15	AT1G20840.1	Tonoplast	6
AT3G11520.1	Cell cycle	13	AT1G78900.1	Tonoplast	6
AT1G76540.1	Cell cycle	4	AT2G02020.1	Tonoplast	5
AT2G23430.1	Cell cycle	2	AT4G25950.1	Tonoplast	5
AT2G32710.1	Cell cycle	2	AT1G19450.1	Tonoplast	4
AT3G24810.1	Cell cycle	2	AT1G20260.1	Tonoplast	4
AT3G48750.1	Cell cycle	2	AT1G31480.1	Tonoplast	4
AT3G50630.1	Cell cycle	2	AT1G77140.1	Tonoplast	4
AT2G06200.1	Growth-related	19	AT2G02040.1	Tonoplast	4
AT2G28950.1	Growth-related	17	AT2G05170.1	Tonoplast	4
AT3G29030.1	Growth-related	17	AT2G21410.1	Tonoplast	4
AT4G37750.1	Growth-related	16	AT2G25610.1	Tonoplast	4
AT1G26770.1	Growth-related	15	AT2G28520.1	Tonoplast	4
AT2G20750.1	Growth-related	15	AT2G38020.1	Tonoplast	4
AT3G52910.1	Growth-related	15	AT3G03090.1	Tonoplast	4
AT2G45480.1	Growth-related	13	AT3G30390.1	Tonoplast	4
AT4G37740.1	Growth-related	13	AT3G42050.1	Tonoplast	4
AT2G37640.1	Growth-related	9	AT3G54860.1	Tonoplast	4
AT3G55500.1	Growth-related	9	AT3G58810.1	Tonoplast	4
AT5G53660.1	Growth-related	6	AT4G02620.1	Tonoplast	4
AT1G65680.1	Growth-related	5	AT4G28770.1	Tonoplast	4
AT2G36400.1	Growth-related	5	AT4G30120.1	Tonoplast	4
AT3G03220.1	Growth-related	5	AT4G38510.1	Tonoplast	4
AT4G35390.1	Growth-related	4	AT4G38920.1	Tonoplast	4
AT3G59900.1	Growth-related	3	AT4G39080.1	Tonoplast	4
AT2G22840.1	Growth-related	2	AT5G14120.1	Tonoplast	4
AT2G40610.1	Growth-related	2	AT5G39040.1	Tonoplast	4
AT2G44080.1	Growth-related	2	AT5G40890.1	Tonoplast	4
AT4G24150.1	Growth-related	2	AT5G46860.1	Tonoplast	4
AT4G38210.1	Growth-related	2	AT1G30400.1	Tonoplast	3
AT5G10140.1	Growth-related	2	AT2G41560.1	Tonoplast	3
AT1G69530.1	Growth-related	1	AT1G06470.1	Tonoplast	2
AT3G13960.1	Growth-related	0	AT1G09960.1	Tonoplast	2
AT3G15370.1	Growth-related	0	AT1G15690.1	Tonoplast	2
AT1G16390.1	Tonoplast	20	AT2G16510.1	Tonoplast	2
AT5G47450.1	Tonoplast	18	AT2G25810.1	Tonoplast	2
AT3G16240.1	Tonoplast	17	AT2G26975.1	Tonoplast	2
AT2G36830.1	Tonoplast	15	AT2G29410.1	Tonoplast	2
AT2G26690.1	Tonoplast	14	AT2G32830.1	Tonoplast	2
AT3G06370.1	Tonoplast	14	AT2G34660.1	Tonoplast	2
AT3G51490.1	Tonoplast	12	AT2G38170.1	Tonoplast	2
AT1G64200.1	Tonoplast	9	AT2G43330.1	Tonoplast	2
AT2G48020.1	Tonoplast	9	AT2G46800.1	Tonoplast	2
AT3G26520.1	Tonoplast	9	AT2G47600.1	Tonoplast	2
AT5G13740.1	Tonoplast	9	AT3G01390.1	Tonoplast	2

Table S4.3 (2/2)

Transcript identifier	Group reference	Score
AT3G03720.1	Tonoplast	2
AT3G05030.1	Tonoplast	2
AT3G13320.1	Tonoplast	2
AT3G26590.1	Tonoplast	2
AT3G28710.1	Tonoplast	2
AT3G28715.1	Tonoplast	2
AT3G51860.1	Tonoplast	2
AT3G58730.1	Tonoplast	2
AT3G62700.1	Tonoplast	2
AT4G01840.1	Tonoplast	2
AT4G03560.1	Tonoplast	2
AT4G11150.1	Tonoplast	2
AT4G17340.1	Tonoplast	2
AT4G18160.1	Tonoplast	2
AT4G23710.1	Tonoplast	2
AT4G26710.1	Tonoplast	2
AT4G35300.1	Tonoplast	2
AT5G13550.1	Tonoplast	2
AT5G17010.1	Tonoplast	2
AT5G20650.1	Tonoplast	2
AT5G27150.1	Tonoplast	2
AT5G39510.1	Tonoplast	2
AT5G45370.1	Tonoplast	2
AT5G46360.1	Tonoplast	2
AT5G46370.1	Tonoplast	2
AT5G47560.1	Tonoplast	2
AT5G55290.1	Tonoplast	2
AT5G55630.1	Tonoplast	2
AT5G67330.1	Tonoplast	2
AT1G12840.1	Tonoplast	1
AT1G17810.1	Tonoplast	1
AT1G73190.1	Tonoplast	1
AT2G38940.1	Tonoplast	1
AT3G08560.1	Tonoplast	1
AT3G12520.1	Tonoplast	1
AT3G47440.1	Tonoplast	1
AT4G32530.1	Tonoplast	1
AT5G01490.1	Tonoplast	1
AT5G26340.1	Tonoplast	1
AT5G48410.1	Tonoplast	1
AT5G64870.1	Tonoplast	1
AT3G53720.1	Tonoplast	0
AT4G01470.1	Tonoplast	0
AT5G03570.1	Tonoplast	0
AT5G43340.1	Tonoplast	0

Transcript	Group	10 mm base	10 mm middle	10 mm tip	15 mm base	15 mm middle	15 mm tip	30 mm base	30 mm middle	30 mm tip
AT1G20930.1	Cell cycle	4.0052	4.3710	5.2631	4.6052	5.0400	5.2250	6.4956	7.4079	7.0383
AT1G76540.1	Cell cycle	2.4976	2.7612	3.5293	2.8757	3.0958	3.5518	3.4468	4.1011	3.8601
AT2G23430.1	Cell cycle	1.1960	0.8794	0.4983	1.3189	1.0511	1.1998	1.0632	1.3279	1.2221
AT2G32710.1	Cell cycle	3.7985	3.8197	3.7786	3.8795	3.8712	3.7222	3.3345	3.5837	3.6296
AT2G38620.1	Cell cycle	9.0322	9.5080	10.9120	9.0067	10.9000	10.9903	13.0403	14.8669	14.5714
AT3G11520.1	Cell cycle	5.1344	5.3918	5.7708	5.3846	5.7173	6.1189	8.7227	10.6096	10.1075
AT3G24810.1	Cell cycle	4.5476	4.5998	4.6449	4.8325	4.9313	5.0768	5.2289	5.9716	5.9985
AT3G48750.1	Cell cycle	0.7174	0.7381	1.2777	0.8734	1.0832	1.0974	0.2136	1.0791	0.4994
AT3G50630.1	Cell cycle	2.9011	3.0311	3.4858	2.8264	3.1312	3.5259	3.4881	3.6122	4.2674
AT3G54180.1	Cell cycle	5.3235	5.5806	7.0575	6.0636	6.2442	6.8286	8.7362	9.8279	10.4987
AT4G37490.1	Cell cycle	6.0780	6.6031	8.1704	7.3422	7.7898	8.3066	10.4210	11.3574	10.8881
AT5G06150.1	Cell cycle	4.1465	4.4781	5.2572	4.6155	4.8825	5.4061	8.7544	9.9068	10.7621
AT1G26770.1	Growth-related	1.4584	1.5902	1.7184	1.7783	1.4156	1.6195	5.1297	4.5280	4.6822
AT1G65680.1	Growth-related	14.5981	12.0000	12.6165	13.3821	12.4164	14.5898	14.1779	14.5663	12.6478
AT1G69530.1	Growth-related	-1.2985	-1.2005	-0.6568	-1.2667	-1.1971	-1.7086	-0.9427	0.2398	-0.4230
AT2G06200.1	Growth-related	10.1524	10.5042	11.0689	10.4782	11.1958	12.8574	13.6538	15.7108	14.7393
AT2G20750.1	Growth-related	2.8064	2.7626	2.9260	3.0643	2.9902	3.2125	9.0466	9.3051	9.1735
AT2G22840.1	Growth-related	6.5014	6.8773	7.0087	6.8718	7.1118	6.7706	5.8829	6.0618	5.9591
AT2G28950.1	Growth-related	-1.1050	-1.1648	0.5013	-0.6060	-0.5636	-0.3872	1.5752	1.8335	1.7693
AT2G36400.1	Growth-related	12.4021	14.1956	13.2952	12.9101	15.4119	15.0084	10.5625	10.4216	10.4527
AT2G37640.1	Growth-related	3.7406	3.8469	4.7562	3.5750	3.6619	4.4684	5.3152	6.5287	7.1409
AT2G40610.1	Growth-related	3.4549	2.5790	3.0924	2.0127	1.5272	1.3198	2.9840	3.5183	3.4802
AT2G44080.1	Growth-related	6.4733	6.1306	6.4615	5.9386	5.3803	5.5683	3.3065	3.8555	3.8205
AT2G45480.1	Growth-related	9.4609	9.6077	9.8160	9.4626	9.8517	10.4161	12.0573	12.3156	12.3804
AT3G03220.1	Growth-related	3.9041	4.0563	4.8064	4.2650	4.2638	4.5479	4.9897	5.3683	5.1724
AT3G13960.1	Growth-related	13.0004	13.2065	NA	14.4696	12.5088	NA	NA	13.7339	NA
AT3G15370.1	Growth-related	14.4382	14.2156	NA	14.7632	14.2537	13.2769	13.3073	14.4048	14.8479
AT3G29030.1	Growth-related	-0.5395	-0.4043	0.6453	-0.4048	-0.4057	0.2358	2.6044	2.8008	3.2388
AT3G52910.1	Growth-related	6.4759	6.9579	7.1268	6.8702	7.2067	7.7239	10.5173	11.0378	11.8370
AT3G55500.1	Growth-related	3.5609	3.5318	2.6232	3.3345	2.6570	2.3340	4.9203	5.3355	5.1569
AT3G59900.1	Growth-related	6.2629	7.1524	7.2515	6.4554	6.3633	7.0647	6.6970	8.0126	7.2836
AT4G24150.1	Growth-related	11.1296	10.7633	10.4668	11.4644	11.1685	10.7084	9.6041	10.4175	10.1898

Transcript	Group	10 mm base	10 mm middle	10 mm tip	15 mm base	15 mm middle	15 mm tip	30 mm base	30 mm middle	30 mm tip
AT4G35390.1	Growth-related	8.7766	9.1201	9.7535	9.1395	9.5914	10.3286	9.6806	10.4114	10.1018
AT4G37740.1	Growth-related	5.7475	6.2810	6.9195	6.1460	6.6290	6.8346	8.1850	8.2922	8.0142
AT4G37750.1	Growth-related	7.0749	7.4305	8.3345	7.7355	8.1125	8.5884	10.4128	11.8688	11.5834
AT4G38210.1	Growth-related	6.8847	6.7769	7.2326	7.2464	7.2949	7.4242	7.1727	7.7137	7.6091
AT5G10140.1	Growth-related	5.3671	5.4715	4.4451	5.6451	5.7594	5.9422	5.8838	6.5047	6.1358
AT5G53660.1	Growth-related	11.0579	12.0693	12.4965	12.7032	13.3231	12.7424	10.1836	10.6083	10.9495
AT1G13320.1	Reference	1.6128	1.5946	2.1177	1.8673	1.7982	2.0024	2.0120	2.1781	1.9917
AT1G58050.1	Reference	4.7974	4.8083	4.8373	5.3105	5.2008	5.4551	5.8548	6.0758	5.9258
AT3G53090.1	Reference	4.7380	4.5344	5.3303	4.9723	5.1638	4.9918	4.4856	5.1154	5.1670
AT4G05320.1	Reference	-3.2538	-3.3929	-3.8450	-3.5321	-3.4341	-3.6213	-3.3938	-3.8620	-3.5421
AT4G34270.1	Reference	2.1203	2.1976	2.3480	2.3431	2.1434	2.2929	2.0408	2.1929	2.0505
AT1G06470.1	Tonoplast	3.7490	3.6938	3.9571	3.7171	3.8833	4.1751	3.6558	3.9074	3.8199
AT1G09960.1	Tonoplast	5.9977	5.8927	6.6621	6.2417	6.2896	6.5031	6.1263	6.6912	6.5525
AT1G11260.1	Tonoplast	1.0378	2.6634	4.3291	1.7681	1.8695	2.1344	0.7749	1.1412	0.6580
AT1G12840.1	Tonoplast	3.9068	2.9427	4.5710	3.3455	3.1628	3.7563	3.4124	4.2677	4.0152
AT1G15690.1	Tonoplast	-2.0050	-2.0236	-1.8469	-1.8086	-2.0163	-1.8422	-1.0042	-0.9288	-1.0426
AT1G16390.1	Tonoplast	5.9852	7.2169	9.6249	6.8419	8.1107	9.5178	9.9031	11.1702	11.7060
AT1G17810.1	Tonoplast	13.5600	13.1643	13.4918	13.3022	13.5303	12.5207	12.3429	13.1353	12.2302
AT1G19450.1	Tonoplast	-0.2958	-0.6342	-0.2290	-0.3231	-0.7106	-0.5190	0.7178	0.6907	0.7307
AT1G19910.1	Tonoplast	-1.9507	-2.0215	-2.1382	-1.8122	-1.7720	-1.4918	-2.5786	-2.4052	-2.4425
AT1G20260.1	Tonoplast	0.8311	0.7468	2.4635	0.8260	0.9530	1.0917	0.9760	0.9757	1.0271
AT1G20840.1	Tonoplast	2.2373	2.6599	4.3138	2.6381	3.1678	3.4126	1.6934	2.3060	1.9165
AT1G30400.1	Tonoplast	3.0542	3.1559	4.6733	3.1314	3.7974	3.4353	3.2083	4.3803	3.9165
AT1G31480.1	Tonoplast	4.8538	4.9330	5.9796	5.1662	5.7716	5.5827	4.4316	4.6420	4.5645
AT1G32410.1	Tonoplast	10.0202	10.2321	10.4951	9.8162	10.7822	11.4109	9.6473	10.4928	11.3830
AT1G53210.1	Tonoplast	-0.1006	-0.2835	0.2843	0.0132	-0.2583	-0.1631	-0.9028	-0.7638	-0.7120
AT1G54370.1	Tonoplast	2.2824	2.1326	2.3393	2.4071	2.3902	2.3365	2.0767	2.5158	2.3577
AT1G58030.1	Tonoplast	3.8559	3.6669	4.0521	3.3236	3.7081	3.3414	1.5831	2.2731	1.6302
AT1G62200.1	Tonoplast	4.7117	4.5108	5.5117	4.7105	4.6279	4.6768	3.3226	4.1223	3.5688
AT1G64200.1	Tonoplast	4.9287	5.4004	4.6784	4.8778	5.3235	5.8106	6.7454	6.4124	6.7366
AT1G64720.1	Tonoplast	-0.3452	-1.2188	-0.1011	-0.5818	-0.4304	-0.6767	-0.9432	-0.3141	-0.6925
AT1G73190.1	Tonoplast	13.3215	15.0738	14.2132	13.7617	13.9967	14.2324	12.6692	13.4486	13.3259

Table S4.4 (3/5)

Transcript	Group	10 mm base	10 mm middle	10 mm tip	15 mm base	15 mm middle	15 mm tip	30 mm base	30 mm middle	30 mm tip
AT1G75220.1	Tonoplast	3.1292	3.1250	3.6066	2.9883	3.0923	3.5960	3.3017	3.4408	3.3451
AT1G75630.1	Tonoplast	-1.0216	-0.9125	-1.0139	-0.5014	-0.5136	-0.2538	-1.5504	-1.3103	-1.3411
AT1G76030.1	Tonoplast	1.1787	1.1196	1.4674	1.1703	1.1213	1.3404	0.5290	1.0604	0.7770
AT1G77140.1	Tonoplast	4.7095	4.5614	6.0132	4.8218	5.2215	5.0823	4.5080	4.6354	4.6152
AT1G78900.1	Tonoplast	-0.0219	0.1029	2.4694	0.1479	0.4619	0.5176	0.2339	0.3154	0.3739
AT1G79610.1	Tonoplast	4.9614	5.0883	5.1205	5.1588	5.1017	5.1814	5.1018	4.6648	5.2420
AT1G80310.1	Tonoplast	5.1399	5.5009	5.9938	5.8364	5.7383	5.5918	3.2548	3.5927	3.3225
AT2G02020.1	Tonoplast	11.5704	11.9096	11.0719	12.2527	11.6648	12.0145	12.9422	13.2913	12.8836
AT2G02040.1	Tonoplast	3.2589	3.0828	5.2509	3.4794	3.6188	3.4463	2.6691	2.8439	2.6876
AT2G05170.1	Tonoplast	3.2651	3.2123	5.0571	3.3306	3.5173	3.4119	3.0727	3.6044	3.5512
AT2G16510.1	Tonoplast	0.4664	0.4728	0.3314	0.5242	0.5382	0.8624	0.4939	0.7929	0.7754
AT2G21410.1	Tonoplast	2.2788	2.0270	3.7198	2.2451	2.4755	2.6768	2.1253	2.3845	2.2935
AT2G23150.1	Tonoplast	4.0257	3.8748	3.8491	4.1628	4.1295	4.5388	5.1438	5.0689	4.6692
AT2G25610.1	Tonoplast	0.8911	0.8510	1.9808	0.9806	1.1213	1.3404	0.8243	1.0508	1.0317
AT2G25810.1	Tonoplast	11.4121	10.5407	12.1937	10.2328	10.0939	8.8652	9.8343	9.7403	10.4624
AT2G26690.1	Tonoplast	0.8109	0.7312	2.7636	0.0595	0.4015	1.0792	2.0377	2.5011	2.5257
AT2G26975.1	Tonoplast	0.5693	0.5674	0.9438	1.0434	0.8879	0.9896	0.3298	0.5880	0.3460
AT2G28520.1	Tonoplast	2.9602	2.8028	4.0785	3.0608	3.1960	3.0856	3.1429	3.0338	3.2147
AT2G29410.1	Tonoplast	6.3336	6.0999	6.2246	6.2097	6.0973	6.1130	6.2686	6.9720	6.3932
AT2G32830.1	Tonoplast	13.8798	12.7313	13.2056	12.9710	11.2701	11.2075	10.5653	11.0265	11.3731
AT2G34660.1	Tonoplast	3.2251	3.0397	3.2928	2.9410	3.4423	3.3379	2.9016	3.5938	3.4881
AT2G36830.1	Tonoplast	-1.9880	-1.8290	-1.4759	-1.9509	-1.8720	-1.1999	1.5196	0.5673	1.3144
AT2G38020.1	Tonoplast	3.6187	3.6694	5.2080	4.0581	4.0455	4.1504	3.3469	3.7713	3.7440
AT2G38170.1	Tonoplast	-0.9901	-0.8235	-0.1715	-0.3017	-0.2829	0.2837	-1.3626	-1.0577	-1.5465
AT2G38940.1	Tonoplast	5.1332	5.3187	5.0099	5.6003	5.9497	5.8894	5.0601	5.1963	6.3772
AT2G41560.1	Tonoplast	1.4876	1.4638	3.0535	2.4749	2.3399	2.5906	0.5296	1.6394	1.4272
AT2G43330.1	Tonoplast	3.7080	3.5788	3.8544	3.4124	3.4005	3.4952	2.7597	2.7478	3.3336
AT2G46800.1	Tonoplast	3.8042	4.0816	4.1379	4.4854	4.3124	4.1919	4.3550	4.3503	4.3062
AT2G47600.1	Tonoplast	4.0705	4.1852	4.3512	4.0888	4.0605	4.2610	2.7605	2.8363	2.9621
AT2G48020.1	Tonoplast	1.1818	1.3345	2.1888	1.9507	1.8084	2.1258	2.5893	2.8919	2.4370
AT3G01390.1	Tonoplast	-1.2584	-1.2568	-0.7043	-1.3136	-1.2402	-0.9755	-0.8123	-0.8046	-0.5700
AT3G03090.1	Tonoplast	3.6701	3.7801	5.0668	3.9795	4.0795	4.2876	3.1116	3.4504	3.2295

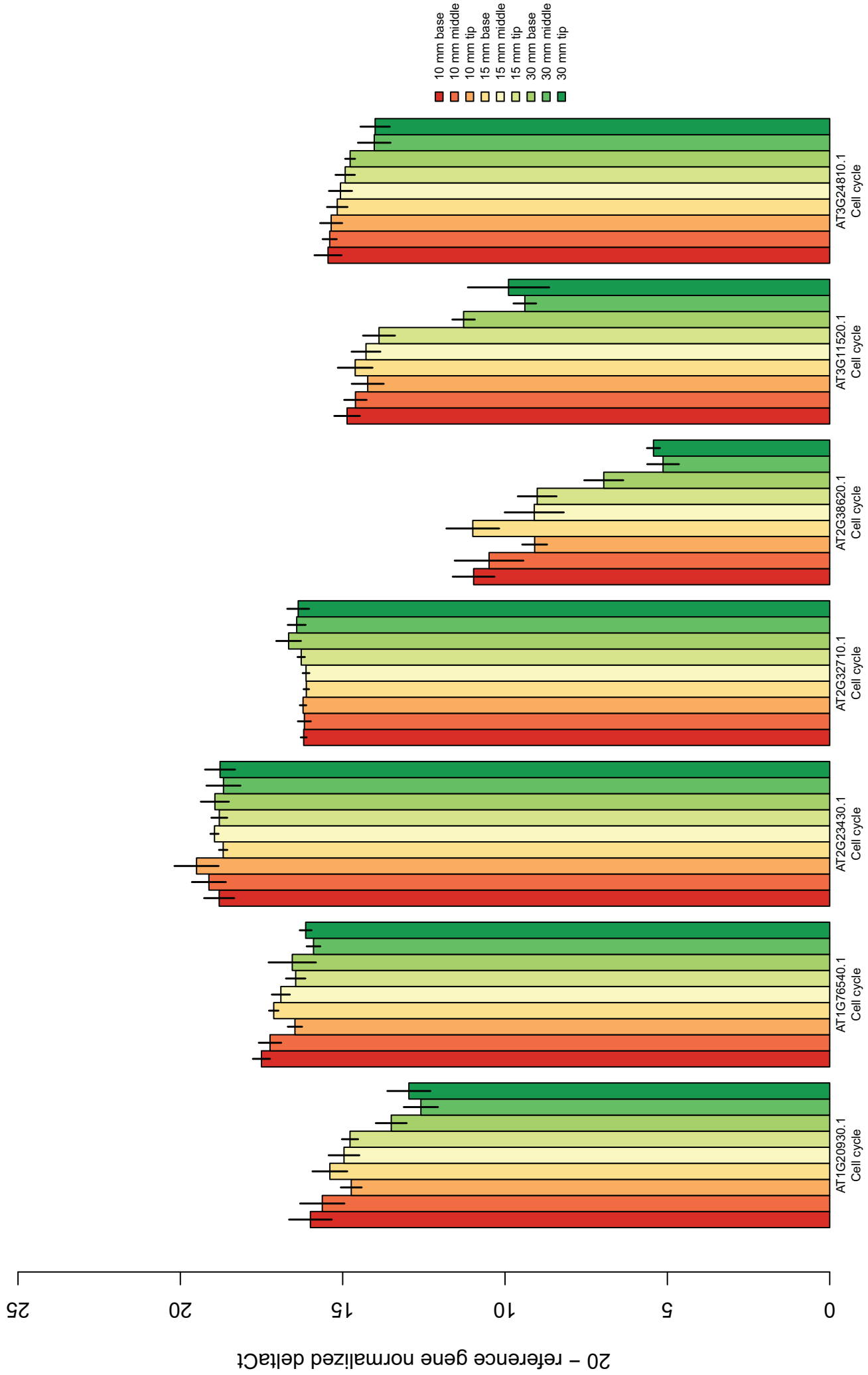
Table S4 (4/5)

Transcript	Group	10 mm base	10 mm middle	10 mm tip	15 mm base	15 mm middle	15 mm tip	30 mm base	30 mm middle	30 mm tip
AT3G03720.1	Tonoplast	3.6240	3.4794	3.9993	3.7245	3.5698	3.6204	2.7923	3.2802	3.1265
AT3G05030.1	Tonoplast	5.7021	5.5562	4.9277	5.1972	4.9524	5.0470	5.0094	4.7040	5.2350
AT3G06370.1	Tonoplast	6.4895	6.6904	8.9707	7.2750	7.7397	8.8118	8.5589	10.1810	10.0991
AT3G08560.1	Tonoplast	12.0758	12.7863	12.1581	12.3029	11.9164	12.3733	13.6436	13.6090	12.6369
AT3G12520.1	Tonoplast	7.0723	7.1100	7.7805	6.5737	6.2478	6.3048	5.4385	6.0311	6.6018
AT3G13320.1	Tonoplast	2.3802	2.3818	2.6114	2.7335	3.0260	3.0488	2.4557	2.3662	2.6169
AT3G16240.1	Tonoplast	-3.4936	-2.8367	-1.8494	-3.2299	-3.0079	-2.3125	0.3341	0.5259	0.3651
AT3G26520.1	Tonoplast	-2.7027	-2.7961	-1.1733	-2.3741	-2.3659	-2.2041	-1.3047	-1.2609	-1.2766
AT3G26590.1	Tonoplast	4.4552	3.8029	4.7524	3.6618	3.5835	3.3526	2.6601	2.8117	2.7187
AT3G28710.1	Tonoplast	0.5039	0.4054	0.5227	0.5527	0.7568	0.8883	0.4993	0.9020	0.6751
AT3G28715.1	Tonoplast	1.5578	1.6354	1.7181	2.1327	1.9998	2.2786	2.5380	2.5173	2.4782
AT3G30390.1	Tonoplast	0.0290	0.1133	1.5324	0.2512	0.3610	0.3153	-1.0312	-0.6908	-0.8089
AT3G42050.1	Tonoplast	0.3108	0.3346	1.8812	0.6783	0.7664	0.8839	0.5140	0.8950	0.9294
AT3G47440.1	Tonoplast	13.4255	12.1163	12.3243	11.9334	11.6387	12.1901	12.8739	12.3746	12.5655
AT3G51490.1	Tonoplast	9.7397	10.4945	10.3724	9.9259	11.2031	11.3981	12.2285	14.8735	14.0520
AT3G51860.1	Tonoplast	7.7583	8.1047	8.9659	8.3087	8.3131	7.8707	2.3189	2.1849	2.0196
AT3G53720.1	Tonoplast	3.4091	2.9161	3.7763	2.5496	2.6359	2.9543	2.2429	3.5986	3.7886
AT3G54860.1	Tonoplast	3.2319	3.5594	4.9950	3.9769	4.1073	4.3385	3.2506	3.8269	3.9173
AT3G58730.1	Tonoplast	-0.0745	-0.1326	0.0962	-0.0304	-0.2483	0.2285	0.0585	0.3668	0.4095
AT3G58810.1	Tonoplast	10.1487	10.8375	11.2350	10.5782	10.4232	10.4189	10.8920	10.9574	11.6507
AT3G62700.1	Tonoplast	2.2149	2.2524	2.8682	2.2008	2.4319	2.3750	1.8141	2.3845	2.2118
AT4G01470.1	Tonoplast	13.1367	13.8549	12.8314	13.3325	13.5696	12.7752	12.8953	14.5763	13.7273
AT4G01840.1	Tonoplast	4.8989	4.8302	4.9875	4.9077	4.8673	4.7360	4.5480	4.7907	4.6944
AT4G02620.1	Tonoplast	0.5613	0.6015	1.5609	0.9012	0.7468	1.1546	0.6664	0.9206	0.7338
AT4G03560.1	Tonoplast	0.8944	0.7999	1.8297	1.1800	1.1349	1.1488	0.0557	0.3449	0.0805
AT4G11150.1	Tonoplast	-1.3223	-1.2997	-1.1277	-1.2711	-1.1623	-1.0333	-1.1137	-0.7239	-1.0009
AT4G17340.1	Tonoplast	1.2179	0.6892	-0.1255	0.1523	-0.5639	-0.5542	1.0981	0.1513	1.1153
AT4G18160.1	Tonoplast	3.7761	3.5689	3.6698	3.8059	3.8805	3.9732	3.3552	3.7831	3.8097
AT4G23710.1	Tonoplast	1.2782	1.3215	1.9625	1.3350	1.4402	1.8413	1.6862	2.2592	2.0929
AT4G25950.1	Tonoplast	8.0222	8.5062	9.3096	8.6320	8.9792	9.6437	8.2090	8.0965	7.8397
AT4G26710.1	Tonoplast	-1.1135	-1.0436	-1.0248	-0.8474	-0.8353	-0.4754	-0.7681	-0.3515	-0.4268
AT4G28770.1	Tonoplast	1.6327	1.8343	2.5200	1.9776	2.0533	2.3732	1.3209	1.6945	1.4664

Table S4.4 (5/5)

Transcript	Group	10 mm base	10 mm middle	10 mm tip	15 mm base	15 mm middle	15 mm tip	30 mm base	30 mm middle	30 mm tip
AT4G30120.1	Tonoplast	12.3953	13.4518	13.4193	13.4717	13.0246	13.4939	11.2461	11.8230	11.7419
AT4G32530.1	Tonoplast	2.2828	1.9456	2.8068	2.0705	2.2542	2.4558	1.7615	2.6207	2.6263
AT4G34720.1	Tonoplast	-1.8011	-1.8101	-1.8273	-1.4529	-1.4988	-1.0822	-0.7933	-0.4301	-0.6116
AT4G35300.1	Tonoplast	3.3028	3.4115	3.4645	3.2173	3.3547	3.7121	3.3919	3.5318	3.4726
AT4G38510.1	Tonoplast	1.1917	1.3529	2.4483	1.5648	1.8887	1.9416	1.8393	2.1663	2.1151
AT4G38920.1	Tonoplast	-0.3976	-0.3963	0.7679	-0.2619	-0.1243	0.2133	-0.8489	-0.5718	-0.7571
AT4G39080.1	Tonoplast	1.0228	0.8091	2.3400	0.9898	1.0483	1.0950	0.7169	1.0330	1.0856
AT5G01490.1	Tonoplast	NA	NA	13.0635	14.1946	NA	13.2638	12.0885	13.1829	12.1464
AT5G03570.1	Tonoplast	NA	NA	14.3416	15.8946	NA	16.0034	13.1634	10.7273	12.8111
AT5G13550.1	Tonoplast	3.8120	3.5802	4.5662	3.7251	3.7360	3.6469	3.1417	3.5636	3.5812
AT5G13740.1	Tonoplast	1.6340	1.2206	1.3306	1.2865	1.4136	1.4163	3.2605	3.6946	3.4701
AT5G14120.1	Tonoplast	0.3734	0.5759	1.9199	0.8583	1.2674	1.9637	1.2698	2.2743	1.7376
AT5G17010.1	Tonoplast	3.1939	2.9634	3.3597	3.1070	3.1801	3.3166	3.1865	3.2876	3.2961
AT5G20650.1	Tonoplast	0.4921	0.5206	0.7782	0.7175	0.8225	0.8625	0.3380	0.6191	0.4744
AT5G26340.1	Tonoplast	7.0529	6.5515	7.1268	5.8783	6.0513	5.1607	1.1682	2.0581	0.9786
AT5G27150.1	Tonoplast	2.7912	2.7961	2.8838	2.0255	2.0159	2.2861	1.6488	1.7913	1.9278
AT5G39040.1	Tonoplast	3.0415	3.0187	4.9537	3.1504	3.2234	3.5086	3.3571	3.6449	3.8672
AT5G39510.1	Tonoplast	1.7188	1.8189	2.0551	1.7768	1.7220	1.8731	1.2999	1.6323	1.6691
AT5G40890.1	Tonoplast	0.9695	0.8041	1.6448	1.1841	1.2350	1.3027	1.7197	1.9069	1.5212
AT5G43340.1	Tonoplast	NA	NA	14.1039	14.3442	NA	NA	11.8542	14.2408	13.8198
AT5G45370.1	Tonoplast	6.0263	5.5863	7.2444	5.9065	6.0046	5.9988	5.7315	6.1870	6.1794
AT5G46360.1	Tonoplast	13.7432	13.1010	13.9330	14.2228	14.1348	12.9499	10.4322	10.6263	10.1999
AT5G46370.1	Tonoplast	11.2184	11.2585	11.0075	11.3242	11.1345	11.7212	10.9948	11.6031	10.7755
AT5G46860.1	Tonoplast	2.2676	2.5169	4.0572	2.7538	2.9384	3.0674	1.4473	1.8438	1.8896
AT5G47450.1	Tonoplast	11.4645	11.9816	12.3525	11.6067	12.8614	12.9874	14.3291	14.6957	15.3570
AT5G47560.1	Tonoplast	2.9241	2.0144	3.0538	1.9856	1.5953	2.0239	0.8844	0.8250	1.5694
AT5G48410.1	Tonoplast	14.3057	6.6941	14.6230	14.7658	NA	NA	9.2888	9.1918	8.3008
AT5G55290.1	Tonoplast	0.5154	0.7068	1.1782	0.7494	0.8571	0.9975	0.2251	0.8357	0.6234
AT5G55630.1	Tonoplast	0.9019	0.7048	0.9193	0.8747	1.0006	1.3219	1.0536	1.4482	1.5292
AT5G62890.1	Tonoplast	1.2857	1.1021	2.2995	1.4083	1.7656	1.7977	2.3970	2.9030	2.9426
AT5G64870.1	Tonoplast	6.4016	6.6605	7.1558	6.3671	7.1104	7.3160	5.3918	6.9698	7.5194
AT5G67330.1	Tonoplast	2.7952	2.3170	1.6873	2.0673	2.2660	2.4100	2.2643	2.4765	2.4698

Figure S4.2 (1/23)



Transcript, Sample

Figure S4.2 (2/23)

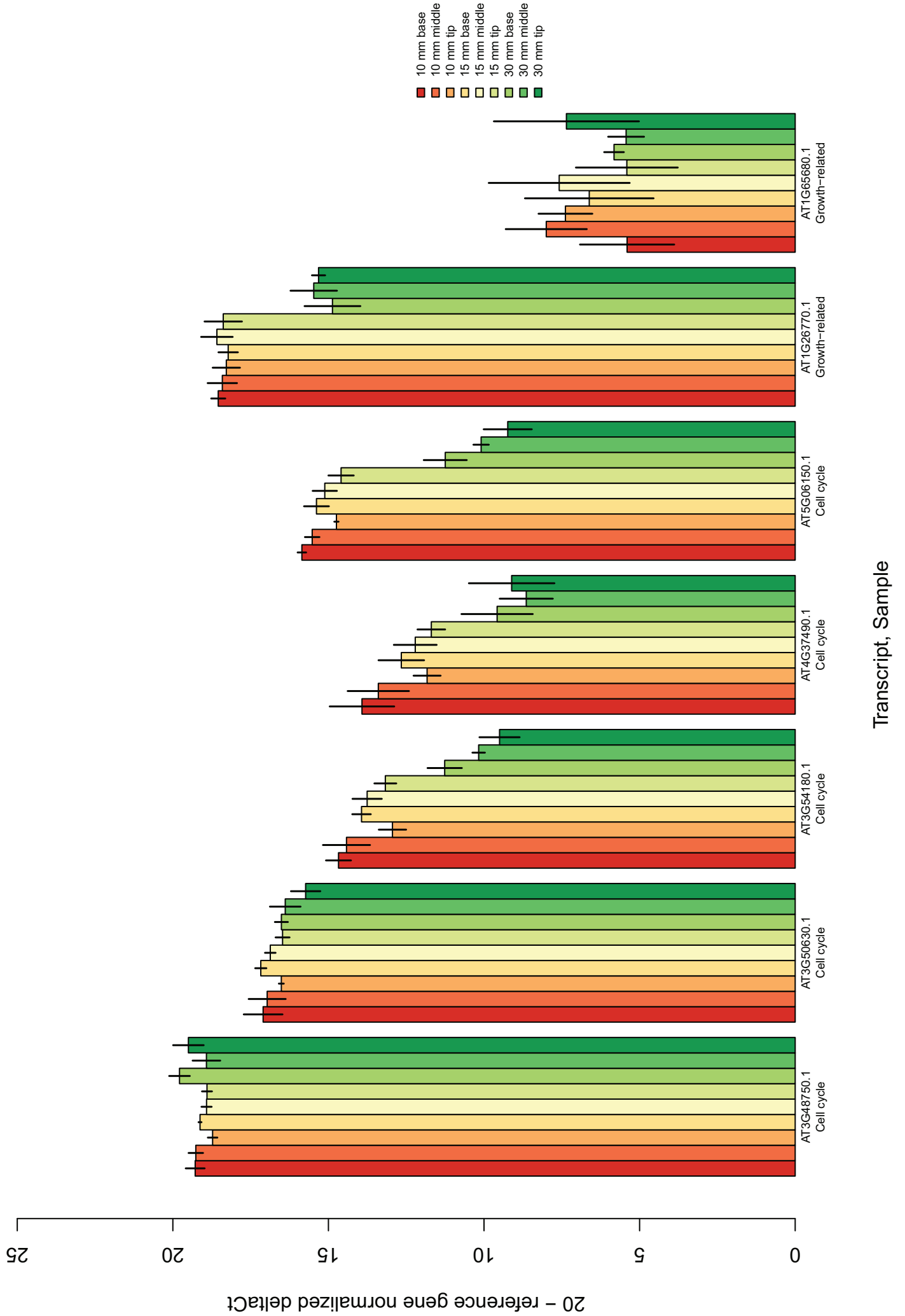
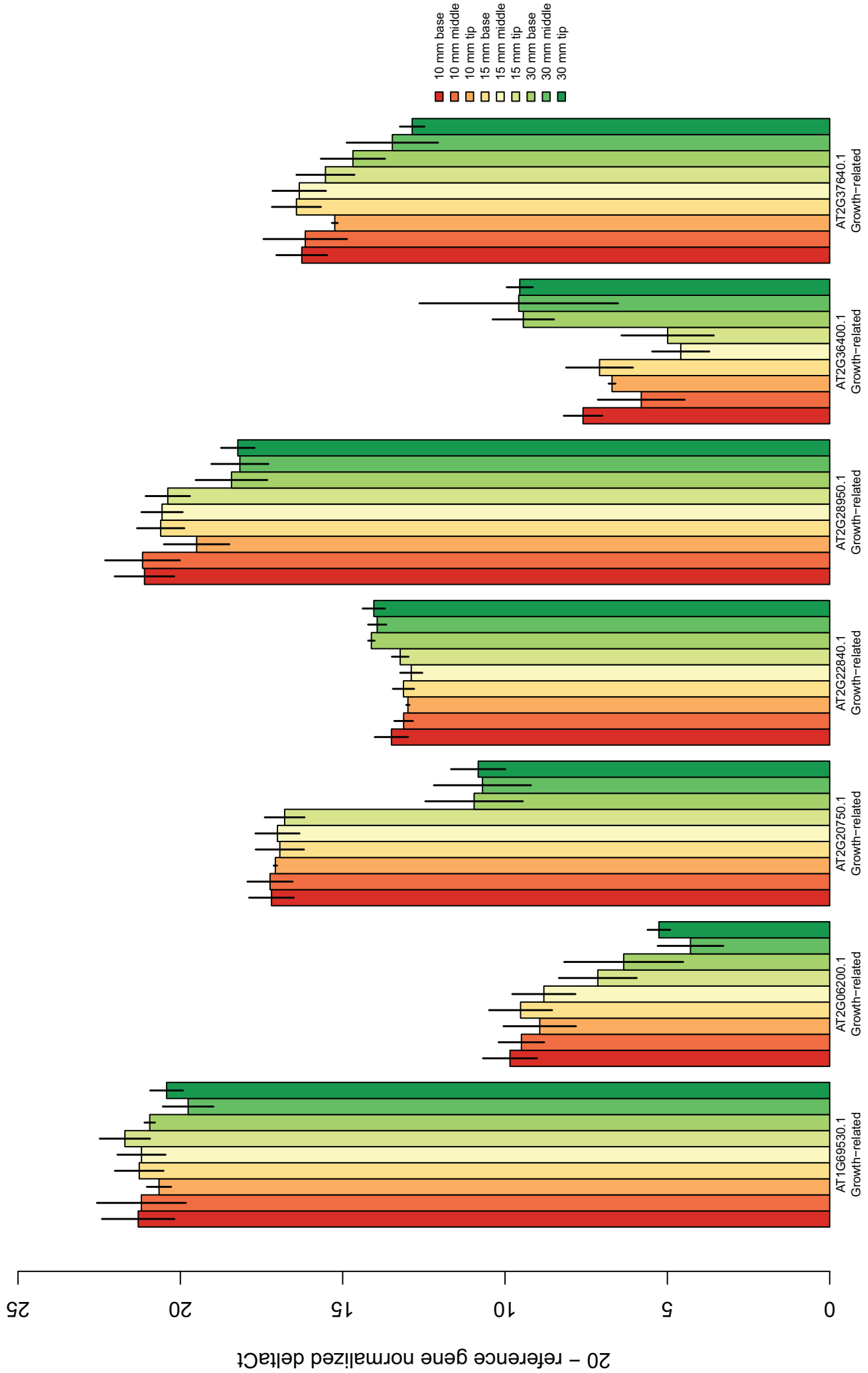


Figure S4.2 (3/23)



Transcript, Sample

Figure S4.2 (4/23)

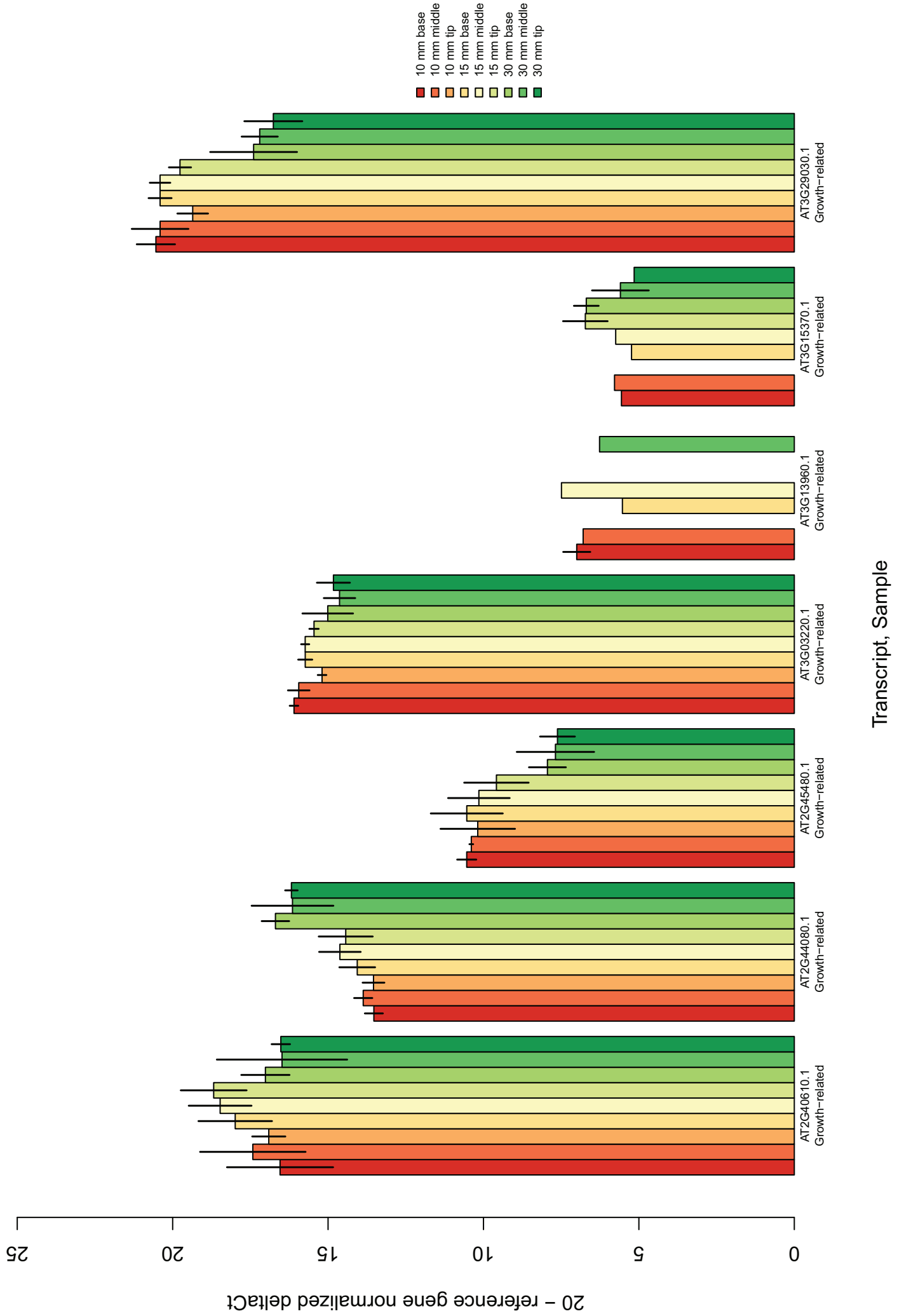
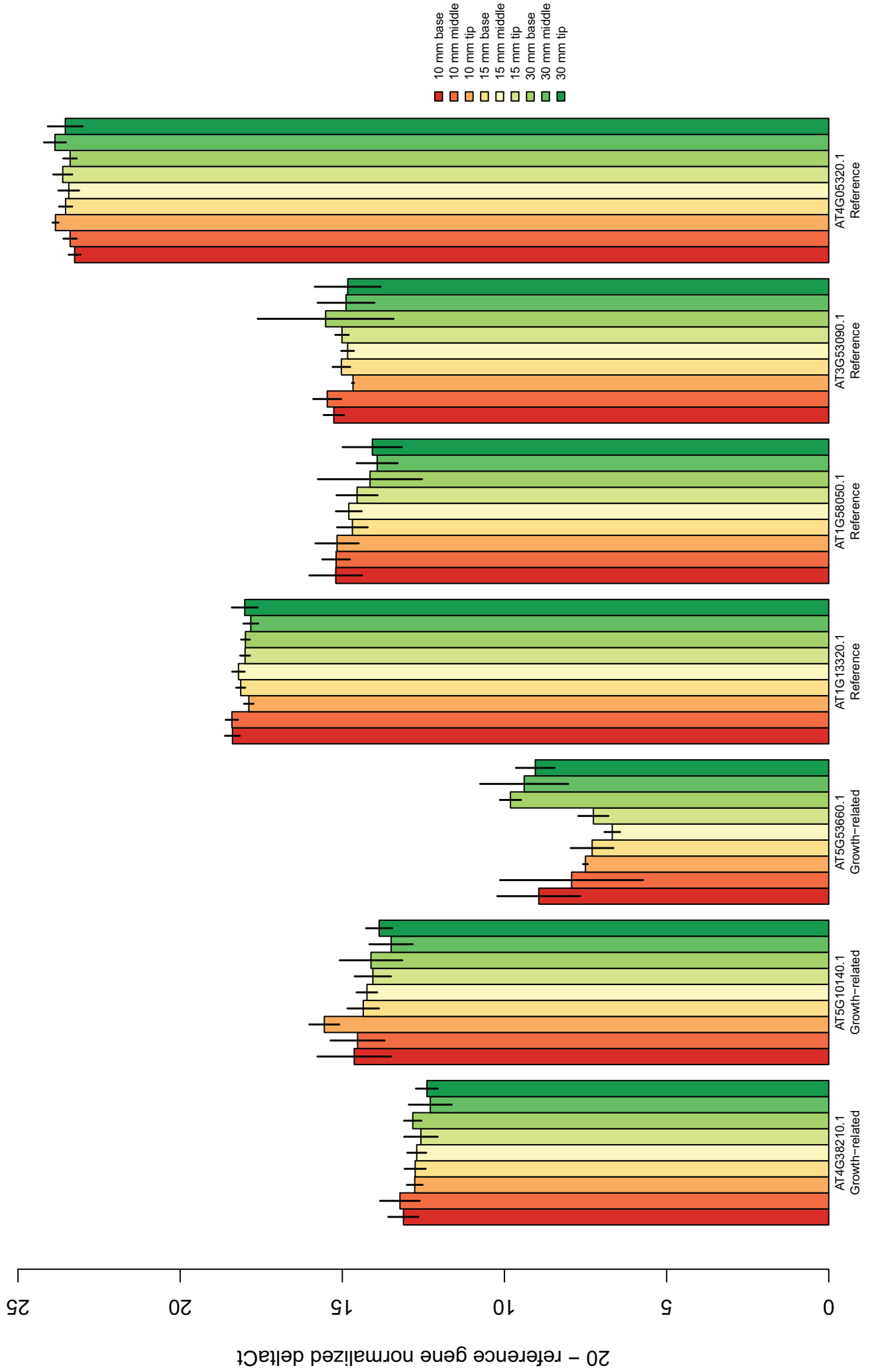


Figure S4.2 (6/23)



Transcript, Sample

Figure S4.2 (7/23)

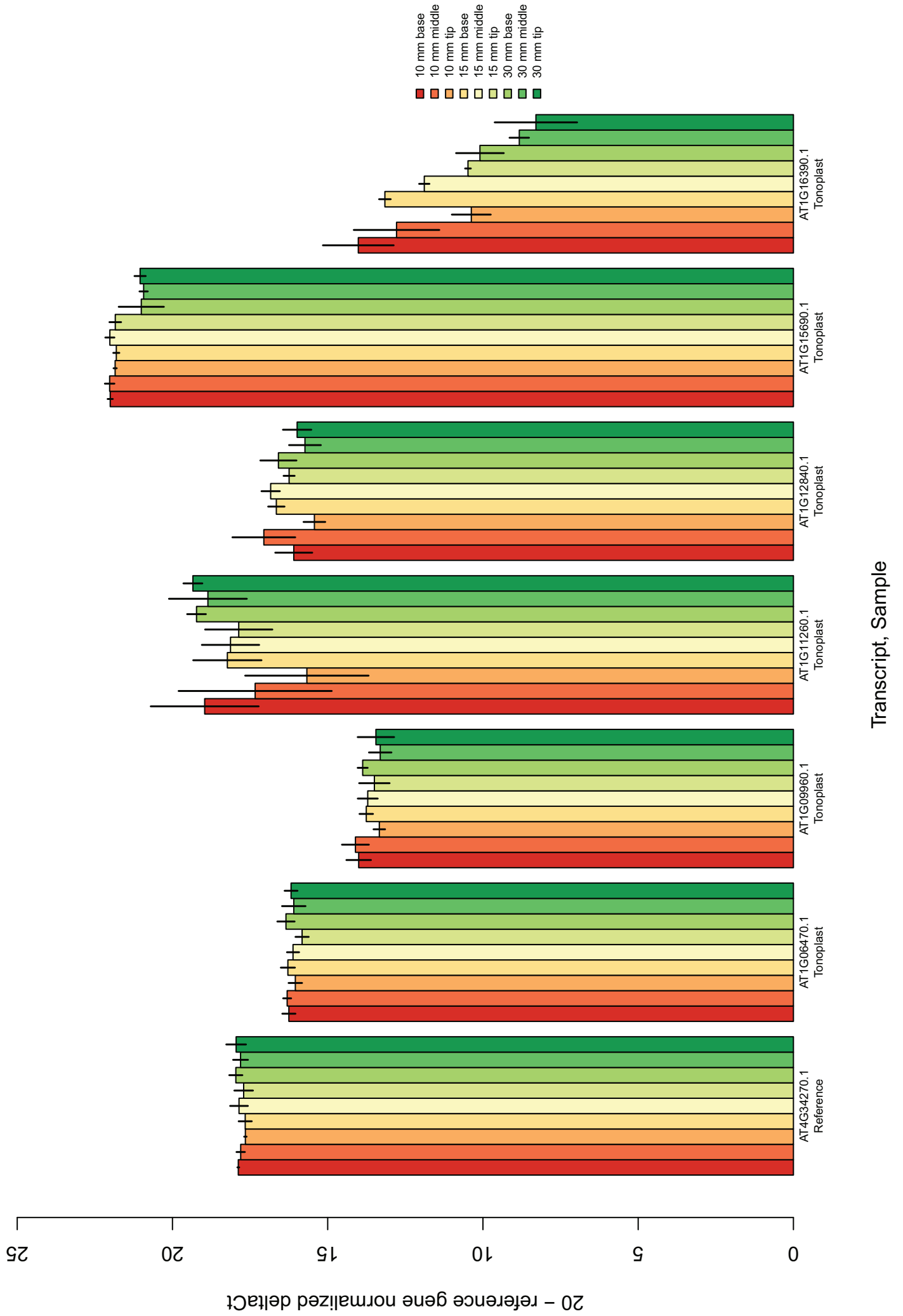
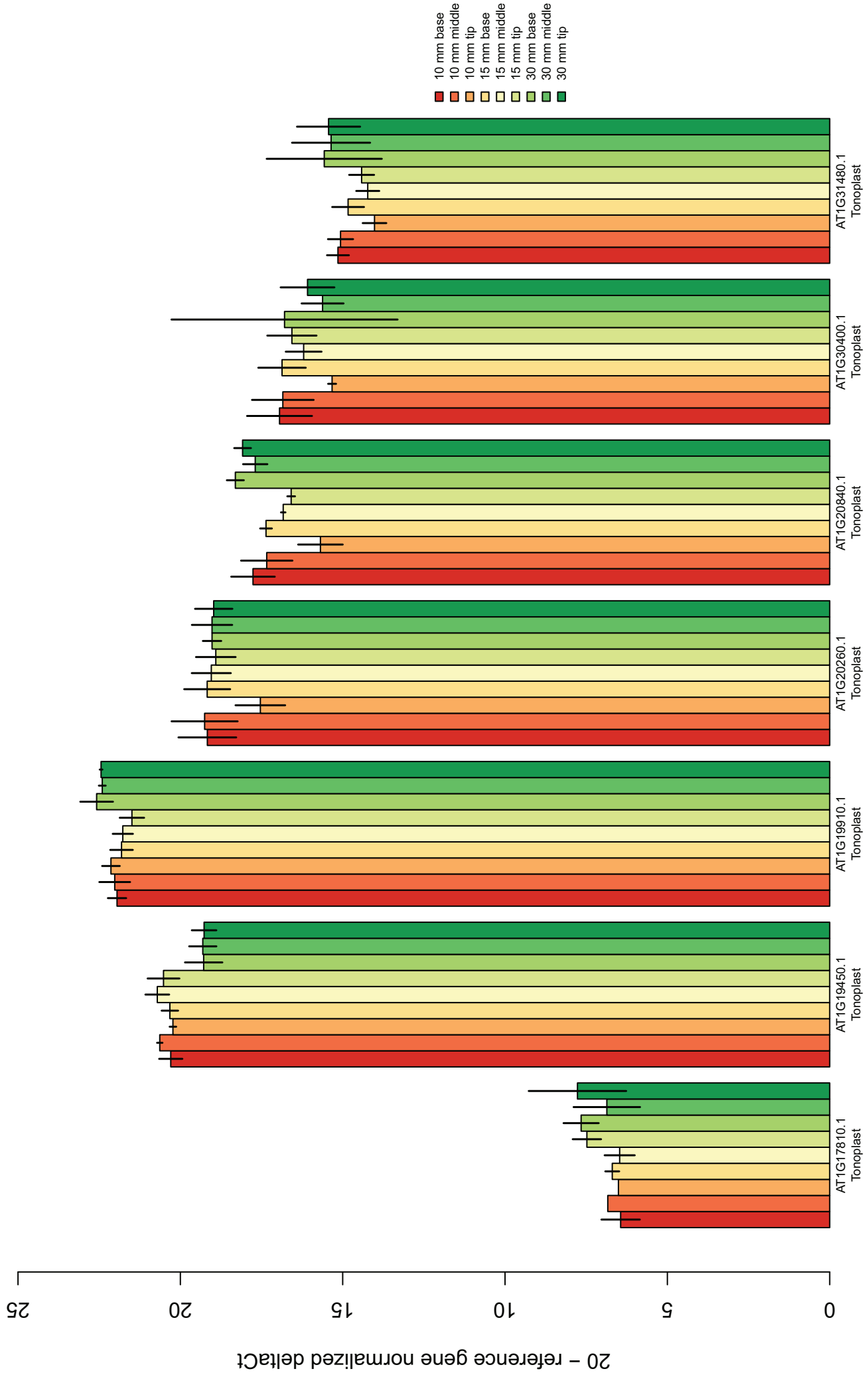
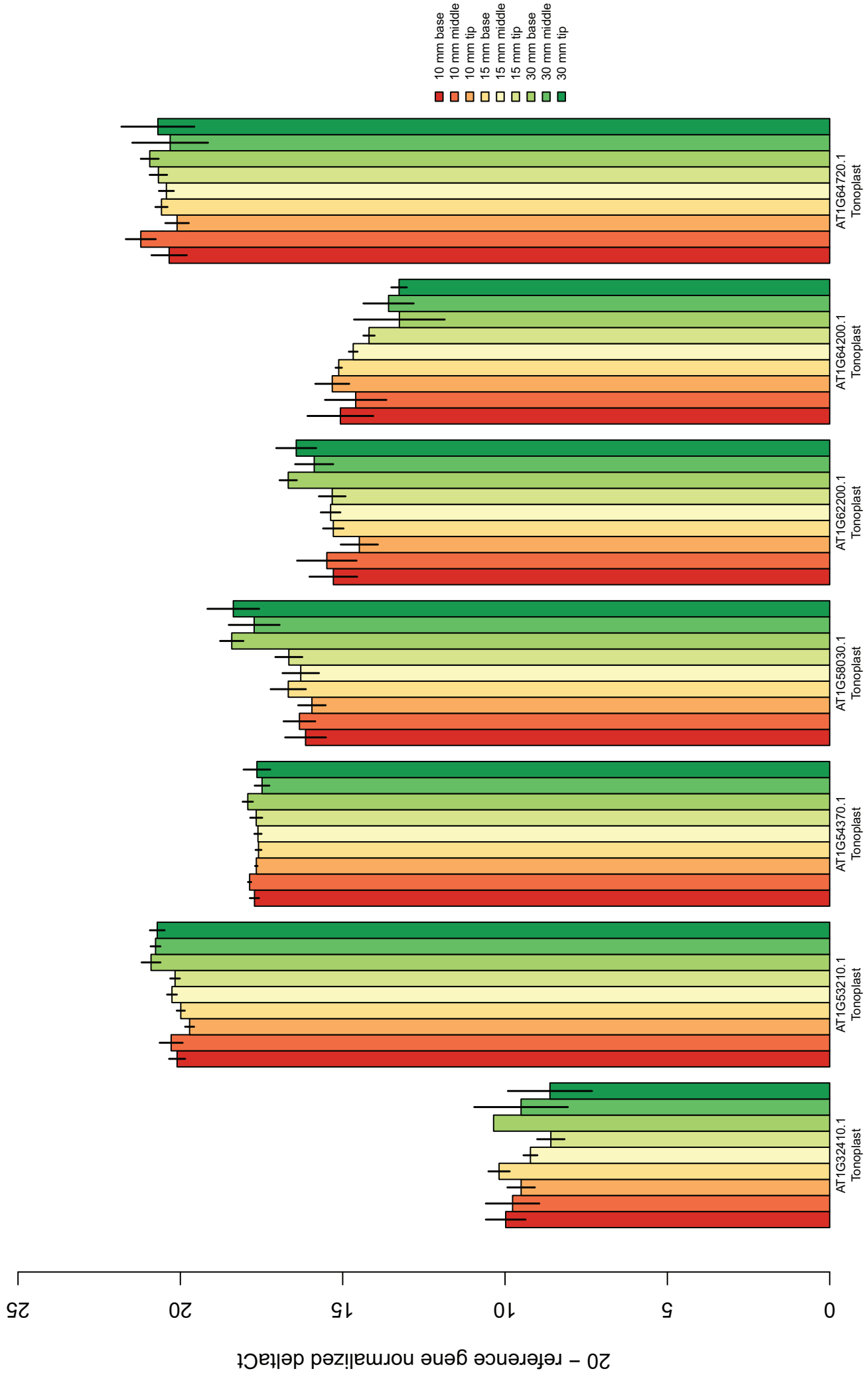


Figure S4.2 (8/23)



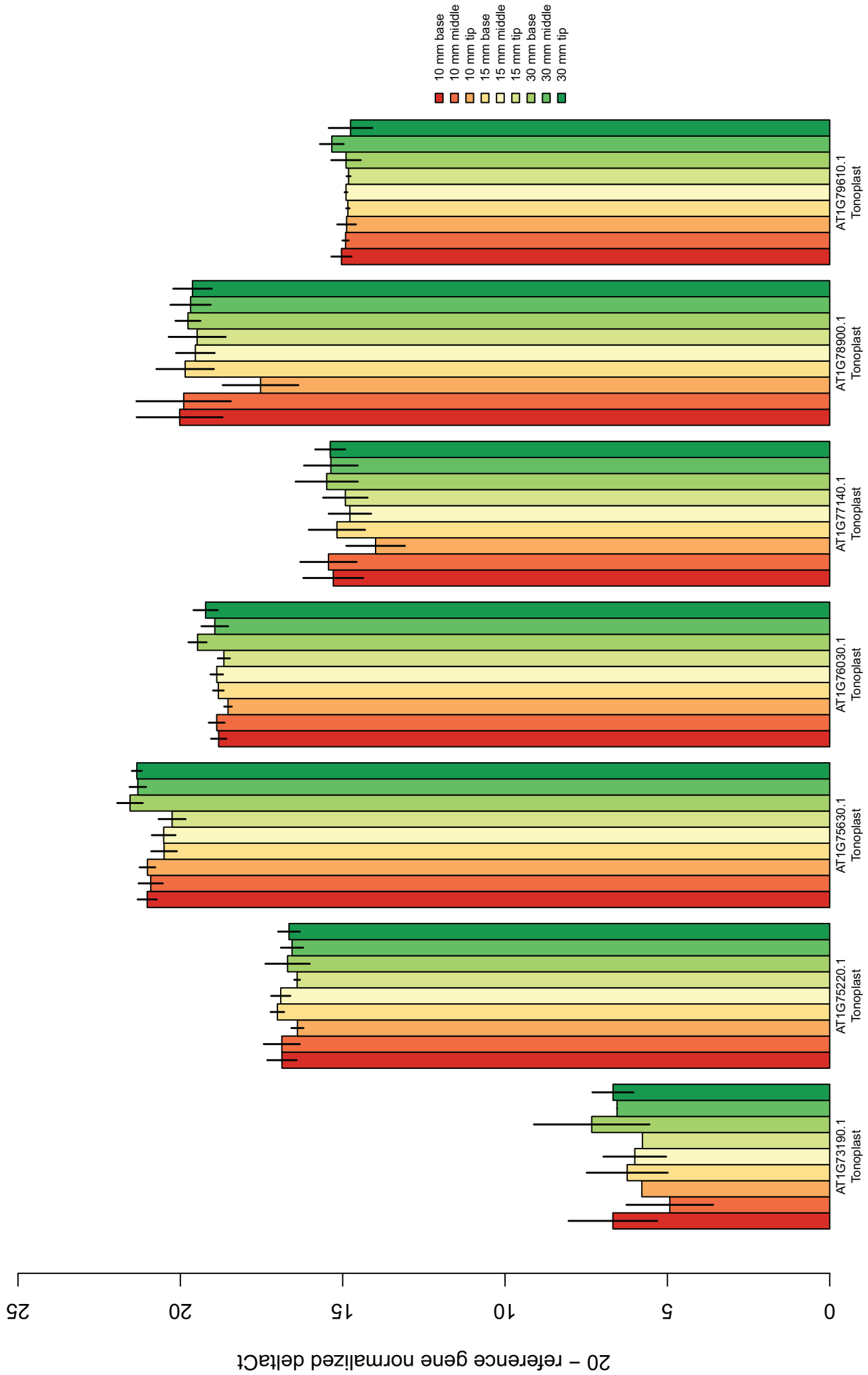
Transcript, Sample

Figure S4.2 (9/23)



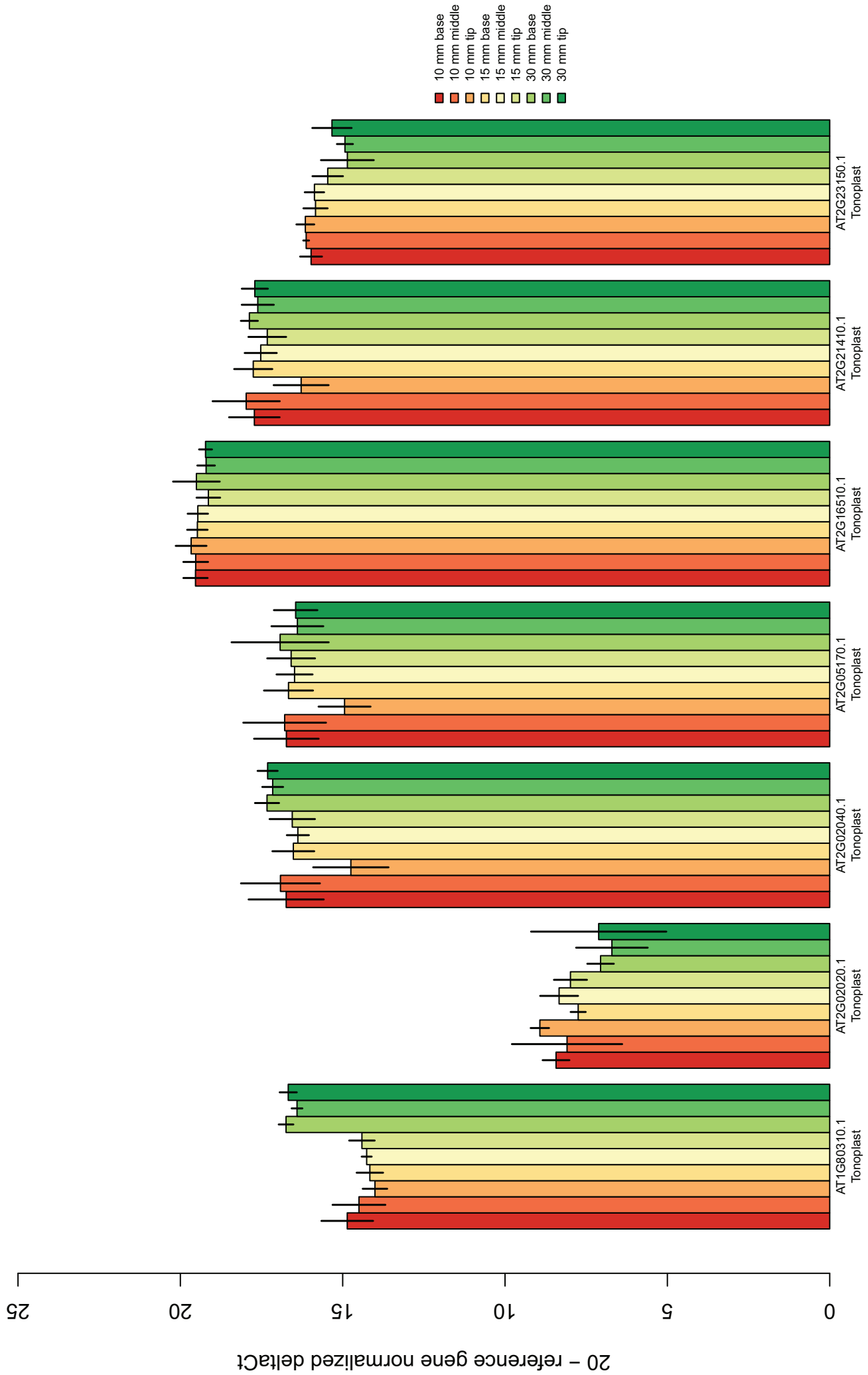
Transcript, Sample

Figure S4.2 (10/23)



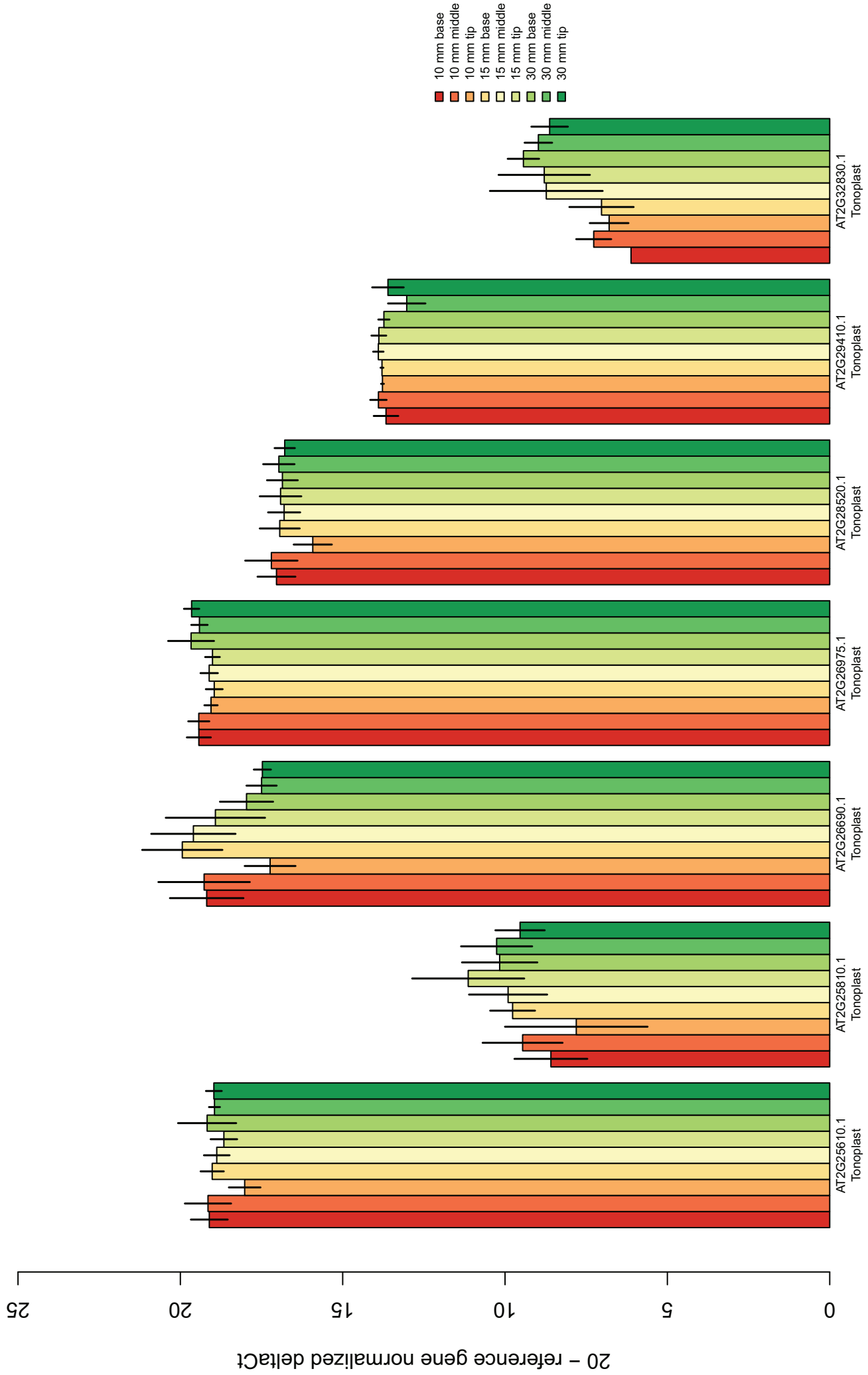
Transcript, Sample

Figure S4.2 (11/23)



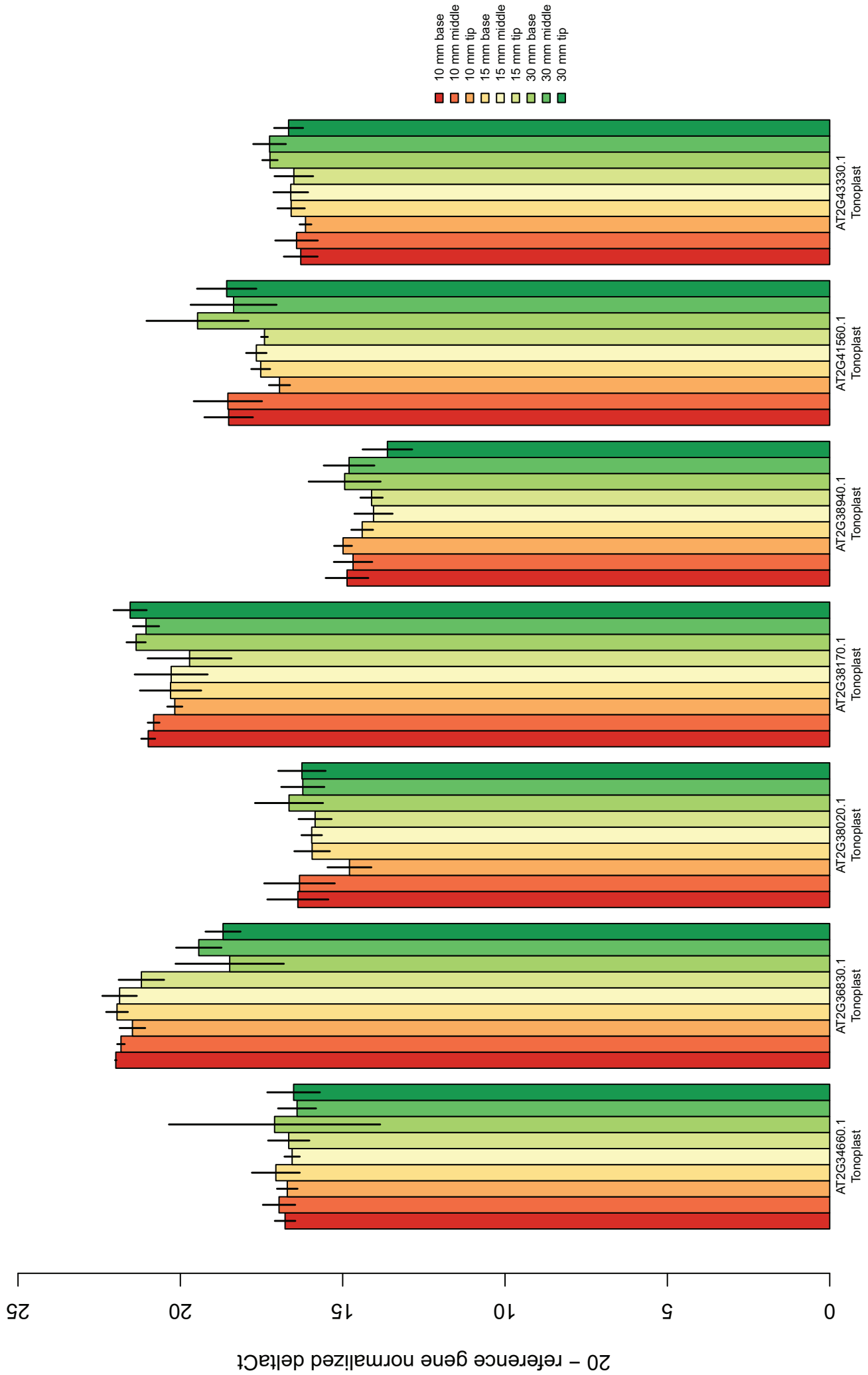
Transcript, Sample

Figure S4.2 (12/23)



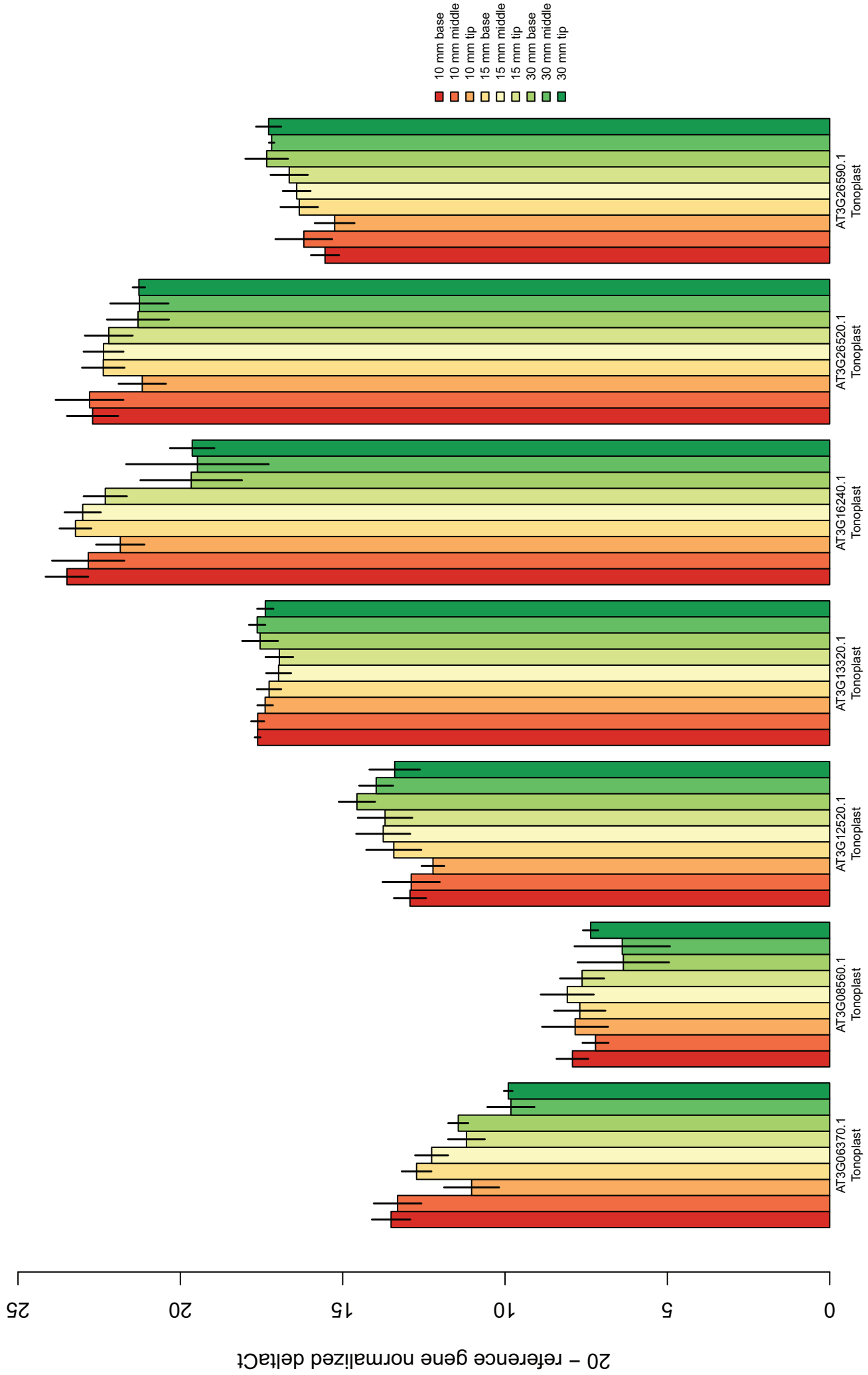
Transcript, Sample

Figure S4.2 (13/23)



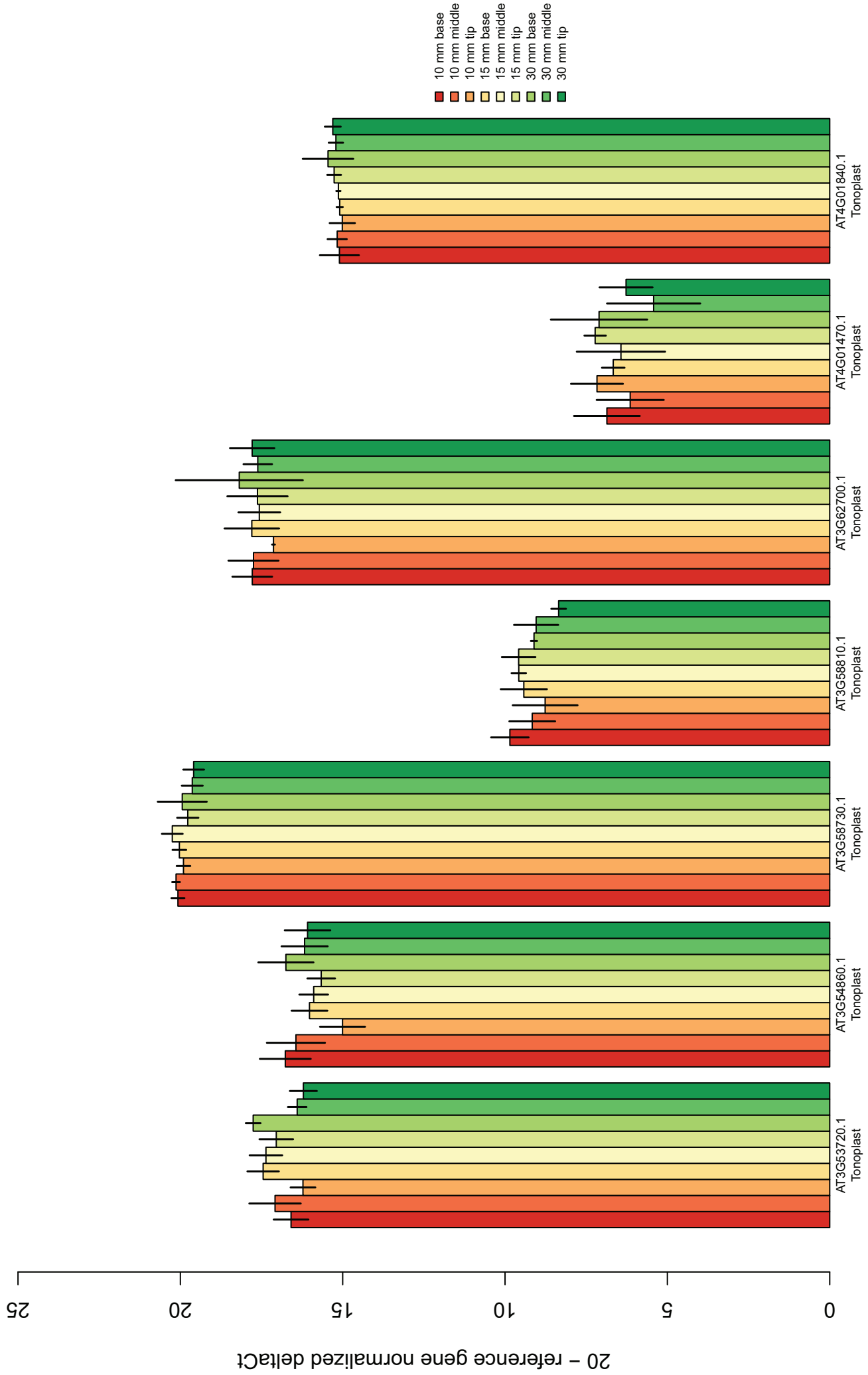
Transcript, Sample

Figure S4.2 (15/23)



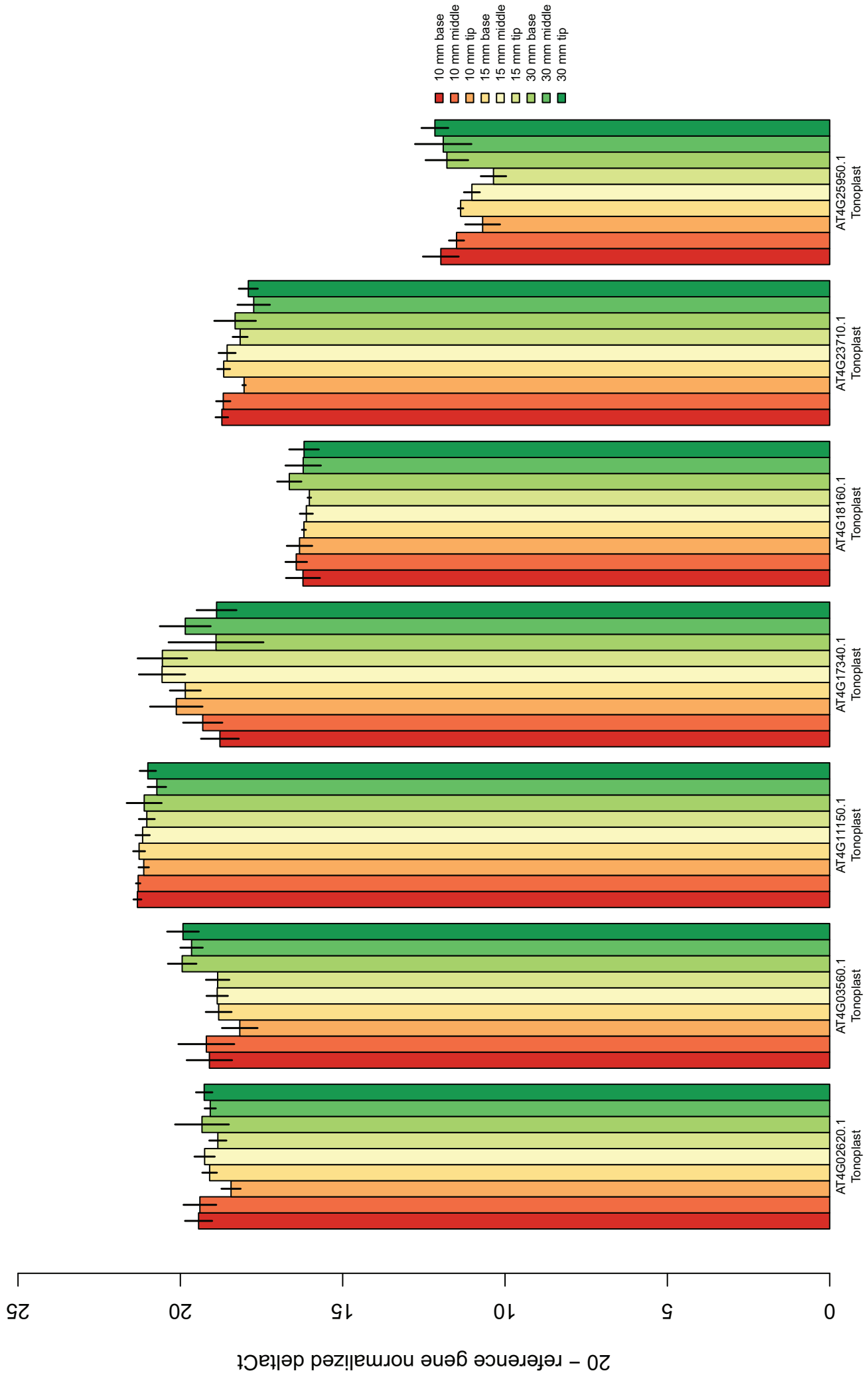
Transcript, Sample

Figure S4.2 (17/23)



Transcript, Sample

Figure S4.2 (18/23)



Transcript, Sample

Figure S4.2 (19/23)

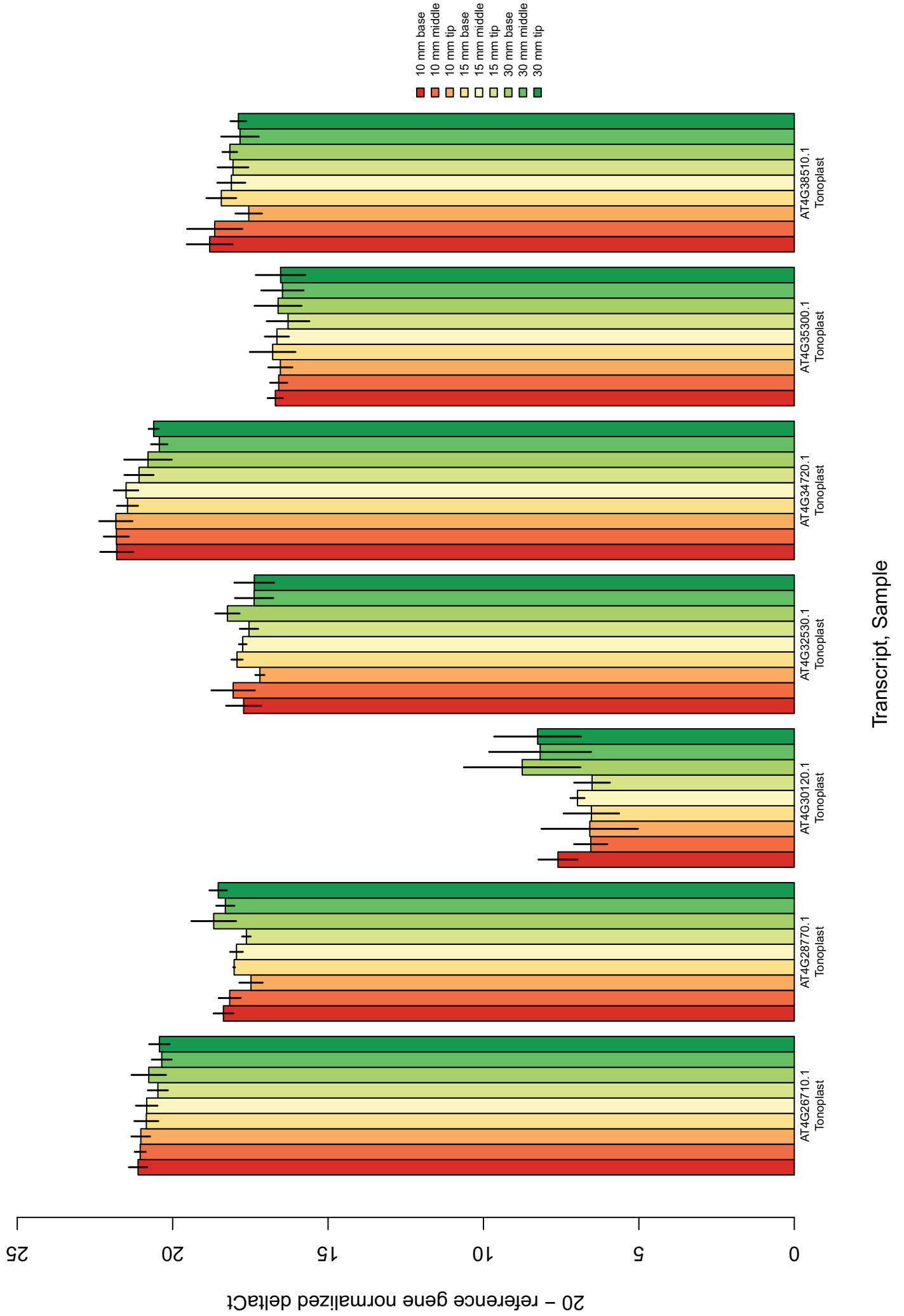
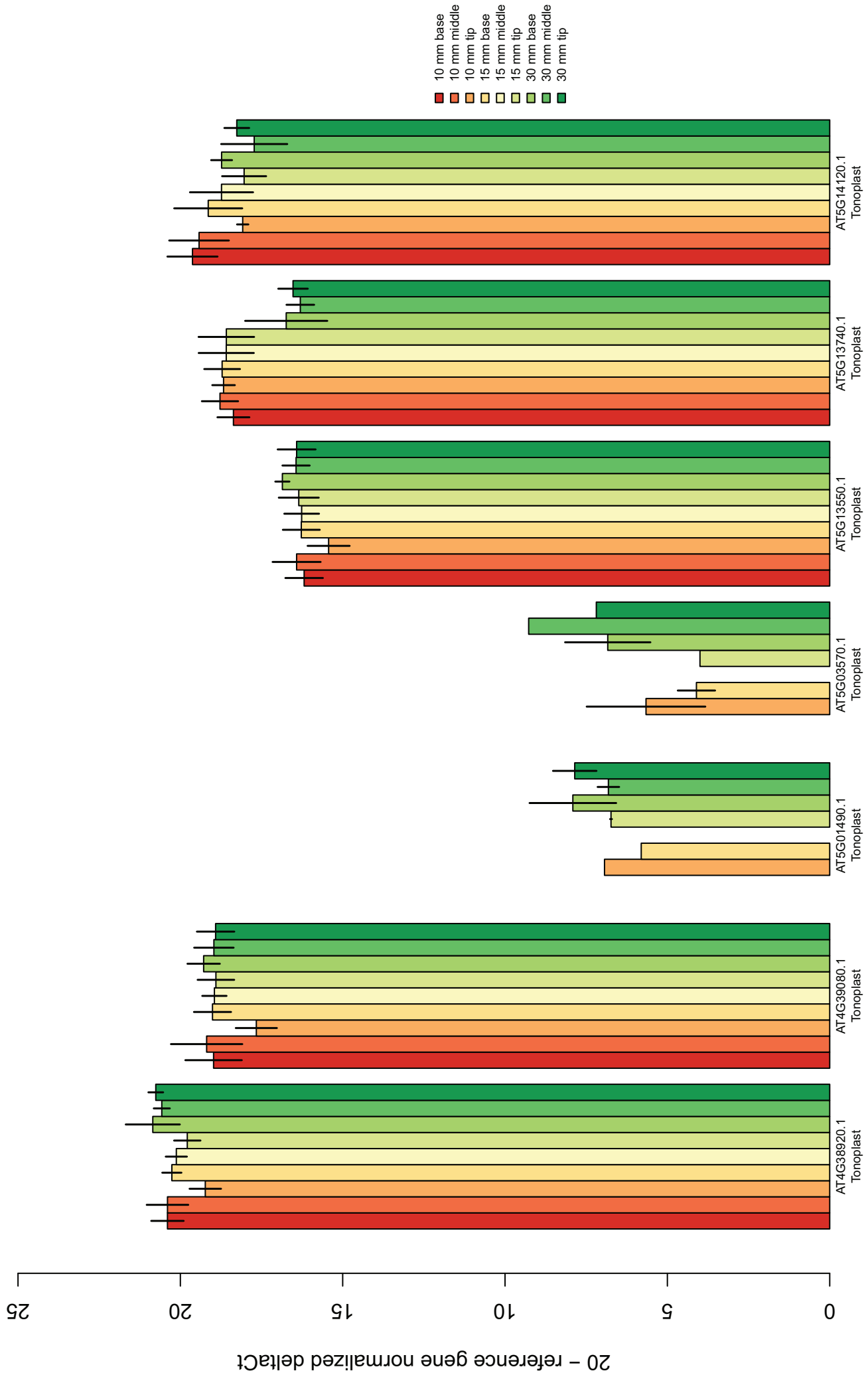


Figure S4.2 (20/23)



Transcript, Sample

Figure S4.2 (22/23)

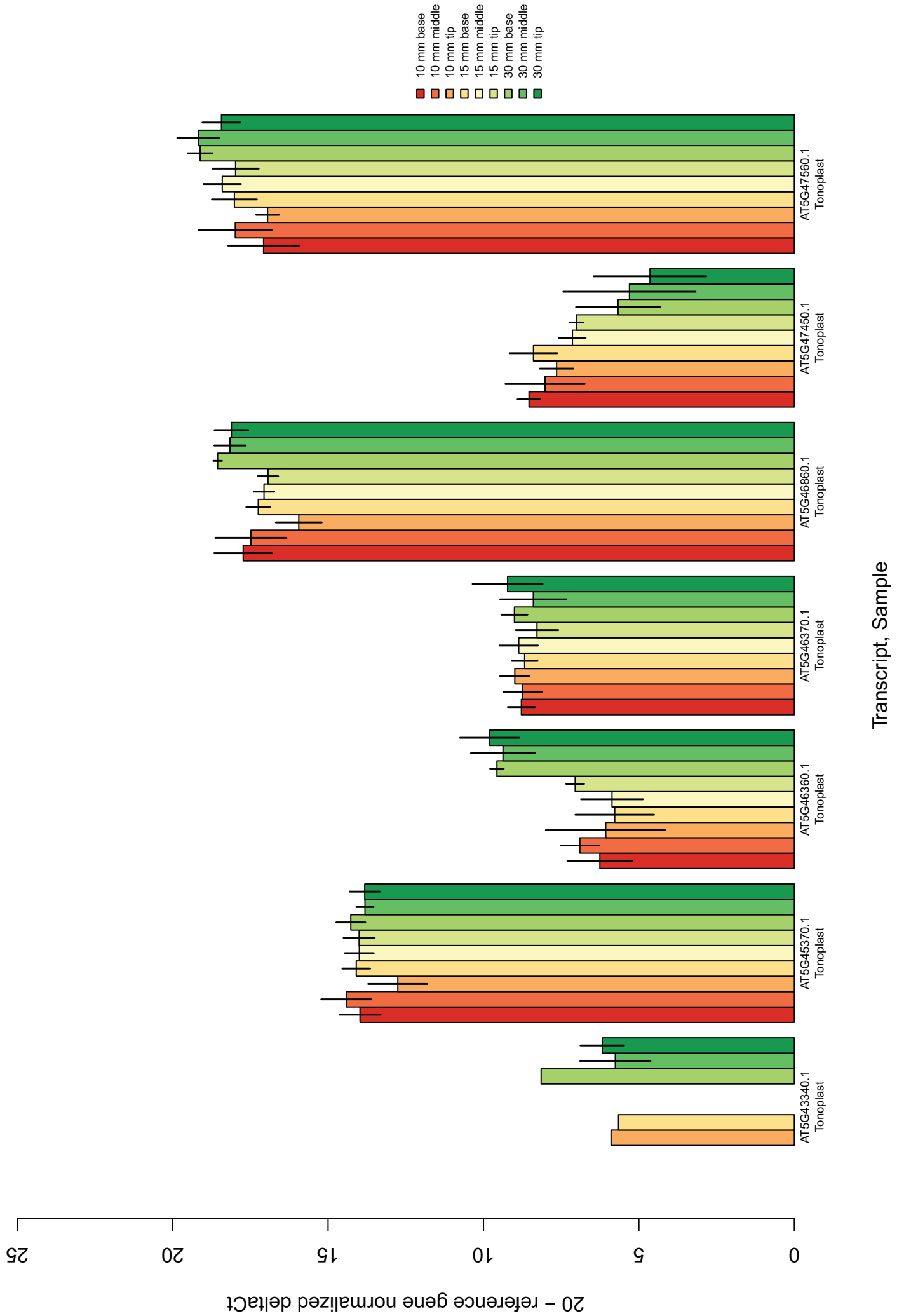


Figure S4.2 (23/23)

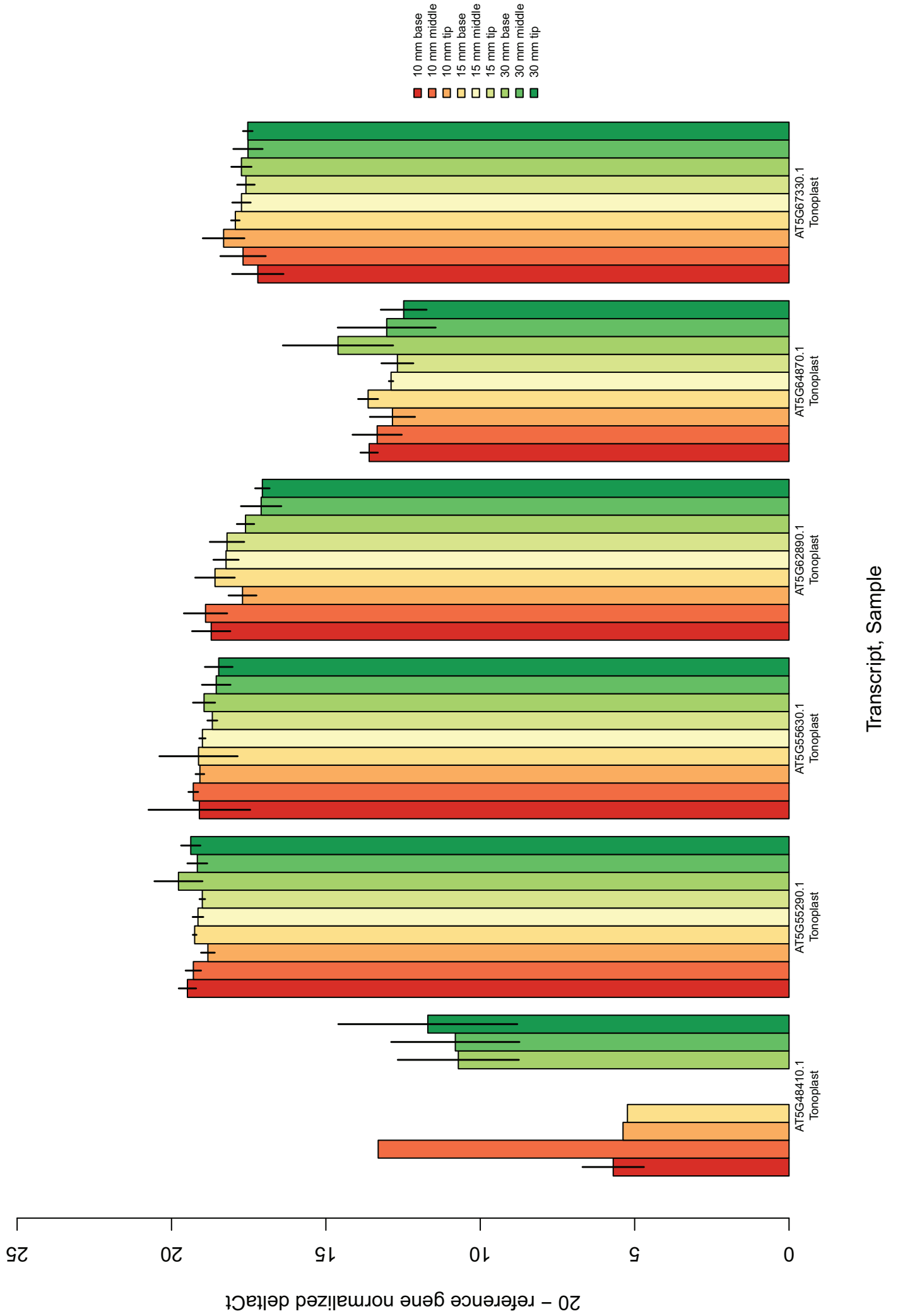


Table S4.5 (1/2)

Locus	Line code	Estimated insert pos ^a	Insertion direction	Gene specific forward primer	Gene specific reverse primer	Ref
AT1G16390	GK-294G01	530	RB-LB	GAGACATGCTCTGTAGCTGG	GCAATAGAAGCAATGCTTTGG	
AT1G16390	SM_3_20320	1100	>	AATGCGCGGCTCGTTCCCTC	ACGCCGCCAAACACCAATGC	
AT1G16390	SM_3_35473	760	>	AATGCGCGGCTCGTTCCCTC	ACGCCGCCAAACACCAATGC	
AT1G19450	GK-033C02	910	LB-RB	TGTGATTTCTCTGCTTTTGG	TGTAGAGGCATACCAACCACC	
AT1G64200	GK-138C07	935	LB-RB	ATGGCATTGAGCGTTTAGATG	CCATGTGTCCGAAGGAGAAG	vha-e3
AT1G64200	SALK_131305	200	RB-LB	TGCGTACGTTTGTAAATCAATG	CATTGACAATGTCATCTTGCG	
AT2G20750	SALK_055401	1390	RB-LB	ACCGTGGCCTAATAAACATCC	CAGGTCAGTTGTTTCGTCTC	
AT2G20750	SALK_088863	1345	RB-LB	ACCGTGGCCTAATAAACATCC	CAGGTCAGTTGTTTCGTCTC	
AT2G23150	SALK_023049	89	LB-RB	TGTCCAGAAAAACAACAACAAC	TCCTGATTTCCAAAAAGACC	
AT2G26690	SM_3_32461	447	<	ATTCGGGAACCTTTGGATTGAC	AACAAAAACAAGGTTTGGAGG	
AT2G26690	SM_3_32461	447	<	TTTTGGATATGTCCCAAAATC	AGGAACAAAAACAAGGTTTGG	
AT2G26690	SM_3_32476	447	<	TTTTGGATATGTCCCAAAATC	AGGAACAAAAACAAGGTTTGG	
AT2G26690	WiscDsLox322_H05	690	RB-LB	TTTTGGATATGTCCCAAAATC	AGGAACAAAAACAAGGTTTGG	
AT2G28950	GK-522C09	1180	LB-RB	AGGTACTAGTCGAGAAGTCTTGG	ACACCACGTCCACGTTTCACAC	exp6
AT2G28950	SALK_137859	1840	RB-LB	AGGTACTAGTCGAGAAGTCTTGG	ACACCACGTCCACGTTTCACAC	
AT2G36830	SM_3_32402	885	<	TCCTCTACTGGATTGCTCAGC	AAAGATATGGTGAATTTGTTAGACG	
AT2G37640	SALK_048023	1800	LB-RB	AGGCTTGTAGCACTTTTAGATAACGGC	ACACGGTCCATTCGACTCGATCAC	exp3
AT2G37640	WiscDsLox289_292N3	1190	RB-LB	TGCAGGCATTTGTCCTCCGCTC	GGCGCCGGCAACGTTAGTTAC	
AT2G45480	SALK_140746C	1621	LB-RB	GTGAGTCACGGCAGAAAAGAAC	ATCCATGGATTCTCAAGAACG	grf9
AT2G48020	SAIL_211_B12	2570	LB-RB	CAGGAATCATCCCATGGCTGCC	ACTGCTTCGTGGCCCATGTCGTG	
AT2G48020	SALK_037799	1700	LB-RB	CAGGAATCATCCCATGGCTGCC	ACTGCTTCGTGGCCCATGTCGTG	
AT2G48020	SALK_037804	1740	LB-RB	CAGGAATCATCCCATGGCTGCC	ACTGCTTCGTGGCCCATGTCGTG	
AT2G48020	SALK_144885	-106	LB-RB	GAAAAAGACAGAAAAGATGCC	CATTCCCTAATTCAGACCTGAG	
AT3G06370	GK-654D09	1749	LB-RB	ACTCTGTTTGGACCTTGCAAG	TCTCATCCGGTGAAGACTTTG	
AT3G06370	GK-770A08	1995	LB-RB	ACTCTGTTTGGACCTTGCAAG	TCTCATCCGGTGAAGACTTTG	nhx4-3
AT3G06370	SAIL_87_A09	2428	RB-LB	CTGGATTGCTCAGTGTCTTTTG	GACTGTTTTGATGAAGGCTGC	nhx4-2
AT3G06370	SALK_112901	42	LB-RB	TACACGAACCTATTTTGGC	GGGGTCAGTTTATGAACCTGG	nhx4-1
AT3G16240	SM_3_32074	410	>	CATATCGGAAAAGACTAGCCGC	CATGTGCTGTTGTCCAAAAAC	
AT3G16240	SM_3_32080	410	>	CATATCGGAAAAGACTAGCCGC	CATGTGCTGTTGTCCAAAAAC	
AT3G16240	SM_3_32083	410	>	CATATCGGAAAAGACTAGCCGC	CATGTGCTGTTGTCCAAAAAC	
AT3G16240	SM_3_32089	410	>	CATATCGGAAAAGACTAGCCGC	CATGTGCTGTTGTCCAAAAAC	
AT3G16240	SM_3_38811	825	>	TGGGTTCAATACTTAATGAATGTGTC	ATTACAACGAAAACCGGACCA	
AT3G16240	SM_3_39039	1360	<	CGGCTGGACTAGGATCGATAG	CGCGAACTGATATTTGAGAAGG	
AT3G26520	SM_3_30104	250	>	TGCTGCTGCAGTTGTAAATTG	GAGACTACCGTCTTTGGGGTC	
AT3G26520	SM_3_30111	250	>	TGCTGCTGCAGTTGTAAATTG	GAGACTACCGTCTTTGGGGTC	
AT3G26520	SM_3_31619	400	>	TGCTGCTGCAGTTGTAAATTG	GAGACTACCGTCTTTGGGGTC	

Table S4.5 (2/2)

Locus	Line code	Estimated insert pos ^a	Insertion direction	Gene specific forward primer	Gene specific reverse primer	Ref
AT3G26520	SM_3_31622	330	>	TGCTGCTGCAGTTGTAAATTG	GAGACTACCGTTCTTGGGGTC	
AT3G26520	SM_3_31771	400	>	TGCTGCTGCAGTTGTAAATTG	GAGACTACCGTTCTTGGGGTC	
AT3G29030	SALK_043239	1224	LB-RB	ACATGCAACATTTGTCAACGG	TTAATTTGCCATTTTGTCTCTCG	
AT3G29030	SALK_043239	1225	LB-RB	TGAACTAAATTAGGTGGAGCATGCGG	ATAAGAGTTGCTTTGCCAATTTTGCCC	
AT3G29030	WiscDsLox495_F06	1275	RB-LB	TGAACTAAATTAGGTGGAGCATGCGG	ATAAGAGTTGCTTTGCCAATTTTGCCC	
AT3G51490	SALK_027520	1138	RB-LB	TTTGGTCTCCCAATGTCTACG	GATCATGACCATCACCCCTCTG	
AT3G52910	SALK_037642C	181	RB-LB	TTGTAGCTCTCTGCAAAATGGAG	AAGCACGACCAACATGTATC	
AT3G52910	SALK_077829C	1011	RB-LB	AACTGCTACTTCTGTCTGCTGC	TCTTGTCACCCTATTGCCAAC	
AT3G55500	GK-863H08	170	LB-RB	ACTCTTCCCTCTCTTGAGCTTCAACCG	ATGTAACAAGGAAAGGAAAGTGACTGAC	
AT4G37740	SALK_003203	2335	LB-RB	ACCGATGAAACAACAATCGAG	TGACTGAACAACCAACCCCTC	
AT4G37740	SALK_035804	2356	LB-RB	ACCGATGAAACAACAATCGAG	TGACTGAACAACCAACCCCTC	
AT4G37740	SALK_035805	2356	LB-RB	ACCGATGAAACAACAATCGAG	TGACTGAACAACCAACCCCTC	
AT4G37750	GK-874H08	660	LB-RB	AAACCAATGATGGGTTTTAGC	TGACTGAACAACCAACCCCTC	
AT4G37750	SALK_022770	2812	LB-RB	TCTTGTCTGGAGAGTTAGCGC	CTGCCCCCATATTAATTTTGG	
AT5G13740	SALK_011451	220	RB-LB	AAACATTTTGAATCCGAAGC	AAATCAATTTCCGGAATCTCG	
AT5G13740	SALK_016418	4366	RB-LB	AGTGAGAGGGGACTAAACGCTGC	AGCCAATCATGCCGGCTTTAACC	
AT5G13740	SALK_052434	4200	RB-LB	AGTGAGAGGGGACTAAACGCTGC	AGCCAATCATGCCGGCTTTAACC	
AT5G13740	SALK_121565	4100	RB-LB	AGTGAGAGGGGACTAAACGCTGC	AGCCAATCATGCCGGCTTTAACC	
AT5G47450	SAIL_724_E08	1080	RB-LB	AATGAGTTGTTTCCCGGTTTG	GATCGACGAGCTCAGATGAAC	
AT5G47450	SALK_010974	240	LB-RB	ATCAAAATAAGAGCGGTTTGG	TCCAGATTTGGACAAGTCAAC	
AT5G47450	SALK_010989	470	RB-LB	ATCAAAATAAGAGCGGTTTGG	TCCAGATTTGGACAAGTCAAC	
AT5G47450	SALK_126757	1130	RB-LB	AATGAGTTGTTTCCCGGTTTG	GATCGACGAGCTCAGATGAAC	
AT5G47450	SALK_127491	520	LB-RB	AGTGGACACCAACAATACACACCGC	GCGTTGGCTCTGCTGTTTGGC	
AT5G47450	SALK_142179	21	LB-RB	ATCAAAATAAGAGCGGTTTGG	TCCAGATTTGGACAAGTCAAC	
AT5G62890	GK-340A03	1540	LB-RB	GTGATGCTTTGGACAAACGGTGC	GCTCAGGCAGTCCAAATCTCTATGC	
AT5G62890	SALK_065004	1450	LB-RB	GTGATGCTTTGGACAAACGGTGC	GCTCAGGCAGTCCAAATCTCTATGC	
AT5G62890	SALK_078079	1150	LB-RB	GTGATGCTTTGGACAAACGGTGC	GCTCAGGCAGTCCAAATCTCTATGC	
AT5G62890	SALK_078239	1150	LB-RB	GTGATGCTTTGGACAAACGGTGC	GCTCAGGCAGTCCAAATCTCTATGC	
AT5G62890	SALK_131213	1450	LB-RB	GTGATGCTTTGGACAAACGGTGC	GCTCAGGCAGTCCAAATCTCTATGC	

^a TAIR9 genomic sequence, position relative to first exon.

Figure S4.3 (1/2)

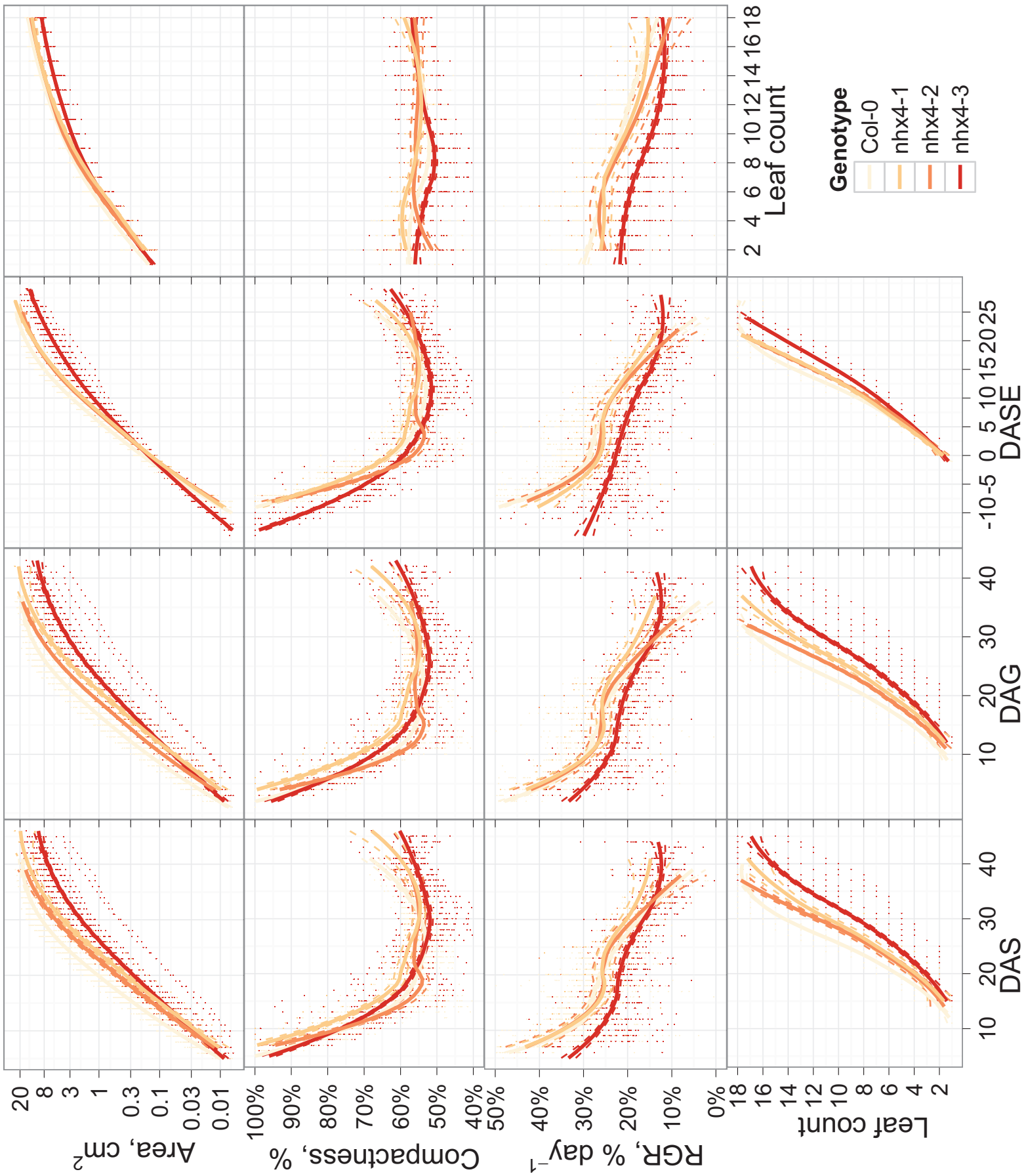
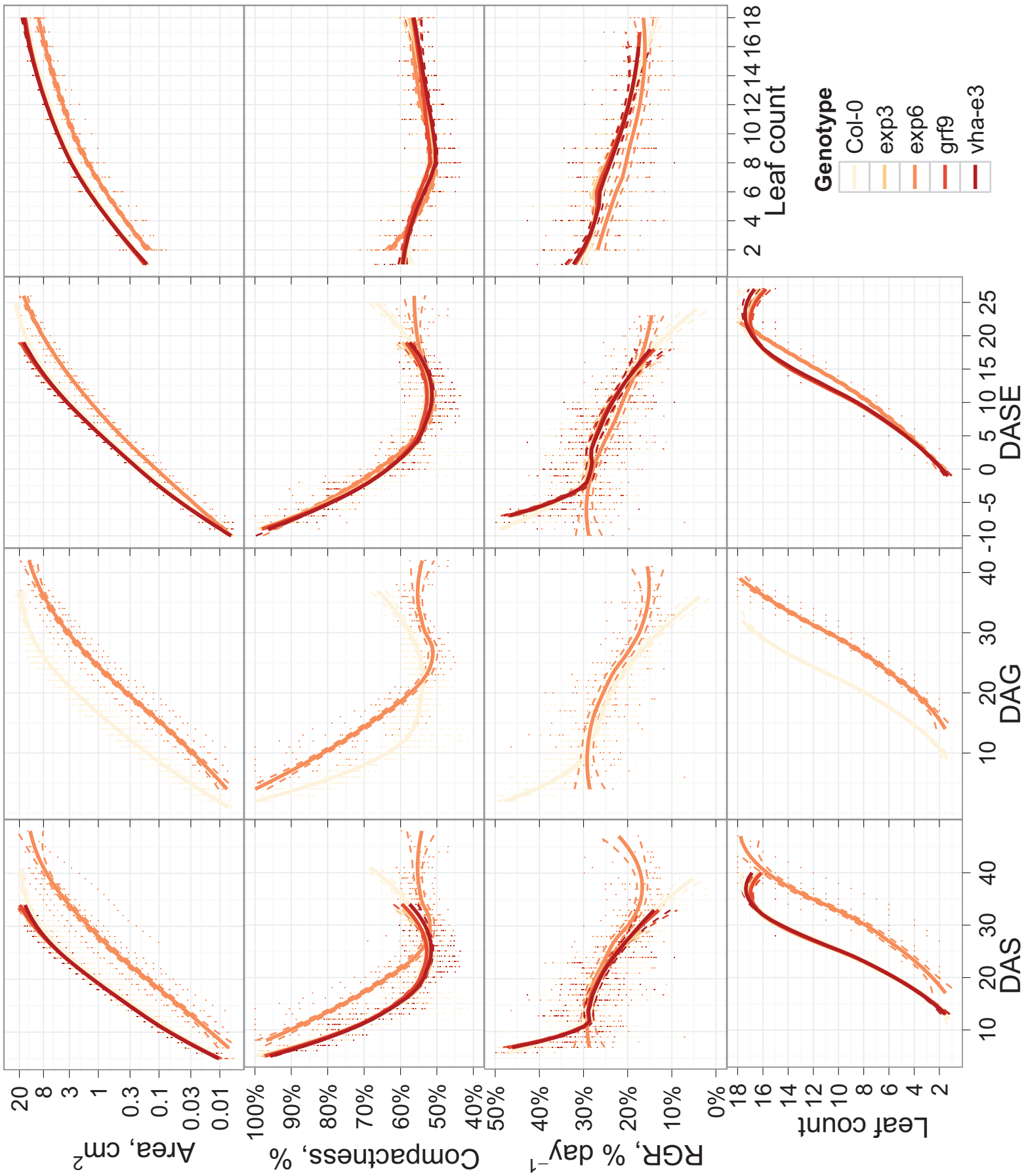


Figure S4.3 (2/2)



Allgemeinverständliche Zusammenfassung

Sehr vereinfacht gesagt kann Blattwachstum auf zwei Prozesse reduziert werden, Zellteilung und Zellexpansion, gefolgt von Zellwandexpansion. Die Vakuole, das größte Organell der Zelle, übt durch die Kontrolle des Wasserhaushaltes der Pflanze eine wichtige Funktion im Zusammenhang mit der Zellexpansion aus. Dies geschieht durch die Regulierung des osmotischen Druckes, durch Import und Export von organischen und anorganischen Ionen über die Vakuolenmembran (den Tonoplast) und durch die Kontrolle ihrer Wasserkanäle (der Aquaporine). Es wird angenommen, dass die Regulierung des vakuolären osmotischen Druckes eine große Rolle bei der Zellexpansion spielt, da der osmotische Druck die Stärke der mechanischen Kraft des Tonoplast auf die Plasmamembran und die Zellwand bestimmt.

In dieser Dissertation wird die Rolle von Tonoplastproteinen und ihrer Gene auf die Zellexpansion anhand der Modellpflanze *Arabidopsis thaliana* (Ackerschmalwand) untersucht, und Kandidaten für wachstumsrelevante Gene werden identifiziert.

Da bisher noch kein Signal für die Lokalisierung von Proteinen im Tonoplast identifiziert wurde, gibt es keine Möglichkeit, genomweite Voraussagen über solche Proteinlokalisierungen zu machen. Daher haben wir eine Reihe von aktuellen Proteom-Studien genutzt, um eine Liste von 117 Genen, die für transmembrane tonoplastproteinkodierende Gene kodieren, zusammenzustellen. Zusätzlich wurden andere wachstumsrelevante Gene und Zellzyklus-Gene in die Liste aufgenommen (38 Gene). Die Expression der Gene während der Blattentwicklung sollte mittels einer sensitiven Technik, der quantitativen Polymerasekettenreaktion (qPCR), untersucht werden. Um rasch die für dieses Verfahren notwendigen Oligonukleotide zu entwerfen, wurde ein Computerprogramm („QuantPrime“) entwickelt. Das Programm entwirft automatisch solche Oligonukleotide und überprüft deren Spezifität *in silico* auf Ebene der Transkriptome und Genome um Kreuz-Hybridisierungen zu vermeiden, die zu unspezifischen Amplifikationen führen würden.

Die qPCR-Plattform wurde in einer Expressions-Studie eingesetzt, um wachstumsrelevante Gen-Kandidaten zu identifizieren. Um wachstumsaktive und nichtaktive Prozesse vergleichen zu können, wurden Proben von unterschiedlichen Bereichen des Blattes zu unterschiedlichen Wachstumsstadien beprobt. Eine musterbasierte Expressionsdatenanalyse wurde eingesetzt, um die Gene hinsichtlich ihrer Assoziation mit der Blattexpansionen in eine Rangordnung zu bringen. Die

Gene mit dem höchsten Rang wurden als Kandidaten für weitere Experimente ausgewählt.

Um die funktionelle Beteiligung dieser Gene auf einer makroskopischen Ebene zu untersuchen, wurden Knockout-Mutanten für die Gen-Kandidaten hinsichtlich ihres Wachstums analysiert. Zu diesem Zweck wurde ein System für die automatisierte Phänotypisierung des Blattwachstums etabliert. Zum einen wurde ein Programm-Paket für detaillierte Annotation von Wachstumsstadien und zum anderen ein Analyse-Paket für automatisierte Datenvorbereitung und statistische Tests entwickelt. Das Analyse-Paket erlaubt die Modellierung und graphische Darstellung verschiedener wachstumsrelevanter Phänotypen. Mit Hilfe dieses Systems wurden 24 Knockout-Mutanten untersucht und signifikante Phänotypen wurden für fünf verschiedene Gene gefunden.

List of publications

Winck, F.V., **Arvidsson, S.**, Riaño-Pachón, D.M., Koseka, A., Donner, S., Rupprecht, J. and Mueller-Roeber, B. (in prep.) Deciphering the gene regulatory network of *Chlamydomonas* under carbon deprivation

Winck, F.V., Riaño-Pachón, D.M., **Arvidsson S.**, Trejos-Espinosa, R. and Mueller-Roeber, B. (in prep.) High light responsive transcription factors in *Chlamydomonas reinhardtii*

Arvidsson, S., Arif M. and Mueller-Roeber, B. (in prep.) Identification and characterization of leaf growth-related tonoplast protein genes in *Arabidopsis thaliana* using transcriptomics and high throughput phenotyping

Rohrmann, J., Tohge, T., Rob, A., Osorio, S., Caldana, C., McQuinn, R., **Arvidsson, S.**, van der Merwe, M.J., Riaño-Pachón, D.M., Mueller-Roeber, B., Nunes Nesi, A., Giovannoni, J.J. and Fernie, A.R. (under revision) Combined transcription factor profiling, microarray analysis and metabolite profiling reveals the transcriptional control of metabolic shifts occurring during tomato fruit development, *The Plant Journal* (under revision)

Arvidsson, S., Pérez-Rodríguez, P. and Mueller-Roeber, B. (accepted) A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects, *New Phytologist* (in press)

Mueller-Roeber, B. and **Arvidsson, S.** (2009) Fertility control: the role of magnesium transporters in pollen development, *Cell Research*, 19, 800–801

Arvidsson, S., Kwasniewski, M., Riaño-Pachón, D.M. and Mueller-Roeber, B. (2008) QuantPrime – a flexible tool for reliable high-throughput primer design for quantitative PCR, *BMC Bioinformatics*, 9, 465

Courbot, M., Willems, G., Motte, P., **Arvidsson, S.**, Roosens, N., Saumitou-Laprade, P. and Verbruggen, N. (2007) A major quantitative trait locus for cadmium tolerance in *Arabidopsis halleri* colocalizes with *HMA4*, a gene encoding a heavy metal ATPase, *Plant Physiology*, 144, 1052-1065

The pages 151-153 (Curriculum vitae) contain personal data. They are therefore not part of the online version of this publication.

The pages 151-153 (Curriculum vitae) contain personal data. They are therefore not part of the online version of this publication.

The pages 151-153 (Curriculum vitae) contain personal data. They are therefore not part of the online version of this publication.