The characterization of the caleosin gene family in Triticeae and their role in G-protein signalling

and

Identification and characterization of rye genes silenced in allohexaploid triticale: A bioinformatic study

Hala Badr Abdel-Sadek Khalil

A Thesis

in

The Department

of

Biology

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy at

Concordia University

Montreal, Quebec, Canada

July 2013

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is certify that the thesis prepared

By: **Hala Badr Abdel-Sadek Khalil**

Entitled: **The characterization of the caleosin gene family in Triticeae and their role in G-protein signalling**

and

Identification and characterization of rye genes silenced in allohexaploid triticale: A bioinformatic study

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

DOCTOR OF PHILOSOPHY (Biology)

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

----------Chair	
Fathy Sarhan	Examiner
Reginald Storms	Examiner
Luc Varin	Examiner
Patrick Gulick	Supervisor

Approved by ----------

Chair of Department or Graduate Program Director

July, 2013 ----------

Dean of Faculty

ABSTRACT

The characterization of the caleosin gene family in Triticeae and their role in G-protein signalling and Identification and characterization of rye genes silenced in allohexaploid triticale: A bioinformatic study

Hala Badr Abdel-Sadek Khalil, Ph.D.

Concordia University, 2013

The caleosin genes encode proteins with a single conserved EF hand calcium-binding domain and comprise small gene families found in a wide range of plant species. In this study,*Clo3*, a member of caleosin family in hexaploid wheat (*Triticum aestivum*), has been shown to play an important role in signaling by both *in vivo* and *in vitro* analyses of its interaction with Gα, the alpha subunit of heterotrimetric GTP binding protein. This interaction increased the GTPase activity of Gα by approximately 25%. Eleven paralogous groups of caleosins, which comprise a total of thirty-four caleosin genes, have been assembled and identified using the*T. aestivum*GenBank EST database and ten gene family members were assembled from *Secale cereale* 454-cDNA sequences. The analysis of caleosin gene expression was assayed by RNA-Seq analysis of 454 sequence sets and members of the gene family were found to have diverse patterns of gene expression in the nine tissues that were sampled in rye and in triticale,a synthetic polyploid species derived from durumwheat(*Triticum turgidum*) and rye (*Secale cereale*).

The impact of the polyploidization event on rye genes in the triticale background was investigated using both caleosin genes and whole transcriptome comparisons. The high-throughput cDNA sequence comparison between the diploid rye and the hexaploid triticale detected suppression of expression of approximately

2% of the rye genes surveyed in the triticale. The expression of 23503 rye cDNA contigs was analyzed in 454-cDNA libraries obtained from anther, root and stem from both triticale and rye as well as in five 454-cDNA data sets created from ovary, pollen, seed, seedling shoot and stigma from triticale. Among these, 112 rye cDNA contigs were found to be totally suppressed in all triticale tissues, although their expression was relatively high in rye tissues. Suppressed rye genes were found to have strikingly low similarity to their closest BASTN matches in a current draft of the wheat genome available through the International Wheat Genome Survey Consortium, IWGSC. The comparison of rye silenced genes to wheat database revealed that 89% of rye genes silenced in triticale do not have a best match in *T. aestivum* with sequence identity higher than 90%, whereas 59% of random rye contigs had a best hit of 90% or higher in *T. aestivum.*The comparisons to the draft genomes of *Triticum.urartu*, and *Aegilops.tausshii,* the A and D genome donors to *T. aestivum*, respectively, support the previous observation. PCR assays found that 6 out of 10 candidate suppressed genes were deleted from the triticale genome.

ACKNOWLEDGMENTS

First of all, I would like to express my deepest thanks to my God, the most merciful who give me ability in all my entire life.

I would like to like to express my gratitude to my supervisors Dr. Patrick Gulick for his incredible patience to guide me through this research. It has been excellent opportunity for me to work in his lab. I also truly appreciate his endless morale and his ability to help his students.

My warmest grateful also goes to my committee members Dr. Reginald Storms and Dr. Varin Luc who gave me suggestions, constructive criticism, and excellent advice in my research.

I would like to express my sincere appreciation to all participants of this work Zhe Jun Wang, Justin Wright, Alexandra Ralevski, Ariel Donayo and Uyen Minh Pham, Dr , André Laroche, Mohammad-Reza Ehdaeivand, Yong Xu and Michele Frick and Samar Elzein. My appreciation is to Dr. Deborah Maret for her editorial of the third chapter of this thesis. I deeply appreciate them and I hope we can keep in touch. I also appreciate and thank all of my current and former lab-mates for their help.

I would like to express my gratitude to the Biology Department at Concordia University for the knowledge that I gained from the professors, staff and graduate students.

My last, but not least, gratitude is for my family and especially to my dear parents who have given me so much and who have encouraged me so much.

My warmest gratitude goes to all my friends, roommates, colleagues and my great neighbor Mustapha who have helped and supported me during my work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Bread wheat (*Triticum aestivum*) is one of the most important cereals grown world-wide and is the most widely adapted major cereal crop with respect to harsh environmental conditions (Tardif *et al.,* 2007). It is among the crops that are most tolerant to low temperatures, with the potential to survive temperatures as low as -21ºC after a period of acclimation, in which plants are exposed to low, non-freezing temperatures. It has also found to tolerate other abiotic stress conditions including salinity (Li *et al.,* 2010; Houde *et al.,* 2006) and drought (Hill *et al.,* 2013), as well as biotic stress (Singh *et al.,* 2013).The adaptation is regulated by sensing mechanisms triggering several cellular components to toleratethe negative impacts ofthese stresses. Analysis of plants has identified a large number of wheat genes that are up-regulated by environmental stress. The tolerance to environmental stresses is a coordinated action of multiple stress responsive genes that cross talk with other components of the stress signal transduction pathway. Signaling proteins may have multiple functions that depend on their cellular context. The cell-type-specific and signaling-dependent functions of these proteins are often determined by their interaction partners (Hu and Kerppola, 2005). Ca^{2+} signaling is a critical element in the response to abiotic and biotic stresses and Ca^{2+} binding proteins that function as signal sensor proteins may regulate specific protein-protein interactions, affect downstream phosphorylation cascades and regulate the response to environmental cues (Kudla *et al.,* 2010).

Gene expression studies and microarray analyses of cold acclimation identified a large number of genes that are more strongly induced or repressed in cold tolerant winter wheat than in freezing sensitive spring wheat (Gulick *et al.,* 2005). The cold acclimation induced calcium-binding protein, Ta-Clo3, formerly J900, was identified

among these and subsequent yeast two-hybrid screening studies identified a protein-protein interaction between Ta-Clo3 and a protein with high sequence similarity to the Gα subunit, TaGA1 (Tardif *et al.,* 2007). Ta-Clo3 has high sequence similarity to members of the caleosin gene family identified in several plant species but little is known about Ta-Clo3, and much less about its potential role in signaling.

Caleosins are calcium-binding proteins encoded by small gene families in plants.Caleosins comprise a gene family of seven members in Arabidopsis andsix membersin rice.In Arabidopsis, the most highly developed model system in plants, insight into a possible role of caleosins in the stress response in plants has been reported and discussed. The genes most similar to Ta-*Clo3* in Arabidopsis are the RD20/At-Clo3 (At2g33380) and At-Clo7 (At1G23240) genes which are members of a small gene family that encodes proteins which contain single EF-hand Ca^{2+}-binding domains. RD20 was first characterized as a drought-induced gene. It was shown to bind Ca^{2+} and to be expressed in aerial tissues, especially in leaves and flowers. RD20 was also induced by dehydration, salt, ABA and cold treatments (Takahashi *et al.,* 2000; Fujita *et al.,* 2004).

The second chapter of this thesis characterizes the interaction of Ta-Clo3 and wheat Gα, Ta-GA3, as well as with the phospholipase C, PI-PLC1. Both *in vivo* and *in vitro* interactions have been characterized between pairwise combinations of the three proteins.In addition, we have found that the GAP activity of Clo3 towards GA3 suggests it may play a role in the inactivation of GA3 as part of the stress response in plants. Three-way affinity characterizations with GA3, Clo3 and PI-PLC1 showed the interaction with Clo3 to be competitive with PI-PLC1, which suggests that Clo3 may play a role in the Ca^{2+}-triggered feedback regulation of both GA3 and PI-PLC1.

The third chapter of this thesis is on the characterization of the caleosin gene family in the Triticeae. This study characterized the diversity of the caleosin gene family in order to facilitate anunderstanding ofthe diverse roles of these proteins in signaling and regulation.In this investigation, we reporta description of the whole gene family in hexaploid wheat and diploid rye based on analysis of high-throughput cDNA sequencing data sets. The expression profiles of the caleosin genes in rye and triticale tissues werealso compared. Gene expression was assayed by RNA-Seq analysis of 454 sequence sets and members of the gene family were found to have diverse patterns of gene expression in the tissues that were sampled in rye and in triticale. Interestingly, the comparison of gene expression between rye and the polyploid triticale indicated that the pattern of expression of several caleosin genes from the rye genome was altered in the polyploid triticale.

The fourth chapter of this thesis is a global analysis of the effect of polyploidization on rye gene expression through comparing expression patterns in the rye parent to that of the allopolyploid triticale.Triticale is a recent man-made polyploid that parallels the generation of bread wheat except that it has rye (*Secale cereale*)as one of its progenitors and *T. turgidum* as the other. The synthetic hybrid was first developed in the early part of the 20th century, and is a useful model for characterizing the alterations in gene expression that resulted from rapid changes in genome structure and gene regulation soon after polyploidization.

Second generation DNA sequencing methods were used to compare gene expression profiles between rye and triticale in different plant tissues. Before the advent of Illumina, 454 and Solexa sequencing,researchers used other methods of global gene expression profiling, including cDNA-AFLP and microarrays, to compare the expression of homeologs in synthetic and natural allopolyploids (Kashkush *et*

*al.,*2002; He *et al.,* 2003; Adams *et al.,* 2004; Wang *et al.,* 2004; Xu *et al.,* 2009). The limitation of cDNA-AFLP is that the detection of the expression level is qualitative, so it will generally detect homeologous genes that have large differences in expression and will underestimate differences relative to more quantitative measures. Microarray studies that compare the relative expression levels between genes duplicated by polyploidy in wheat (Pumphrey *et al.,* 2009; Akhunova *et al.,* 2010), cotton (Hovav *et al.,* 2008; Rapp *et al.,* 2009) and Brassica (Gaeta *et al.,* 2009) generally reported a higher fraction of differentially expressed homeologs. Because of the quantitative nature of microarrays, the analysis parameters such the number of replicates used in the experiments, the fold-change threshold cut-off, and P-value cut-off, will influence the percentage of the genes that are reported to show differential expression. In addition, microarrays including the short oligonucleotide arrays of Affymetrix arrays may not be completely discriminating between highly similar sequences; mismatching short oligonucleotide probes may still cross hybridize to the homeologs but at a reduced level and thus the quantification may not be accurate. In contrast, estimating gene expression using second generation high-throughput DNA sequencing techniques offers the advantage of increased accuracy of transcript identification directlyfrom the sequence rather than by DNA or RNA hybridization, thus next generation sequencing data help to overcome limitations of these previous techniques.

REFERENCES

Adams KL, Percifield R and Wendel JF (2004). Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid.*Genetics*, 168: 2217-2226

Akhunova AR, Matniyazov RT, Liang H, and Akhunov ED (2010). Homoeolog-specific transcriptional bias in allopolyploid wheat.*BMC Genomics*, 11: 1-16

Feldman M, (2001). Origin of cultivated wheat. In: Bonjean, A.P., Angus, W.J.(Eds.), The World Wheat Book; a History of Wheat Breeding. Lavoisier Publishing, Paris, pp. 3–56

Fujita M, Fujita Y, Maruyama K, Seki M, Hirratsu K, Ohme-Takagi M, Tran LSP, Yamaguchi-Shinozaki K and Shinozaki K (2004). A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J.,* 39:863–876

Gaeta RT and Chris Pires J (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet.*New Phytol.,* 186: 18-28

Gulick PJ, Drouin S, Yu Z and Danyluk J (2005). Transcriptome comparison of winter and spring wheat responding to low temperature.*Genome,* 48: 913–923

He P, Friebe BR, Gill BS and Zhou JM (2003). Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant MolBiol*, 52: 401–414

Hill CB, Taylor JD, Edwards J, Mather D, Bacic A, Langridge P and Roessner U (2013). Whole genome mapping of agronomic and metabolic traits to identify novel quantitative trait loci in bread wheat grown in a water-limited environment.*Plant Physiol*. 162:1268-1281

Houde M, Belcaid M, Ouellet F, Danyluk J, Monroy AF, Dryanova A, Gulick P, Bergeron A, Laroche A, Links MG MacCarthy L, Crosby WL and Sarhan F (2006). Wheat EST resources for functional genomics of abiotic stress.*BMC Genomics,* 7: 149

Hovav R, Udall JA, Chaudhary B, Flagel L and Rapp R (2008). Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant.*Proc Natl Acad Sci* (USA) 105: 6191–6195

Hu CD, Kerppola TK (2005). Direct visualization of protein interactions in living cells using bimolecular fluorescence complementation analysis. In: Adams P, Golemis E (eds) Protein–protein interactions, vol 34. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 1–20

Kashkush K, Feldman M and LevyAA (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid.*Genetics,* 160: 1651–1659

Kudla J, Batisti O and Hashimoto K (2010). Calcium signals: the lead currency of plant information processing. *Plant Cell*, 22: 541–563

Li C, Lv J, Zhao X, Ai X , Zhu X , Wang M, Zhao S and Xia G (2010). TaCHP: A Wheat Zinc Finger Protein Gene Down-Regulated by Abscisic Acid and Salinity Stress Plays a Positive Role in Stress Tolerance. *Plant Physol.,* 154: 211–221

Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O and Gill BS (2009). Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. *Genetics,*181: 1147–1157

Rapp RA, Udall JA and Wendel JF (2009). Genomic expression dominance in allopolyploids.*BioMed Central*, 7: 1-10

Sing A, Knox RE, Depauw RM, Sing AK, Cuthbert RD, Campbell HL, Sing D, Bhavani S, Fetch T, Clark F (2013). Identification and mapping in spring wheat of genetic factors controlling stem rust resistance and the study of their epistatic interactions across multiple environments. *Theor Appl Genet* [Epub ahead of print]

Takahashi S, Katagiri T, Yamaguchi-Shinozaki K and Shinozaki K (2000). An Arabidopsis gene encoding a Ca^{+2}-binding protein is induced by abscisic acid during dehydration. *Plant Cell Physiol.*, 41: 898–903

Tardif G, Kane NA, Adam H, Labrie L, Major G, Gulick P, Sarhan F and Laliberte JF (2007). Interaction network of proteins associated with abiotic stress response and development in wheat. *Plant Mol Biol.,* 63: 703–718

Van-Ginkel M and Ogbonnaya F (2007). Novel genetic diversity from synthetic wheats in breeding cultivars for changing production conditions. *Field Crops Res.,*104: 86–94

Wang J, Tian L, Madlung A, Lee H and Chen M (2004). Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids.*Genetics,* 167: 1961–1973

Xu Y, Zhong L, Wu X, Fang X and Wang J (2009). Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta,* 229: 471–483

Chapter 2 Heterotrimeric Gα subunit from wheat (*Triticum aestivum*), GA3, interacts with the calcium-binding protein, Clo3, and the phosphoinositide-specific phospholipase C, PI-PLC1

2.1. Outline and contribution of colleagues

Chapter 2 characterizes the *in vivo* and *in vitro* protein–protein interactions between the *T. aetivum*heterotrimeric Gα subunit, GA3, the calcium-binding protein, Clo3 and the phosphoinositide-specific phospholipase C, PI-PLC1. The canonical Gα subunit of the heterotrimeric G protein complex from wheat (*Triticum aestivum*), GA3, and *T. aestivum* Clo3, were shown to interact both *in vivo* and *in vitro* and Clo3 was shown to enhance the GTPase activity of GA3. The GAP activity of Clo3 towards GA3 suggests that it may play a role in the inactivation of GA3 as part of the stress response in plants. Three-way affinity characterizations with GA3, Clo3 and PI-PLC1 showed the interaction with Clo3 to be competitive, which suggests that Clo3 may play a role in the Ca^{2+}-triggered feedback regulation of both GA3 and PI-PLC1. Most of the findings described were published in Plant Molecular Biology (Khalil *et al.,* 2011). I am the first author of this publication. I have participated in the experimental design and writing the manuscript. All *in vitro* protein-protein interaction experiment was performed by me. I have also cloned *PI-Plc* genes to study their subcellular localization using tobacco as a plant model and to investigate their interaction with both GA3 and Clo3 *in vivo* and *in vitro.* All of the microcopy which demonstrates the *in vivo* protein-protein interactions and the protein subcellular localizations presented in the manuscript were carried out by me.

2.2. Abstract

The canonical Gα subunit of the heterotrimeric G protein complex from wheat (*Triticum aestivum*), GA3, and the calcium-binding protein, Clo3, were revealed to interact both *in vivo* and *in vitro* and Clo3 was shown to enhance the GTPase activity of GA3. Clo3 is a member of the caleosin gene family in wheat with a single EF-hand domain and is induced during cold acclimation. Bimolecular Fluorescent Complementation (BiFC) was used to localize the interaction between Clo3 and GA3 to the plasma membrane (PM). Even though heterotrimeric G-protein signaling and Ca^{2+} signaling have both been shown to play a role in the response to environmental stresses in plants, little is known about the interaction between calcium-binding proteins and Gα. The GAP activity of Clo3 towards GA3 suggests it may play a role in the inactivation of GA3 as part of the stress response in plants. GA3 was also shown to interact with the phosphoinositide-specific phospholipase C, PI-PLC1, not only in the PM but also in the endoplasmic reticulum (ER). Surprisingly, Clo3 was also shown to interact with PI-PLC1 in the PM and ER. *In vitro* analysis of the protein–protein interactions showed that the interaction of Clo3with GA3 and PI-PLC1 is enhanced by high Ca^{2+}levels. Three-way affinity characterizations with GA3, Clo3 and PI-PLC1 showed the interaction with Clo3 to be competitive, which suggests that Clo3 may play a role in the $Ca2^{+}$-triggered feedback regulation of both GA3 and PI-PLC1. This hypothesis was further supported by the demonstration that Clo3 has GAP activity with GA3.

2.3. Introduction

Heterotrimeric GTP-binding proteins (G proteins) are involved in multiple signaling pathways in plants and animals initiated through G-protein-coupled receptors (GPCR). The G protein complex is composed of three subunits, Gα, Gβ and Gγ. In animals the genes encoding these subunits comprise gene families; for example, the human genome has 23 Gα, 6 Gβ, and 12 Gγ encoding genes. In contrast, plant genomes contain relatively few heterotrimeric G proteins, with the Arabidopsis genome encoding one Gα, one Gβ and two Gγ subunits. In addition, there are 37 genes encoding regulator of G-protein- signaling (RGS) proteins and about 800 genes with significant sequence similarity to GPCRs (Jones and Assmann, 2004). G-protein signaling has been implicated in the plant response to pathogens (Blumward et al., 1998), light (Lapik and Kaufman, 2003), high salinity, drought (Misra et al. 2007), hypoxia and ethylene signaling (Steffens and Sauter,, 2010), and in signaling pathways regulated by jasmonic acid (Okamoto et al. 2009), gibberellin (Ullah et al., 2003) and abscisic acid (ABA) (Ritche and Gilroy 2000; Wang et al. 2001; Pandey et al. 2006). In the classical model of heterotrimeric G-protein signaling, the receptor G-protein complex transmits signals through guanine nucleotide exchange and hydrolysis at the inside surface of the cell membrane. In the inactive state, Gα tightly binds the Gβγ subunits and a G-protein-coupled receptor (GPCR). When a ligand activates the GPCR, G proteins bound to the GPCR undergo conformational changes. The Gα subunit exchanges GDP for GTP, which results in the dissociation of the Gα subunit from the Gβγ dimer and GPCR. Gα-GTP and the Gβγ dimer can then activate downstream signaling cascades and effectors. The hydrolysis of the bound GTP to GDP by the GTPase activity of Gα allows the reconstitution of the inactive heterotrimeric complex (Sprang, 1997; Hamm, 1998). In plants, in contrast to animals,

Gα was found to have a high intrinsic GDP/GTP exchange rate that was predicted to favor a free GTP-bound resting state which suggests that GTPase activating proteins (GAPs) would play a critical role in the regulation of Gα-signaling (Johnston et al. 2007). The GAP activity of Arabidopsis RGS1 in association with GPA1 supports this notion, and is hypothesized to shift the G protein to a GDP-bound state (Chen, 2008; Jones *et al.,* 2011). The low number of genes encoding G proteins in plant genomes suggests that the multiplicity of signaling associated with G proteins is modulated through multiple interacting proteins, and it is likely that other GAPs exist in plants that are not part of the RGS-protein family. One such protein, PLDα1, in spite of its lack of anRGS-box, was shown to have GAP activity with the Arabidopsis GPA1 (Zhao and Wang, 2004). The role of effector proteins associated with Gα is an important area of research and the number of identified interacting proteins is growing. To date, five upstream G-protein-coupled receptors have been shown to interact with the Gα subunit, AtGCR1 (Pandey and Assmann, 2004), AtRGS1 (Chen et al., 2006), the pea GPCR (Misra et al. 2007), and two GPCR-type G proteins, GTG1 and GTG2 (Pandey et al., 2009). There are also few known downstream effectors of the plant Gα subunit. The Arabidopsis Gα (GPA1) interacts with Atpirin1, a cupin domain protein, which regulates seed germination and seedling development (Lapik and Kaufman, 2003), phospholipase D (PLDα1) (Zhao and Wang, 2004), prephenate dehydratase protein (PD1) (Warpeha et al., 2006), and the plastid protein thylakoid formation 1 (THF1) (Huang et al., 2006). Gαhas been shown to interact with phospholipase A2 (PLA2) in *Eschscholzia californica* (Heinze et al., 2007) and phospholipase C in *Pisum sativum* (Misra et al., 2007).

Bread wheat (*Triticum aestivum*) is one of the two major cereals worldwide and is the major crop that is most widely adapted to harsh environmental conditions

(Tardif et al., 2007). It is among the few crops that are tolerant to low temperatures, with the ability to survive at temperatures as low as -21°C expression studies and microarray analyses of cold acclimation identified a large number of genes that were more strongly induced or repressed in cold tolerant winter wheat than in freezing sensitive spring wheat (Gulick et al., 2005; Monroy et al., 2007). A cold acclimation-induced calcium binding protein, Clo3, formerly J900, was identified among these and subsequent yeast two-hybrid screening studies identified a protein–protein interaction between Clo3 and a protein with high sequence similarity to the Gα subunit, GA1 (Tardif *et al.,* 2007). This suggested that the Gα protein may be involved in signaling pathways regulating cold acclimation. This was one of the first reports of a protein–protein interaction between a plant Gα and calcium-binding protein. Little is known about Clo3, and much less about its potential role in signaling. Its closest homologs in Arabidopsis are RD20/AtClo3 (At2g33380) and AtClo7 (At1G23240), which are members of a small gene family that contains single EF-hand Ca^{2+} binding domains. RD20 was first characterized as a drought-induced gene that was shown to bind $Ca^{2+.}$ It was expressed in aerial tissues, mainly in the leaves and flowers. RD20 was also induced by dehydration, salt, ABA and cold treatments (Takahashi et al., 2000; Fujita et al., 2004).

In animal systems, activated Gα proteins are known to activate effector enzymes including phosphoinositide-specific phospholipase C (PI-PLC), which cleaves the phosphodiester bond of phosphatidylinositol 4,5-bisphosphate (PIP2) in the inner leaflet of the plasma membrane (PM), releasing inositol triphosphate (IP3) and diacylglycerol (DAG). IP3 is a primary effector of calcium release, interacting with receptors on intracellular calcium storage sites to release calcium into the cytoplasm (Berridge, 1993). The existence of homologous genes for these proteins

suggests that similar signaling also functions in plants. The PI-PLCs identified in plants are closely related to mammalian PLCδ (Munnik et al., 1998); they contain the two subdomains, X and Y, that are necessary for phosphoesterase activity and a calcium-binding C2 domain. PI-PLCs have been shown to play a role in signaling involved in disease resistance in tomato (Vossen et al., 2010). Misra et al. (2007) demonstrated that PLCδ functions as an intracellular signaling enzyme for the Gα-subunit of pea (*Pisum sativum*), and that Gα interacts with the calcium-binding C2 domain of the pea PLCδ.

Signaling proteins may have multiple functions that depend on their cellular context. The cell-type-specific and signaling-dependent functions of these proteins are often determined by their interaction partners (Hu and Kerppola, 2005). Ca^{2+} plays a critical role in abiotic and biotic responses, and Ca^{2+}-binding proteins that function as signal sensor proteins may regulate specific protein–protein interactions, affect downstream phosphorylation cascades and regulate the response to environmental cues (Kudla et al., 2010).

The present study was conducted to characterize *in vivo* and *in vitro* protein–protein interactions between the wheat heterotrimeric Gα subunit, GA3, the calcium-binding protein, Clo3 and the phosphoinositide-specific phospholipase C, PI-PLC1.

2.4. Materials and Methods

Plant materials and growth conditions

Tobacco seeds (*Nicotiana benthamiana*) were germinated in 10 cm pots in a potting mixture with equal volumes of peat moss, vermiculite and soil. The plants were grown for2–4 weeks in the greenhouse with supplemental light to extend the day length to 16h light/8h dark at 20^{o}C.

Ga3, Clo3, Pi-Plc1 and Pi-Plc2 full length cDNA cloning and expression clones

The full length cDNA clone of Ga3-CS, from wheat cultivar Chinese Spring, encoding the Gα subunit of wheat was obtained from the Arizona Genomics Institute (AGI). This cDNA clone had a 17 bp duplication 91 nucleotides upstream of the normal stop codon. The intact version of the Ga3 cDNA was generated by combining the 358 nucleotides of the 3‘end of the cDNA PCR-amplified from a cDNA library of cultivar Norstar and 1027 nucleotides from the 5‘end of the Ga3-CS cDNA. Gene-specific primers for the cloning are listed in Supplemental Table S1. A full-length Clo3 cDNA, from the *T. aestivum* cultivar Norstar, was obtained from the Functional Genomics of Abiotic Stress (FGAS) cDNA clone collection (Houde et al., 2006). The partial cDNA clone of PI-Plc1 corresponding to GenBank EST accession GI: 55684870, also from the FGAS clone set, had a 1192 nucleotide ORF, but lacked its 5’ end. The 5’ end of the cDNA was cloned from a cDNA library by PCR amplification using gene-specific and vector-specific primers (Supplemental Table S1) and two rounds of cloning were carried out to obtain the missing portion of the cDNA. The full ORF of PI-Plc1 was subcloned into pDONR207 by Gateway® BP clonase II Enzyme mix (Invitrogen) and subsequently transformed into the TOP10 *E. coli* strain. The coding regions of Ga3, Ga3^{Q223L}, Ga3-CS, Clo3, Pi-Plc1, Pi-Plc2, Pi-Plc2ΔEF-

hand, Pi-Plc2ΔC2 and C2 domain of Pi-Plc2 were cloned as fusions with fluorescent proteins in plant expression vectors using the Gateway® cloning system (Invitrogen). Gateway®LR reactions wereused to transfer the inserts of the entry clones of Ga3, $Ga3^{Q223L}$, Ga3-CS, Clo3, Pi-Plc 1, Pi-Plc 2, Pi-PlcΔEFhand Pi-Plc2ΔC2 and C2 domain to the plant destination binary vector, PK7FWG2, to generate enhanced Green Fluorescent Protein (eGFP) C-terminal fusions. They were also transferred to BiFC binary Yellow Fluorescent Protein (YFP) vectors, pBatL-B-sYFP-N and pBatL-B-sYFP-C.

For protein expression in *E. coli,* PI-PLC1 and Clo3 were cloned in pDEST17, a Gateway® vector with a 6-His-tag N-terminal fusion (Invitrogen). PI-PLC1 was also cloned into pDEST15, a Gateway vector with a GST N-terminal fusion. Ga3 was amplified using a pair of primers, digested with BamHI and SalI (Supplemental Table S1) and directionally cloned into a modified pGEX-2T, an N-terminal GST-fusion expression vector to which a SalI restriction site had been added.

Agrobacterium transformation and agroinfiltration

Electrocompetent *Agrobacterium* strain AGL1 was transformed with plant gene expression constructs as follows: overnight cultures were centrifuged at 4000 x*g* for 15 min. at 4° C and resuspended in 10 mM $MgCl_2$ and 150 μM acetosyringone to an OD_{600} of 0.1 and incubated at room temperature for 2 hours (Walter *et al.,* 2004). The leaf infiltration suspensions were a mixture of equal volumes of three or four *Agrobacterium* culture suspensions containing expression vectors for the experimental samples, the mCherry fluorescent cellular marker proteins (Nelson *et al.,* 2007) (Supplemental Table S2) and a culture expressing P19 of tomato bushy stunt virus, to suppress gene silencing (Voinnet *et al.,* 2003). The agroinfiltration solution was co-infiltrated into the leaf abaxial air space of two to four week old *N. benthamiana*

plants. The plants were incubated in environmental growth chambers under long dayconditions(16h light / 8h dark)at 20º C for two to four days and subsequently analyzed by microscopy.

Confocal laser scanning microscopy

The epidermal tissues of tobacco leaves were examined using the Spinning Disk2 (SD2) confocal microscope, at The Cell Imaging and Analysis Network (CIAN) laboratory, McGill University. GFP was excited at a wavelength of 491 nm by the diode laser and the emitted fluorescence was collected through a 520/535 nm band-pass. YFP was excited at the same wavelength laser, and the emitted fluorescence was collected through a 543 nm long-pass filter. The Cherry Red Fluorescent Protein (cRFP) was excited at a wavelength of 561 nm and the emitted fluorescence was collected through a 624/640 nm long-pass filter.

***In vitro* protein-protein interaction between PI-PLC1, GA3 and Clo3**

For protein-protein interaction characterization and expression in *E. coli,* BL21 cells transformed with the recombinant plasmids were grown at 37^{o}C with 100 µg/ml ampicillin for 6 hours and then induced with 1 mM IPTG for 12 hours at room temperature. Cultures expressing GA3-GST were lysed in buffer: 50 mM Tris-HCl, pH 8.8, 100 mM NaCl, 2% Triton X-100, 5% Glycerol and 1 mM PMSF. The recombinant GA3-GST protein was purified with Glutathione Sepharose^{TM} (GE Healthcare Life Sciences), according to the manufacturer's protocol. Cultures expressing PI-PLC1-His or Clo3-His were lysed in the same buffer, centrifuged at 12,000 x*g* for 15 min. and protein was recovered from the pellet by solubilization in 50 mM Tris-HCl, pH 8.8, 0.5% SDS and 5% Glycerol, 1 mM PMSF. PI-PLC1 and

Clo3 were purified using Ni-NTA Agarose (Qiagen) according to Todorova (2009) with some modifications: the Ni-NTA Agarose was equilibrated with the wash buffer containing 50 mM Tris-HCl, pH 8.8, 0.05% SDS, and 1 mM PMSF. The protein sample was immobilized on the equilibrated beads, subsequently rinsed twice with wash buffer and then eluted with 2X SDS-PAGE buffer (100 mM Tris-HCl, pH 7.0, 200 mM DTT, 4% (W/V) SDS, 0.2% (W/V) bromophenol blue, 30% glycerol and 100 mM imidazole).

The His-tag pull-down assay (Todorova, 2009) was also used to study the pairwise interaction among the three proteins, GA3, PI-PLC1 and Clo3. A mixture of 50 µl of crude extract of the Clo3 or PI-PLC1 culture and 50 µl of Ni-NTA Agarose beads, previously equilibrated with the wash buffer (50 mM Tris-HCl, pH 8.8, 0.05% SDS and 1 mM PMSF), were incubated for two hours at 4° C with gentle shaking. The mixture was rinsed with the wash buffer, centrifuged at 2000 xg for 1 min and the supernatant was discarded. The Ni-NTA Agarose beads bound with PI-PLC1-His or Clo3-His proteins were incubated with 100 µl of GA3-GST crude *E. coli* lysate in 50 mM Tris-HCl, pH 8.8, 100 mM NaCl, 2% Triton X-100, 5% Glycerol, 1 mM PMSF and either 10 mM $CaCl_2$ or 1 mM EGTA for two hours at 4° C with gentle shaking. Samples were centrifuged at 2000 xg, washed twice with wash buffer, then eluted in 50 µl of 2X SDS-PAGE buffer. To assay GTP/GDP bound forms of Gα, GA3 lysate was pre-treated with 5 mM EDTA and 5 mM of either GDP or GTP for 10 min. followed by the addition of 20 mM $MgCl_2$.

The three-way interaction between Clo3-His, GA3-GST, and PI-PLC1-GST was studied by first binding Clo3-His to Ni-NTA Agarose beads, then GTP-loaded GA3-GST in 50 mM Tris-HCl, pH 8.8, 100 mM NaCl, 2% Triton X-100, 5% Glycerol, 1 mM PMSF buffer was added at levels to achieve saturation of the binding

to Clo3-His. Subsequently-increasing amounts of PI-PLC1-GST (in buffer: 50 mM Tris-HCl, pH 8.8, 0.5% SDS, 5% Glycerol, 1 mM PMSF) were added to the interaction mixture. Bound protein was rinsed with wash buffer and eluted with 2X SDS-PAGE buffer and samples were analyzed by SDS-PAGE electrophoresis on 12% polyacrylamide gels subsequently stained with Coomassie Brilliant Blue R-250.

GTPase assay for GA3 and Clo3

GA3 and Clo3 proteins were expressed in *E. coli* as described above. GA3-GST was purified as described in Willard and Siderovski (2004), with modified lysis buffer: 25 mM Tris–HCl, pH 7.6, 100 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 30 µM $AlCl_3$, 20 mM NaF, 5 mM PMSF, 50 µM GDP, 1 mM DTT, and 39 Complete EDTA-free protease inhibitor (Roche) at 3 tablets/l. Subsequently, 1 mg/ml lysozyme and 20 mM $MgCl_2$ were added and the solution was rocked for 30 min and sonicated for 2 min 200 mM NaCl was added, and the solution was rocked again for 30 min. The sample was centrifuged for 45 min at 13,000 xg and the recombinant GA3-GST protein was purified from the supernatant with Glutathione Sepharose™ 4B beads (GE Healthcare Life Sciences) according to the manufacturer's protocol by incubation for 90 min at 4° C with rocking. The matrix was rinsed twice in a chromatography column (Bio-Rad) with 137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, and 1.76 mM KH_2PO_4 and eluted with 25 mM glutathione, 50mM Tris–HCl, pH 8.0, and 150 mM NaCl. Clo3-His was isolated from a bacterial pellet from a 1-liter culture with 10 ml of B-PER lysis buffer (Pierce) and centrifuged at 13,000 xg for 30 min. Clo3 protein was purified from the supernatant using Ni–NTA Agarose (Qiagen) according to Pandey *et al.* (2009) and stored in aliquots of 100 µl at -80° C. Purified GA3 protein was incubated with 2mM EDTA for five min, and GTP was added to 5mM final concentration and incubated for 30 min at 4° C. Both GA3 and Clo3 protein samples

were buffer-exchanged three times in 50mM Tris–HCl, pH 7.5, and 100 mM NaCl at 4^{o} C, with Viva Spin 500 columns (Stedim) according to the manufacturer's protocol. GTPase reactions were performed in the exchange buffer with GA3 at 49 μM and Clo3 at 48 μM with 10 μM GTP and 10 mM $CaCl_2$; the reaction was activated with 10 mM $MgCl_2$ and incubated for 30 min at room temperature. Inorganic phosphate, the GTPase reaction product, was assayed using the PiPer Phosphate assay kit (Invitrogen) according to the manufacturer's protocol, with samples being analyzed over a 1 h time course.

2.5. Results

The heterotrimeric Gα subunit, Ga3, calcium-binding protein, Clo3, and phosphoinositide-specific phospholipase C, Pi-Plc, genes from wheat

The Ga3 cDNA from the wheat cultivar Norstar has 98% nucleotide sequence identity with the coding regions of two previously reported wheat Gα genes from the cultivar S615, GA1 and GA2 (Hossain *et al.,* 2003). Ga3 likely represents a homeologous copy of GA1, whereas GA1 and GA2 appear to be recent gene copies with small rearrangements. GA1 is 99.4% identical to GA2 but GA2 has a 21 bp insertion. A second version of Ga3 was found as a cDNA clone from the wheat cultivar Chinese Spring, Ga3-CS, has a 17 bp insertion located 1038 bp downstream from the start codon which causes a frame shift and results in an ORF encoding a protein with 367 amino acids instead of the 382 amino acids of the normal allele, Ga3. In addition, the last 20 amino acids of the shortened protein are different from a similar clone identified in the cultivar Norstar. Except for the insertion, the nucleotide sequence is 100% identical to the Ga3 clone from Norstar and thus it is likely an allele of Ga3; it is referred to as Ga3-CS. $GA3^{Q223L}$, a constitutive GTP-bound form of GA3, was also constructed by replacing adenosine with thymine at position 668 of the Ga3 coding region according to Ullah *et al.* (2003) by PCR-based site directed mutagenesis to create a glutamine to lucine change. A cDNA clone for Clo3 (previously referred to as J900) (Tardif *et al.,* 2007) which belongs to a caleosin superfamily, has high sequence similarity to caleosin-like genes identified in wheat and Arabidopsis. It has 63% amino acid sequence similarity with At-Clo3 of Arabidopsis (also known as RD20), a calcium-binding protein gene family member that is strongly induced by drought and ABA treatments (Partridge and Murphy 2009). Clo3 contains an EF-hand motif (Supplemental Fig. S1), has 76% amino acid

sequence similarity with a Ca^{+2}-binding protein in rice (GB acc. BAD45228) and 69% amino acid sequence similarity with a barley Ca^{+2}-binding protein (GB acc. CAB71337). cDNA clones encoding phosphoinositide-specific phospholipase C protein, Pi-Plc1 and Pi-Plc2 cDNAs were identified among a FGAS wheat EST clone collection (Houde *et al.,* 2006). The full-length cDNA clone of Pi-Plc1 was derived from the combination of a partial-length cDNA and the 5' end of the cDNA including the 5' UTR and the independently cloned 5' end of the ORF of PI-Plc1. This 5' end was cloned by PCR amplification from a plasmid mixture of a complete cDNA library using gene specific and vector specific primers.The 1860 nucleotide cDNA clone encodes a 506 amino acids protein with a molecular mass of 57 kDa that has 80% amino acid sequence identity with *Oryza sativa* PI-PLC1 (GB acc. AAK01711). PI-PLC1 contains X and Y domains that are necessary for the phosphoesterase activity as well as a C2 calcium-binding domain, but does not have an EF-hand calcium-binding domain in the N-terminal region of the protein that is found in many plant PI-PLCs. There are other examples of PI-PLC genes that lack an N-terminal EF-hand domain from *Zea mays* (Zhai *et al.*, 2005) and *Vigna unguiculata* (GB acc. AAB41107) (El-Maarouf *et al.,* 2001). The full-length cDNA sequence of Pi-Plc2 is 1761 nucleotides long and encodes a 586 amino acid protein. It has 59% amino acid identity to Pi-Plc1 and contains an EF-hand in the N-terminal region, the X and Y phosphoesterase domains, as well as the C2 calcium-binding domain in the C-terminal region of the protein.

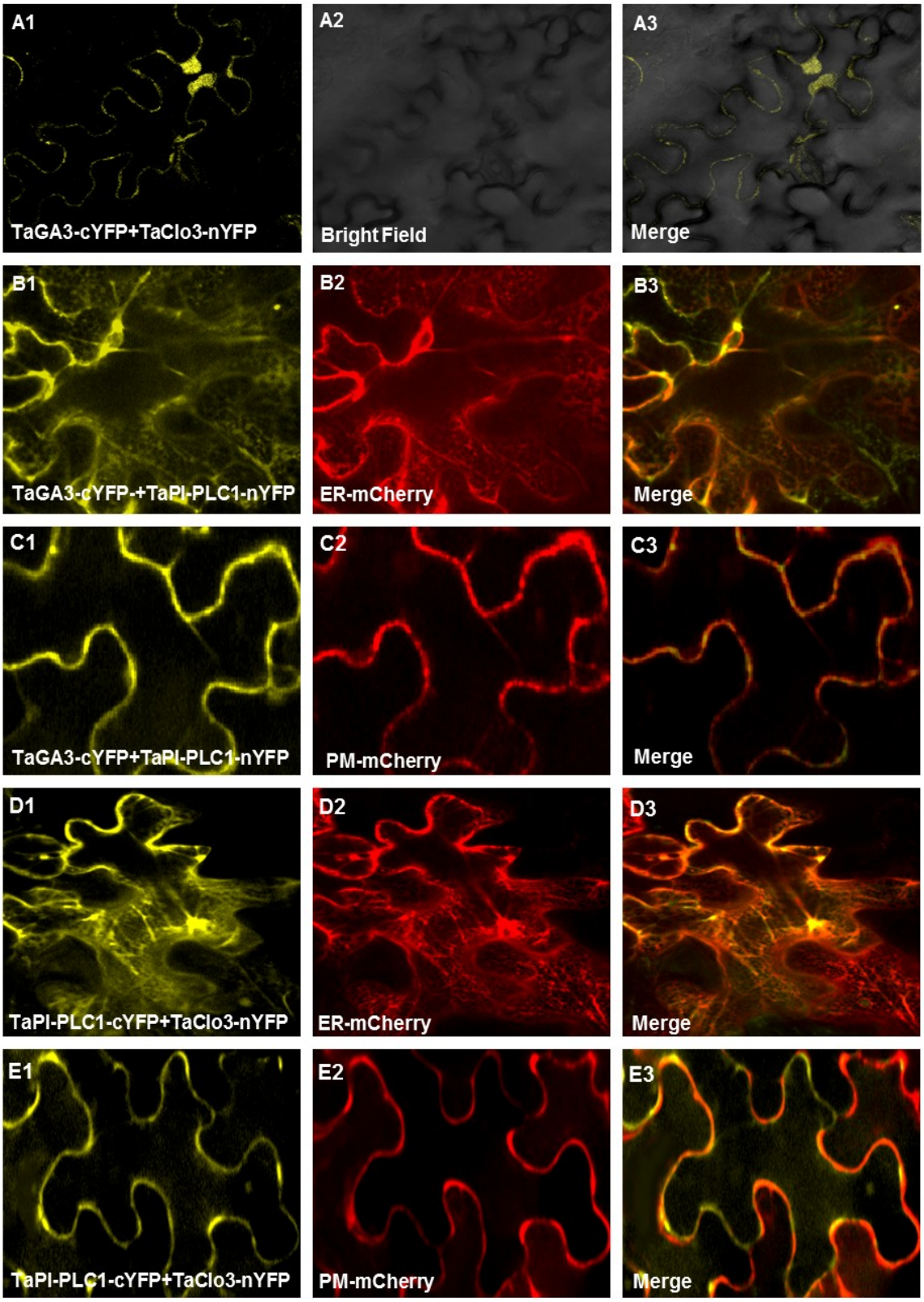


Fig. 1 BiFC interactions of TaGA3, TaClo3 and TaPI-PLC1 in tobacco epidermal leaf tissue.

A: Co-localization of TaGA3-CYFP and TaClo3 -NYFP interaction to the PM; Scale bar = 20 μM; B: co-localization of TaGA3-CYFP and TaPI-PLC1-NYFP interaction to the ER; C: co-localization of TaGA3-CYFP and TaPI-PLC1-NYFP interaction to the PM; D: co-localization of TaPI-PLC1-CYFP and TaClo3-NYFP interaction to the ER; E:co-localization of TaPI-PLC1-CYFP and TaClo3-NYFP interaction to the PM; Scale bar = 24 μm

GA3 interacts with Clo3

The three-dimensional reconstruction of the confocal images of the BiFC interaction assay of Clo3 and GA3 clearly show a pattern of localization with the PM (Fig. 1a) and not with any other structures. Though tonoplast-localized proteins would also be expected to localize to the periphery of the cell, they are also expected to be seen in internal membranes due to the presence of multiple adjacent vacuoles in a cell; however, this was not observed. The BiFC interaction assay of Clo3 with the constitutively GTP-bound form of GA3, $GA3^{Q223L}$, and the truncated version of GA3, GA3-CS, did not show significant fluorescence (data not shown). The specificity of the BiFC assay was supported by positive and negative controls including the interaction of Arabidopsis Gα, At-GPA1-C-terminal-YFP, and At RGS1-N-terminal-YFP (Grigston *et al.,* 2008), which showed a clear interaction in the PM. No interaction was observed between negative controls, Arabidopsis Ptdlns synthase, At-PIS-C-terminal-YFP, and an Arabidopsis ABA-responsive protein, At HVA22d-N-terminnal-YFP. GA3-eGFP and GA3-CS-eGFP fusions to full length eGPF proteins were localized to the PM (Fig. 2a, b) and ER (Supplemental Fig. S2a, S2b) in tobacco epidermal cells. GA3-eGFP and GA3-CS-eGFP did not co-localize with any of the other 5 cellular makers (Supplemental Fig. S3, S2).In contrast to the PM localization of the interaction between Clo3 and GA3, Clo3-eGFP expressed in *N. benthamiana* epidermal cells as a fusion to the full-length eGFP was localized to the ER and tonoplast (Fig. 2c, d, Supplemental Fig. S4). The localization to the ER is seen by both co-localization with the ER marker and by the clear network structure that is seen in the upper focal planes of the cell. In mid-focal planes the labeling appears as a punctuate signal seen at the edges of the cells, which is expected in the cross-section of a network. The punctuate label co-localized with ER makers and was

distinguishable from the PM and tonoplast makers. Gα in Arabidopsis has previously been reported to localize to the PM and the ER (Weiss et al., 1997) and Clo3 homologs in Arabidopsis and Brassica have been localized to the ER (Hernandez-Pinzon et al., 2001). The difference between the localization of Clo3-eGFP and

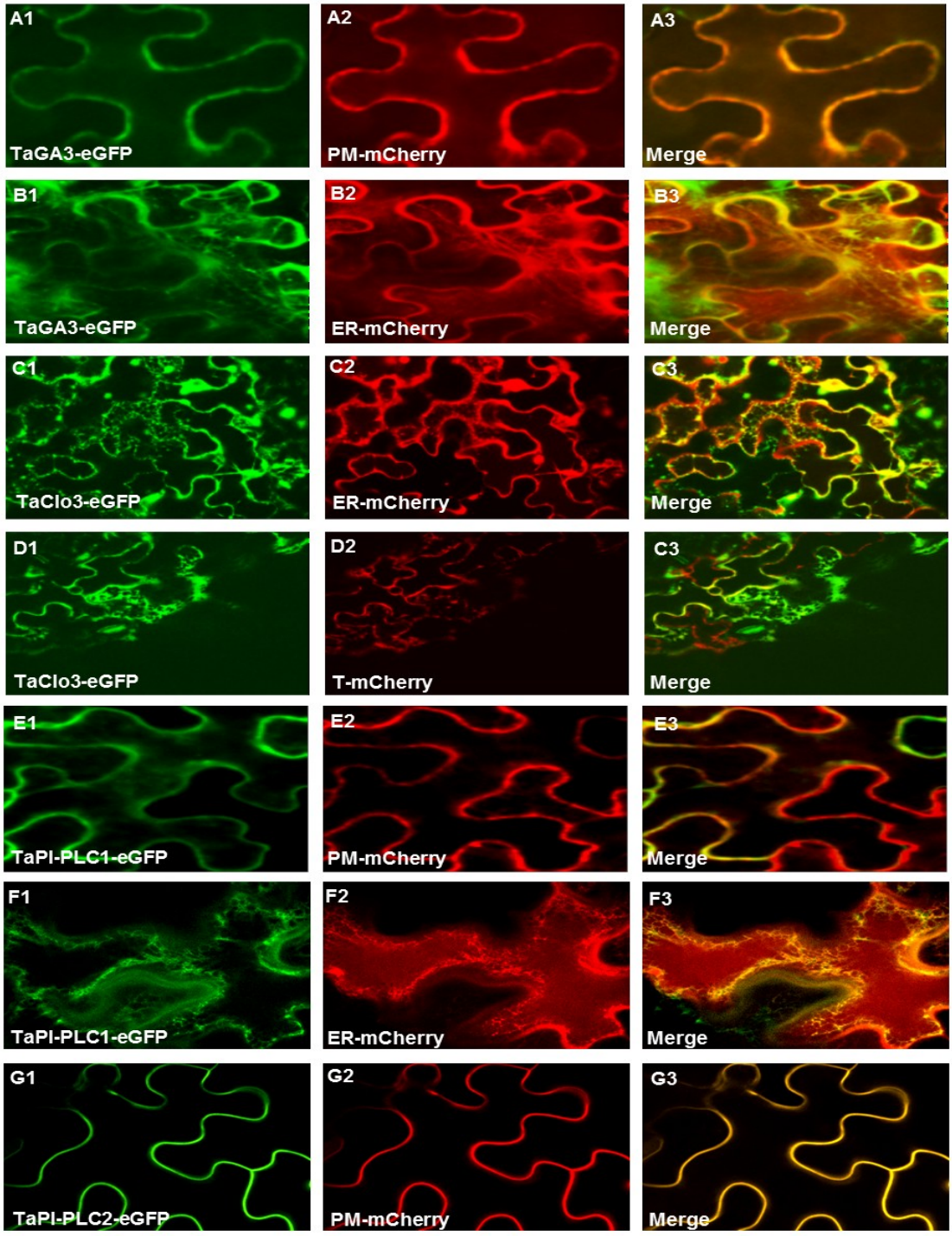


Fig. 2 Subcellular localization of TaGA3, TaClo3, TaPI-PLC1 and TaPI-PLC2 eGFP fusions in tobacco epidermal leaf tissue.
A and B: co-localization of TaGA3-eGFP to PM and ER; C and D: co-localization of TaClo3-eGFP to ER and tonoplast; E and F: co-localization of TaPI-PLC1-eGFP to PM and ER; G: co-localization of TaPI-PLC2-eGFP to PM; Scale bar = 24 μm

the localization of its interaction with GA3 suggests that there is a small subpopulation of Clo3 in the PM that interacts with GA3 that is not visible when the protein is expressed as a Clo3-eGFP construct.

To test for*in vitro* protein–protein interaction between GA3 and Clo3, Clo3 expressed as a His-tagged protein fusion in *E. coli* was immobilized on Ni–NTA Agarose beads and incubated with GA3 expressed as a GST-fusion. The interaction of GA3 and Clo3 was shown *in vitro* by a His-tag pull-down assay of Clo3, which resulted in the co-purification of GA3-GST (Fig. 3a). There was approximately a five-fold higher ratio of the GA3-GST/Clo3 recovery in the presence of 10 mM $CaCl_2$, than when proteins had been pretreated with 1 mM EGTA to remove bound calcium ions. GA3 interacted more strongly with Clo3 when it was preloaded with GTP than when it was preloaded with GDP in the absence of calcium ions, whereas no difference was found in the affinity between the GTP- or GDP-loaded forms of GA3 in its binding to Clo3 in the presence of 10 mM calcium ions (Fig. 3a).

The specificity of the interaction between GA3 and PI-PLCs

The interaction of GA3 and PI-PLC1 was observed by a BiFC assay as a reconstitution of YFP, and was localized to the ER and PM (Fig. 1b, c). The subcellular localization of the fusion to the full-length eGFP, PI-PLC1-eGFP, was also seen on the PM and ER (Fig. 2e, f, Supplemental Fig. S5). In contrast, the second phospholipase C, PI-PLC2, did not show any interaction with GA3 in the same assay. The PI-PLC2-eGFP fusion was localized predominantly to the PM (Fig. 2g, Supplemental Fig. S6). The localization of three truncations of PI-PLC2 was also carried out. PI-PLC2ΔEF-hand-eGFP, which lacked an EF-hand, was detected on the PM.This was similar to the full length construct; however, the level of expression

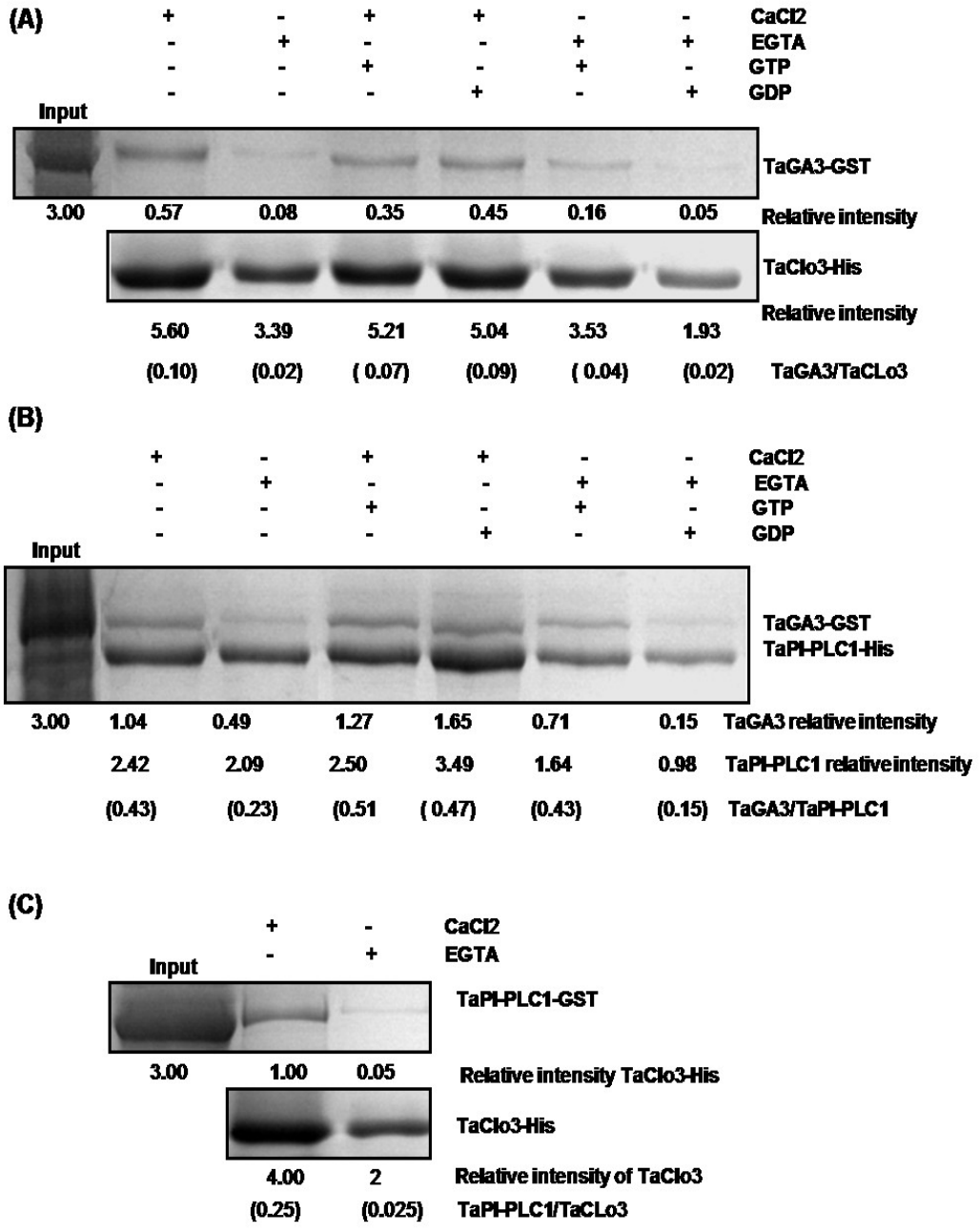


Fig. 3 The pair wise *in vitro* protein-protein interactions of TaGA3, TaClo3 and TaPI-PLC1 using His pull-down assay.

(A) TaClo3-His (26KD) was immobilized on Ni NTA-agrose beads and incubated with TaGA3-GST (70KD) in the presence of EGTA,GDP, GTP or $CaCl_2$.

(B) TaPI-PLC1-His (59KD) was immobilized on Ni NTA-agrose beads and incubated with TaGA3-GST (70KD) in the presence of EGTA,GDP, GTP or $CaCl_2$.

(C) TaClo3-His (26K D) was immobilized on Ni NTA-agrose beads and incubated with TaPI-PLC1-GST (83 KD) in the presence of either $CaCl_2$ or EGTA.

appeared to be lower than the full-length version of PI-PLC2 (Supplemental Fig. S7a). Both PI-PLC2ΔC2 (which lacked the C2 domain) and C2 domain fused to eGFP, were not targeted to the PM but were found on the ER (Supplemental Fig. S7b). Though some deletions affected the localization of the PI-PLC2 protein to the PM, none of the truncated versions of PI-PLC2 showed interaction with GA3. PI-PLC1 was also found to form homopolymeric complexes on the ER and PM when two fusion products, one as a fusion with the N-terminus of YFP and the other with the C-terminus of YFP were expressed together in the BiFC assay (Supplemental Fig. S8a). On the other hand, PI-PLC2 did not show evidence of dimerization in the same assay. The difference in the interaction and localization of PI-PLC1 and PI-PLC2 demonstrates the specificity of the interaction assay and suggests that the two proteins play different roles in plant metabolism and signaling.

In vitro binding experiments were also performed to confirm the interaction between GA3 and PI-PLC1. PI-PLC1 was expressed as a His-tagged protein fusion in *E. coli* and Ni–NTA Agarose-bound-PI-PLC1 protein was incubated with GA3 expressed as a GST-fusion. GA3 was co-purified with PI-PLC1-His when it was isolated in a His-tag pull-down assay using Ni–NTA Agarose beads (Fig. 3b). The recovery of GA3 from the *in vitro* interaction between GA3 and PI-PLC1 was approximately two times higher in the presence of 10 mM $CaCl_2$ than when calcium was sequestered by the addition of 1 mM EGTA to the interaction solution.

When the His-tag pull-down assay was employed after preloading GA3 with either GDP or GTP, the GTP-bound form of GA3 showed three times more binding with PI-PLC1 than the GDP-bound form of the protein in the absence of calcium ions. In the presence of calcium there was no difference between the interactions of PI-PLC1 with the GA3-GTP- or GA3-GDP-bound forms (Fig. 3b). The BiFC *in*

*vivo*interaction between a constitutive GTP bound form of wheat Gα, GA3Q223L, and PI-PLC1 also appeared to have higher levels of fluorescence than the interaction between the wild type GA3 and PI-PLC1 (Supplemental Fig. S8b, S8c).

PI-PLC1 also interacts with Clo3

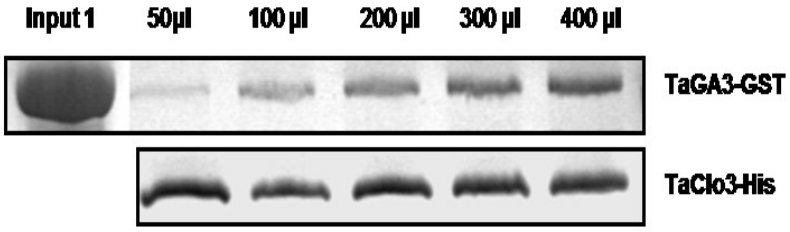
PI-PLC1 was also found to interact with the calcium binding protein, Clo3, on the ER and PM (Fig. 1d, e). The localization of the interaction between the two proteins was seen by co-localization with PM and ER markers and by the labeling of the ER network structure in the upper focal planes of the cell (Supplemental Fig. S8d). A second phospholipase C, PI-PLC2, did not show any interaction with Clo3. The interaction between PI-PLC1 and Clo3 was verified by an *in vitro* pull-down assay; Clo3 expressed as a His-tagged protein was effective in co-purifying a PI-PLC1-GST-fusion protein in the presence of 10 mM $CaCl_2$ (Fig. 3c). The interaction between the two proteins was ten times higher in the presence of 10 mM calcium ions than when calcium ions were sequestered by the addition of EGTA.

The PI-PLC1-GA3-Clo3 interactions are competitive

Since Clo3 was found to interact with both GA3 and PI-PLC1, the dynamics of the three-way interaction were investigated by *in vitro* interaction studies to assess whether the interaction was synergistic or competitive. Clo3-His-tag-bound Ni–NTA Agarose beads were incubated with increasing amounts of GA3 preloaded with GTP in the presence of 10 mM $CaCl_2$, to reach saturation of GA3 binding to Clo3 (Fig. 4a). With GA3 at the saturation level, increasing amounts of PI-PLC1-GST were subsequently added to the interaction complex of Clo3 and GA3. The addition of PI-PLC1 resulted in decreased binding of GA3 to Clo3, while increasing amounts of PI-

PLC1 were recovered as binding partners of Clo3 (Fig. 4b). This indicates competitive binding between the three proteins.

(A)



(B)

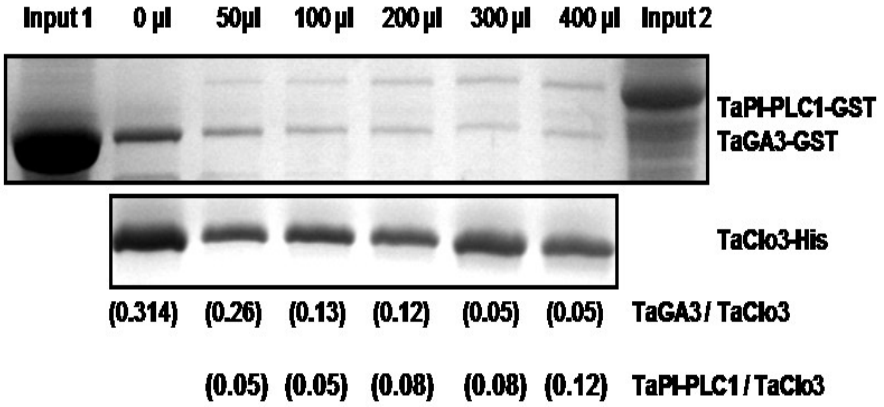


Fig. 4 The three proteins, TaGA3, TaClo3 and TaPI-PLC1, *in vitro* interaction.

(A) TaClo3-His immobilized on Ni NTA-agrose beads was incubated for 2 hours at 4oC with 50, 100, 200, 300 and 400 µl of TaGA3-GST preloaded GTP in lysate with 10 mM $CaCl_2$.

(B) TaClo3-His -TaGA3-GST preloaded GTP complex immobilized on Ni NTA-agrose beads were incubated for 2 hours at 4oC with 50, 100, 200, 300 and 400 µl of TaPI-PLC1-GST in lysate with 10 mM $CaCl_2$.

Clo3 has GAP activity

Calcium release is a downstream signaling consequence of PLC activation, and the hydrolysis of GTP is the critical regulatory step in inactivation of G proteins. Consequently, the role of the Clo3 interaction with GA3 was further investigated by characterizing the effect of protein–protein interactions on the GTPase activity of GA3. GA3-GST and Clo3-His-fusion proteins were purified from *E. coli* cultures and assayed for GTPase activity in the presence of calcium. GA3-GST alone showed intrinsic GTPase activity *in vitro* and the activity was stimulated 25% by the equimolar addition of Clo3-His in the presence of calcium as shown in Fig. 5. Clo3 alone showed no GTPase activity.This level of GAP activity was similar to that reported for the Arabidopsis PLDα1 (Zhao and Wang, 2004).

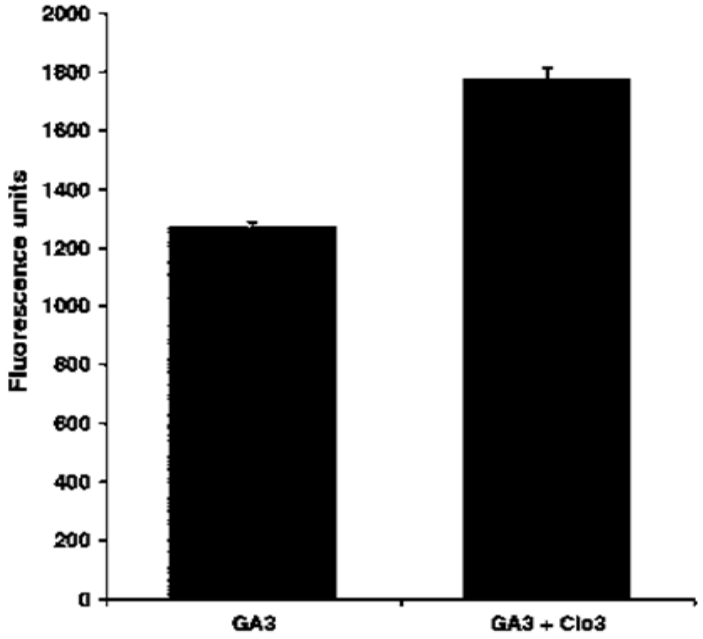


Fig. 5 Clo3 stimulation of GA3 GTPase activity. GA3-GST and Clo3-His were expressed in *E. coli* and purified with affinity matrices for the GST and 6-His-tags, respectively. The proteins were incubated in equimolar amounts with 10 lM GTP in the presence of 10 mM$CaCl_2$ for 30 min. GTPase activity was assayed by measuring Pi-released in the reaction mix and measured by a fluorescence emission assay. Clo3 showed no GTPase activity when it was assayed alone. (assay performed by Justin Wright)

2.6. Discussion

Clo3 interaction with GA3

The low number of Gα, Gβ and Gγ genes in plants as compared to animals raises questions of how heterotrimeric G proteins might be involved in multiple signaling pathways in plants. The additional gene copies and allelic variations for Gα in wheat reported here indicate a greater complexity and divergence within this gene family than in Arabidopsis and rice, phenomena reflected also in the recent analysis of the gene family in soybean (Bisht *et al.,* 2011). However, the complexity of the family is modest relative to that in animal species. The complexity of signaling in plants may be derived from a variety of effector molecules that interact with the heterotrimeric G proteins (Assmann, 2005; Pandey *et al.,* 2010); thus the analysis of the interaction between Gα, PI-PLC and Clo3 from wheat makes an important contribution towards understanding the complexity of protein–protein interactions involved in G-protein signaling. Studies in non-plant systems have shown that heterotrimeric G-protein signaling is intrinsically linked to calcium signaling by activation of PI-PLCs, which in turn leads to calcium release into the cytoplasm. In spite of the complex array of signaling pathways in which both calcium and G proteins are implicated, the direct interaction between calcium-binding proteins and heterotrimeric G proteins has not been reported in plants. In non-plant systems, there are few examples of calcium-binding proteins known to interact with G proteins directly. The calciumsensing receptor (CaR) in human pituitary cells is a G-protein receptor (Mamillapalli and Wysolmerski, 2010), and the rat calcium-binding protein, calnuc, binds to Gα, though its function is not clear (Kanuru *et al.,* 2009). More significantly, the avian regulator of G-protein-signaling, RGS3, which has the hallmark EF-hand domain of calcium-binding proteins, was shown to bind Gα and interrupt Gα-protein-mediated-signaling

in neural cells in a Ca^{+2} -dependent manner (Tosetti *et al.,* 2003). With so few calcium-binding proteins known to interact with Gα, the possible role of the calcium-binding protein Clo3 interacting with Gα is intriguing. $Ca^{+2}$is a central regulator in cell physiology and plays an important role in the response to abiotic stress including low temperature, salt and water stress. Release of Ca^{+2} from the apoplast or intercellular stores to the cytoplasm gives rise to signature patterns of oscillations of cytosolic Ca^{+2} concentrations in response to different environmental cues (Dodd *et al.,* 2007). Several classes of calcium-binding regulatory proteins including calcium-dependent protein kinases (CDPK), calcium-regulated transcription factors, calmodulins and calcineurins have been implicated in the response to abiotic stress (reviewed by Tuteja and Sopory,2008). Overexpression of the CDPK, At CPK6 (Xu *et al.,* 2010), and of the calcineurin B-like protein, CLB5, (Cheong *et al.,* 2010) have both been shown to enhance salt and drought tolerance in Arabidopsis. RD20/Clo3, the ortholog of Clo3 in Arabidopsis, is strongly induced by ABA treatment (Takahashi *et al.,* 2000) and by overexpression of RD26, an ABA and drought inducible NAC transcription factor (Fujita *et al.,* 2004). GPA1, the Arabidopsis ortholog of Ga3, is also implicated in ABA signaling; the *gpa1* mutant has enhanced ABA suppression of germination (Ullah *et al.,* 2003), reduced stomatal closure in response to ABA treatment (Fan *et al.,* 2008) and increased transpiration efficiency related to a reduced density of stomates (Nilson and Assmann, 2010). Thus, the physical interaction of the ABA-induced Clo3 and GA3 present important components to investigate the role of Gα in ABA signaling. The sequence of events in the well-known models of Gα activation of PI-PLC and subsequent IP3-trigged release of calcium suggests that the Clo3 interaction with GA3 could play a role in the

feedback inactivation of Gα. This model is supported by the GTPase Activating Proteins (GAP) activity demonstrated by Clo3 in the present work.

GA3 interaction with Clo3 is enhanced by high Ca^{+2} concentrations, and at high Ca^{+2} concentrations the binding to Clo3 is similar to the GTP- and GDP-bound state of Gα (Fig. 3a). Though the competitive interaction between GA3, Clo3 and PI-PLC1 demonstrated by *in vitro* interactions alone does not predict the chronological order of binding events, existing models of PI-PLC activation by Gα suggest that PI-PLC binds to Gα and may be displaced by Clo3, since binding and displacement is enhanced by Ca^{+2}. GAP proteins promote the inactivation of G proteins by GTP hydrolysis, and the inactivation of GA3 by Clo3 is consistent with a number of stress-related responses in which Gα mutants have been shown to have phenotypes linked to increased stress tolerance, including enhanced ABA sensitivity and stomatal closure (Fan *et al.,* 2008), increased transpiration efficiency (Nilson and Assmann, 2010), the regulation of root proliferation in Arabidopsis (Chen *et al.,* 2006) and adaptation to hypoxia in rice (Steffens and Sauter, 2010). The fact that Clo3 also binds to PI-PLC1 suggests that it has functions other than GTPase activation. If Clo3 is a negative regulator of GA3, the logical hypothesis is that it is also a negative regulator of PI-PLC.

GA3 and PI-PLC1

The activation of PI-PLC by Gα has long been known in animal systems, but the interaction of these two classes of proteins has only recently been reported in plants. The interaction of PI-PLC and Gα proteins from *Pisum* was characterized *in vivo* and by yeast two-hybrid analyses (Misra *et al.,* 2007). Here we report the localization of the PI-PLC1 interaction with GA3 on the PM and ER, which raises the possibility that the two locations may relate to different functions. In addition to the

classical model of G-protein-coupled receptor signaling that has been described for PM receptors, the localization of different Gα gene family members to endomembranes in animal systems has also been associated with the regulation of membrane trafficking (Marrari *et al.,* 2007). The activity and localization of PI-PLCs and other phosphoinositide-metabolizing enzymes has also been shown to regulate membrane trafficking (Thole and Nielsen, 2008), indicating that the localization of the PI-PLC interaction with GA3 to the ER and the PM may contribute to the multiplicity of signaling from G-protein receptors. The interaction of GA3 with PI-PLC1 and its lack of interaction with PI-PLC2 demonstrates the specificity of the interaction. The C2 domain of PI-PLC is found in several classes of proteins involved in signaling and membrane trafficking and has been implicated in binding with phospholipids and other proteins (Nalefski and Falke, 1996). The C2 domain of a PI-PLC from *Pisum* was shown to be sufficient for PI-PLC binding with Gα (Misra *et al.,* 2007). The 131 amino acid C2 domain of the *Pisum sativum* PI-PLC has 74% amino acid sequence identity with the *T. aestivum* PI-PLC1 but only 68% identity with PI-PLC2 over the same region. PI-PLC1 and the *Pisum* PI-PLC share 13 amino acids within the sequence of their C2 domains that are not shared by the two wheat PLCs; these are candidates for the critical amino acids mediating the interaction of PI-PLC1 and GA3. The same 13 amino acids are also conserved in two stress induced PI-PLCs identified in other species, the systemic acquired resistance induced *Oryza sativa* PI-PLC (Song and Goodman, 2002) and the drought-, salinity- and ABA induced *Vigna radiata* PI-PLC3 (Kim *et al.,* 2004) (Fig. 6).

```
###
   TaPI-PLC2          KTRLKVTVYMGDGWRFDFRKTHFDKCSPPDFYARVGIAGVVADTMMKETK 50
   TaPI-PLC1          KKTLKVKVYMGDGWRMDFKQTHFDQYSPPDFYARVGIAGVPADSVMKKTK 50
   PsPI-PLC           KTTLKVTVYMGEGWYYDFDHTHFDQFSPPDFYARVGIAGVPFDTIMKKTK 50
   OsPI-PLC           KKTLKVKVYMGDGWRMDFTQTHFDQYSPPDFYARVGIAGVPADSVMKRTR 50
   VrPI-PLC3          KKTLKVTIYMGEGWFHDFKHTHFDQYSPPDFYARVGIAGVPYDTVMKKTK 50
                      *. ***.:***:**  ** :****: **************  *::**.*:

######
   TaPI-PLC2          VIMDNWIPTWDHEFEFPLSVPELALLRVEVHESDNHQKDDFAGQTCLPVW 100
   TaPI-PLC1          AVEDNWVPVWGEEFSFDLTVPELALLRVEAHEYDMSEKDDFAGQTVLPVS 100
   PsPI-PLC           TVEDSWLPSWNEVFEFPLSVPELALLRIEVHEYDMSEKDDFGGQTCLPVW 100
   OsPI-PLC           AIEDNWVPVWEEDFTFKLTVPEIALLRVEVHEYDMSEKDDFGGQTVLPVS 100
   VrPI-PLC3          SVEDNWSPSWNEEFKFPLSVPELALLRVEVHEYDMSEKDDFGGQTCLPVW 100
                       : *.* * * . * * *:***:****:*.** *  :****.*** ***

####
   TaPI-PLC2          ELRSGIRSVRLYARDGEVLRSVKLLMRFEFS 131
   TaPI-PLC1          ELQPGIRAVALFDRKGNKLPNVKLLMRFEFV 131
   PsPI-PLC           ELRTGIRAVPLHSRKGDKYNNVKLLMRFEFI 131
   OsPI-PLC           ELIPGIRAVALHDRKGIKLNNVKLLMRFEFE 131
   VrPI-PLC3          ELRSGIRAVPLYSRKGEKYHNVKLLMRFEFI 131
                      ** .***:* *. *.*    .*********
```

Fig. 6 Multiple sequence alignment of PI-PLC-C2-2 domains.
The alignment between 131 amino acidsof the C2-2 domain of PI-PLCs from *Triticum aestivum*, *Pisum sativum*, *Vigna radita* and*Oryza sativa*. Thirteen amino acids (**#**) are in common between two PI-PLC's known to interact with Gα's, wheat PI-PLC1 and the *Pisum* PI-PLC, and are not shared between TaPI-PLC1 and TaPI-PLC2, the latterdoes not interact with TaGA3.

Clo3 and PI-PLC interaction

The significance of the interaction between Clo3 and PI-PLC is open to speculation as there have been relatively few reports of protein–protein interactions for PI-PLCs. The overexpression of a PI-PLC in maize enhanced drought tolerance (Wang *et al.,* 2008) and the overexpression of NtC7, a protein that anchors PI-PLC to the plasma membrane in tobacco, was found to enhance salt tolerance (Nakamura and Sano, 2009). PI-PLC was also reported to interact with actin in the cytoskeleton of oat roots (Huang and Crain, 2009). In addition to their role in cleaving phosphoinositides from phospholipids, PI-PLCs have been studied for their role in releasing glycosylphosphatidylinositol (GPI)-anchored proteins from membranes (Coonrod *et al.,* 1999). The role of the interaction with GPI proteins implies a large number of potential interaction partners; nevertheless there is little known about the interaction between a PI-PLC and a calcium-binding protein in plants. Clo3 is predicted by TopPred 5 (Claros and Von-Heijne, 1994) to have one membrane-spanning domain and was localized as a GFP-fusion protein to the ER and tonoplast; thus it may play a role in anchoring PI-PLC to the membrane.

Our results indicate a role for the calcium-binding protein Clo3 in the regulation of the signaling proteins Gα and PI-PLC. The GAP activity of Clo3 indicates that it plays a role in attenuating the activity of GA3. In addition, the phenotypes of Gα mutants in Arabidopsis and rice suggest that such attenuation could contribute to a number of physiological and developmental adaptations to environmental stress. The localization of the interactions to both the PM and the ER suggest diverse roles for the interaction. The effect of the protein–protein interaction on the activity of GA3 and PI-PLC and the role of Clo3 is the subject of further investigation.

REFERENCES

Assmann SM (2005) G proteins go green: a plant G protein signaling FAQ sheet. Science 310:71–73

Berridge MJ (1993) Inositol trisphosphate and calcium signalling. Nature 361:315–325

Bisht NC, Jez JM, Pandey S (2011) An elaborate heterotrimeric G protein family from soybean expands the diversity of plant G-protein networks. New Phytol 190:35–48

Blumward E, Aharon GS, Lam BCH (1998) Early signal transduction pathways in plant-pathogen interactions. Trends Plant Sci 3:342–346

Chen JG (2008) Heterotrimeric G-protein signaling in Arabidopsis: puzzling G-protein-coupled receptor. Plant Signal Behav 3:1042–1045

Chen JG, Gao Y, Jones AM (2006) Differential roles of Arabidopsis heterotrimeric G-protein subunits in modulating cell division in roots. Plant Physiol 141:887–897

Cheong YH, Sung SJ, Kim BG, Pandey GK, Cho JS, Kim KN, Luan S (2010) Constitutive overexpression of the calcium sensor CBL5 confers osmotic or drought stress tolerance in Arabidopsis. Molec Cell 29:159–165

Claros MG, Von-Heijne G (1994) TopPred II, an improved software for membrane protein structure predictions. Comput Appl Biosci 10:685–686

Coonrod SA, Naaby-Hansen S, Shetty J, Shibahara H, Chen M, White JM, Herr JC (1999) Treatment of mouse oocytes with PI-PLC releases 70-kDa (pI 5) and 35- to 45-kDa (pI 5.5) protein clusters from the egg surface and inhibits sperm-oolemma binding and fusion. Dev Biol 207:334–349

Dodd AN, Gardner MG, Hotta CT, Hubbard KE, Dalchau N, Love J, Assie J, Robertson FC, Jakobsen MK, Gonc¸alves J, Sanders D, Webb AAR (2007) The Arabidopsis Circadian Clock incorporates a cADPR-based feedback loop. Science 318:1789–1792

El-Maarouf H, Lameta A, Gareil M, Zuily-Fodil Y, Pham-Thi A (2001) Cloning and expression under drought of cDNAs coding for two PI-PLCs in cowpea leaves. Plant Physiol Biochem 39:167–172

Fan LM, Zhang W, Chen JG, Taylor JP, Jones AM, Assmann SM (2008) Abscisic acid regulation of guard-cell K? and anion channels in G beta- and RGS-deficient Arabidopsis lines. Proc Natl Acad Sci USA 105:8476–8481

Fujita M, Fujita Y, Maruyama K, Seki M, Hirratsu K, Ohme-Takagi M, Tran LSP, Yamaguchi-Shinozaki K, Shinozaki K (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J 39:863–876

Grigston JC, Osuna D, Scheible WR, Liu C, Stitt M, Jones AM (2008) D-Glucose sensing by a plasma membrane regulator of G signaling protein, AtRGS1. FEBS Lett 582:3577–3584

Gulick PJ, Drouin S, Yu Z, Danyluk J (2005) Transcriptome comparison of winter and spring wheat responding to low temperature. Genome 48:913–923

Hamm HE (1998) The many faces of G protein signaling. J Biol Chem 273:669–672

Heinze M, Steighardt J, Gesell A, Schwartze W, Roos W (2007) Regulatory interaction of the Galpha protein with phospholipase A2 in the plasma membrane of Eschscholzia californica. Plant J 52:1041–1051

Hernandez-Pinzon I, Pateland K, Murphy DJ (2001) The Brassica napus calcium-binding protein, caleosin, has distinct endoplasmic reticulum- and lipid body-associated isoforms. Plant Physiol Biochem 39:615–622

Hossain MS, Koba T, Harada K (2003) Cloning and characterization of two full-length cDNAs, TaGA1 and TaGA2, encoding G-protein alpha subunits expressed differentially in wheat genome. Genes Genet Syst 78:127–138

Houde M, Belcaid M, Ouellet F, Danyluk J, Monroy AF, Dryanova A, Gulick P, Bergeron A, Laroche A, Links MG, MacCarthy L, Crosby WL, Sarhan F (2006) Wheat EST resources forfunctional genomics of abiotic stress. BMC Genom 7:149

Hu CD, Kerppola TK (2005) Direct visualization of protein interactions in living cells using bimolecular fluorescence complementation analysis. In: Adams P, Golemis E (eds) Protein–protein interactions, vol 34. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 1–20

Huang CH, Crain RC (2009) Phosphoinositide-specific phospholipase C in oat roots: association with the actin cytoskeleton. Planta 230:925–933

Huang J, Taylor JP, Chen JG, Uhrig JF, Schnell DJ, Nakagawa T, Korth KL, Jones AM (2006) The plastid protein THYLAKOID FORMATION1 and the plasma membrane G-protein GPA1 interact in a novel sugar-signaling mechanism in Arabidopsis. Plant Cell 18:1226–1238

Johnston CA, Taylor JP, Gao Y, Kimple AJ, Grigston JC, Chen JG, Siderovski DP, Jones AM, Willard FS (2007) GTPase acceleration as the rate-limiting step in Arabidopsis G protein-coupled sugar signaling. Proc Natl Acad Sci USA 104:17317–17322

Jones AM, Assmann SM (2004) Plants: the latest model system for G-protein research. EMBO Rep 5:572–578

Jones JC, Temple BR, Jones AM, Dohlman HG (2011) Functional reconstitution of an atypical G protein heterotrimer and regulator of G protein signaling protein (RGS1) from Arabidopsis thaliana. J Biol Chem 286:13143–13150

Kanuru M, Samuel JJ, Balivada LM, Aradhyam GK (2009) Ionbinding properties of Calnuc, Ca^{2+} versus Mg^{2+} Calnuc adopts additional and unusual Ca^{2+}-binding sites upon interaction with G-protein. FEBS J 276:2529–2546

Kim Y, Jee K, Myoung L, Ho J, Young B, Byung H, Inhwan H, Woo K (2004) The Vr-PLC3 gene encodes a putative plasma membrane-localized phosphoinsitide specific phospholipase C whose expression is induced by abiotic stress in mung bean Vigna radiate L. FEBS Lett 556:127–136

Kudla J, Batisti O, Hashimoto K (2010) Calcium signals: the lead currency of plant information processing. Plant Cell 22:541–563

Lapik YR, Kaufman LS (2003) The Arabidopsis cupin domain protein AtPirin 1 interacts with the G protein a-subunit GPA1 and regulates seed germination and early seedling development. Plant Cell 15:1578–1590

Mamillapalli R, Wysolmerski J (2010) The calcium-sensing receptor couples to Galpha(s) and regulates PTHrP and ACTH secretion in pituitary cells. J Endocrinol 204:287–297

Marrari Y, Crouthamel M, Irannejad R, Wedegaertner PB (2007) Assembly and trafficking of heterotrimeric G proteins. Biochemistry 46:7665–7677

Misra S, Yuliang W, Gayatri V, Sopory KS, Tuteja N (2007) Heterotrimeric G-protein complex and G-protein-coupled receptor from a legume (Pisum sativum): role in salinity and heat stress and cross-talk with phospholipase C. Plant J 51:656–669

Monroy AF, Dryanova A, Malette B, Oren DH, Ridha Farajalla M, Liu W, Danyluk J, Ubayasena LWC, Kane K, Scoles GS, Sarhan F, Gulick PJ (2007) Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. Plant Mol Biol 64:409–423

Munnik T, Irvine RF, Musgrave A (1998) Phospholipid signalling in plants. Biochem Biophys Acta 1389:222–272

Nakamura K, Sano H (2009) A plasma-membrane linker for the phosphoinositide-specific phospholipase C in tobacco plants. Plant Signal Behav 4:26–29

Nalefski EA, Falke JJ (1996) The C2 domain calcium-binding motif: structural and functional diversity. Protein Sci 5:2375–2390

Nelson KB, Xue C, Andreas N (2007)A multicolored set of in vivo organelle markers for co-localization studied in Arabidopsis and other plants. Plant J 51:1126–1136

Nilson SE, Assmann SM (2010) The alpha-subunit of the Arabidopsis heterotrimeric G protein, GPA1, is a regulator of transpiration efficiency. Plant Physiol 152:2067–2077

Okamoto H, Gobel C, Capper RG, Saunders N, Feussner I, Knight MR (2009) The alpha-subunit of the heterotrimeric G-protein affects jasmonate responses in Arabidopsis thaliana. J Exp Bot 60:1991–2003

Pandey S, Assmann SM (2004) The Arabidopsis putative G proteincoupled receptor GCR1 interacts with the G protein alpha subunit GPA1 and regulates abscisic acid signaling. Plant Cell 16:1616–1632

Pandey S, Chen JG, Jones AM, Assmann SM (2006) G-protein complex mutants are hypersensitive to abscisic acid regulation of germination and postgermination development. Plant Physiol 141:234–256

Pandey S, Nelson DC, Assmann SM (2009) Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. Cell 136:136–148

Pandey S, Wang RS, Wilson L, Li S, Zhao Z, Gookin TE, Assmann SM, Albert R (2010) Boolean modeling of transcriptome data reveals novel modes of heterotrimeric G-protein action. Mol Syst Biol 6:372–402

Partridge M, Murphy D (2009) Roles of a membrane-bound caleosin and putative peroxygenase in biotic and abiotic stress responses in Arabidopsis. Plant Physiol Biochem 47:796–806

Ritche S, Gilroy S (2000) Abscisic acid stimulation of phospholipase D in the barley aleurone is G-protein-mediated and localized to the plasma membrane. Plant Physiol 124:693–702

Song F, Goodman MR (2002) Molecular cloning and characterization of a rice phosphoinositide-specific phospholipase C gene, OsPIPLC1, that is activated in systemic acquired resistance. Physiol Mol Plant Pathol 61:31–40

Sprang SR (1997) G protein mechanisms: insights form structural analysis. Annu Rev Biochem 66:639–678

Steffens B, Sauter M (2010) G proteins as regulators in ethylenemediated hypoxia signaling. Plant Signal Behav 5:375–378

Takahashi S, Katagiri T, Yamaguchi-Shinozaki K, Shinozaki K (2000) An Arabidopsis gene encoding a Ca2?-binding protein is induced by abscisic acid during dehydration. Plant Cell Physiol 41:898–903

Tardif G, Kane NA, Adam H, Labrie L, Major G, Gulick P, Sarhan F, Laliberte JF (2007) Interaction network of proteins associated with abiotic stress response and development in wheat. Plant Mol Biol 63:703–718

Thole JM, Nielsen E (2008) Phosphoinositides in plants: novel functions in membrane trafficking. Curr Opin Plant Biol 11:620–631

Todorova R (2009) In vitro interaction between the N-terminus of the Ewing's sarcoma protein and the subunit of RNA polymerase II hsRPB7. Mol Biol Rep 36:1269–1274

Tosetti P, Pathak N, Jacob MH, Dunlap K (2003) RGS3 mediates a calcium-dependent termination of G protein signaling in sensory neurons. Proc Natl Acad Sci USA 100:7337–7342

Tuteja N, Sopory SK (2008) Chemical signaling under abiotic stress environment in plants. Plant Signal Behav 3:525–536

Ullah H, Chen JG, Temple B, Boyes DC, Alonso JM, Davis JM, Ecker JR, Jones AM (2003) The b-subunit of the Arabidopsis G protein negatively regulates auxin-induced cell division and affects multiple developmental processes. Plant Cell 15:393–409

Voinnet O, Rivas S, Mestre P, Baulcombe DC (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J 33:949–956

Vossen JH, Abd-El-Haliem A, Fradin EF, Berg GC, Ekengren SK, Meijer HJ, Seifi A, Bai Y, Have A, Munnik T, Thomma BP, Joosten MH (2010) Identification of tomato phosphatidylinositol-specific phospholipase-C (PI-PLC) family members and the role of PLC4 and PLC6 in HR and disease resistance. Plant J 62:224–239

Walter M, Chaban C, Schutze K, Batistic O, Weckermann K, Nake C, Blazevic D, Grefen C, Schumacher K, Oecking C, Hater K, Kudla J

(2004) Visualization of protein interaction in living plant cells using bimolecular fluorescence complementation. Plant J 40:428–438

Wang XQ, Ullah H, Jones AM, Assmann SM (2001) G protein regulation of ion channels and abscisic acid signaling in Arabidopsis guard cells. Science 292:2070–2072

Wang CR, Yang AF, Yue GD, Gao Q, Yin HY, Zhang JR (2008) Enhanced expression of phospholipase C1 (ZmPLC1) improves drought tolerance in transgenic maize. Planta 227:1127–1140

Warpeha KM, Lateef SS, Lapik Y, Anderson M, Lee BS, Kaufman LS (2006) G-protein-coupled receptor 1, G-protein G alphasubunit 1, and prephenate dehydratase 1 are required for blue light-induced production of phenylalanine in etiolated Arabidopsis. Plant Physiol 140:844–855

Weiss CA, White E, Huang H, Ma H (1997) The G protein a subunit (GPa1) is associated with the ER and the plasma membrane in meristematic cells of Arabidopsis and cauliflower. FEBS Lett 407:361–367

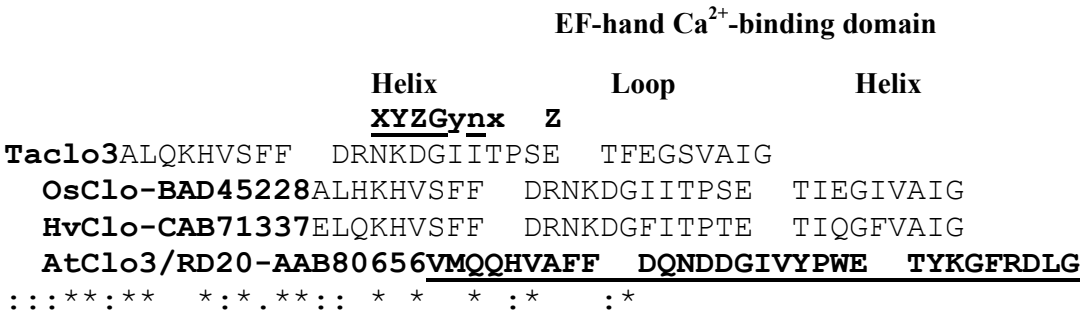
Willard FS, Siderovski DP (2004) Purification and in vitro functional analysis of the Arabidopsis thaliana regulator of G-protein signaling-1. Methods Enzymol 389:320–338

Xu J, Tian YS, Peng RH, Xiong AS, Zhu B, Jin XF, Gao F, Fu XY, Hou XL, Yao QH (2010) AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in Arabidopsis. Planta 231:1251–1260

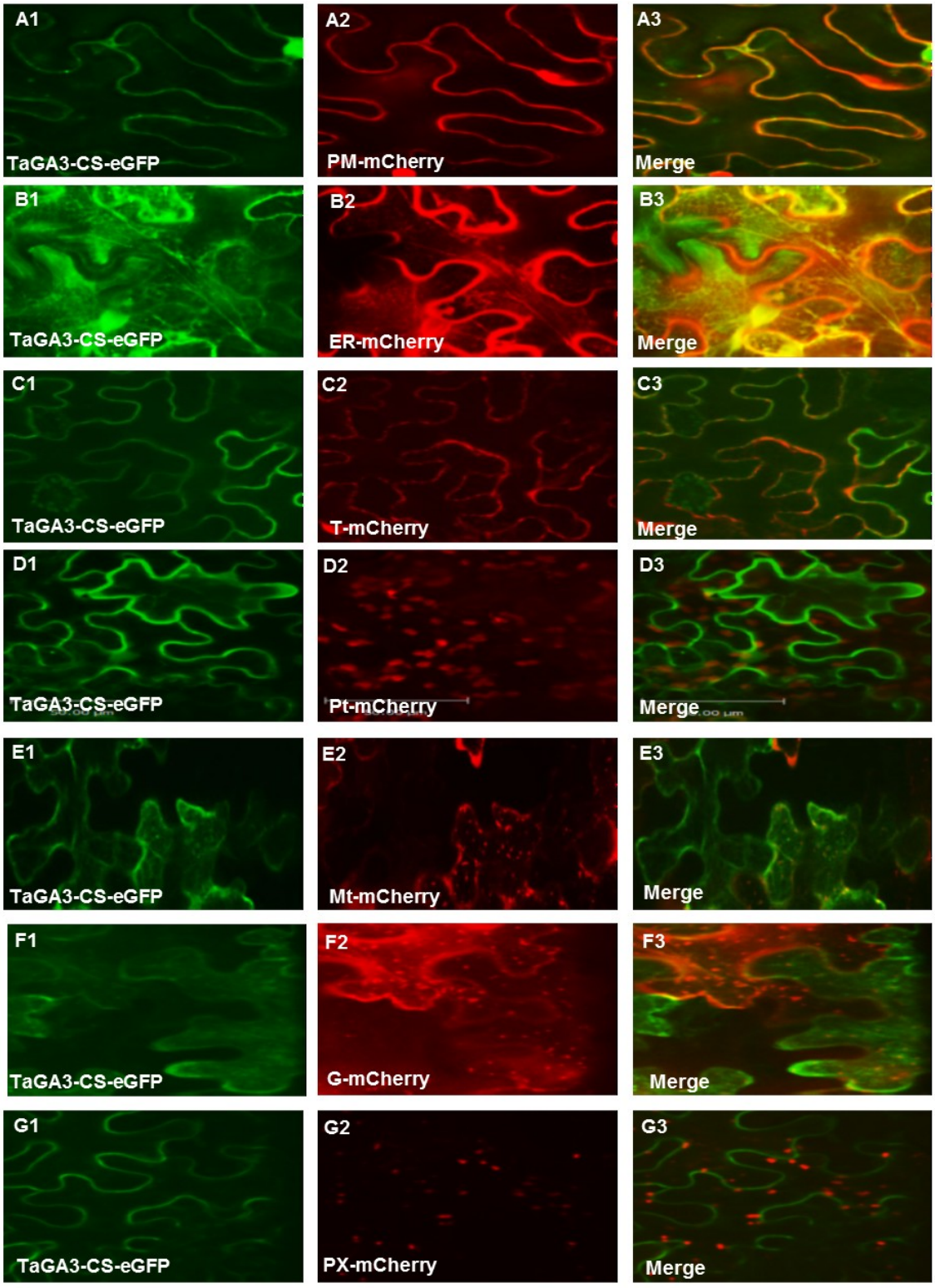
Zhai S, Zhenhua S, Aifang Y, Zhang J (2005) Characterization of a novel phosphoinositide-specific phospholipase C from Zea mays and its expression in Escherichia coli. Biotech Lett 27:799–804

Zhao J, Wang X (2004) Arabidopsis phospholipase Dalpha1 interacts with the heterotrimeric G-protein alpha-subunit through a motif analogous to the DRY motif in G-protein-coupled receptors. J Biol Chem 279:1794–1800 158 Plant Mol Biol (2011) 77:145–158

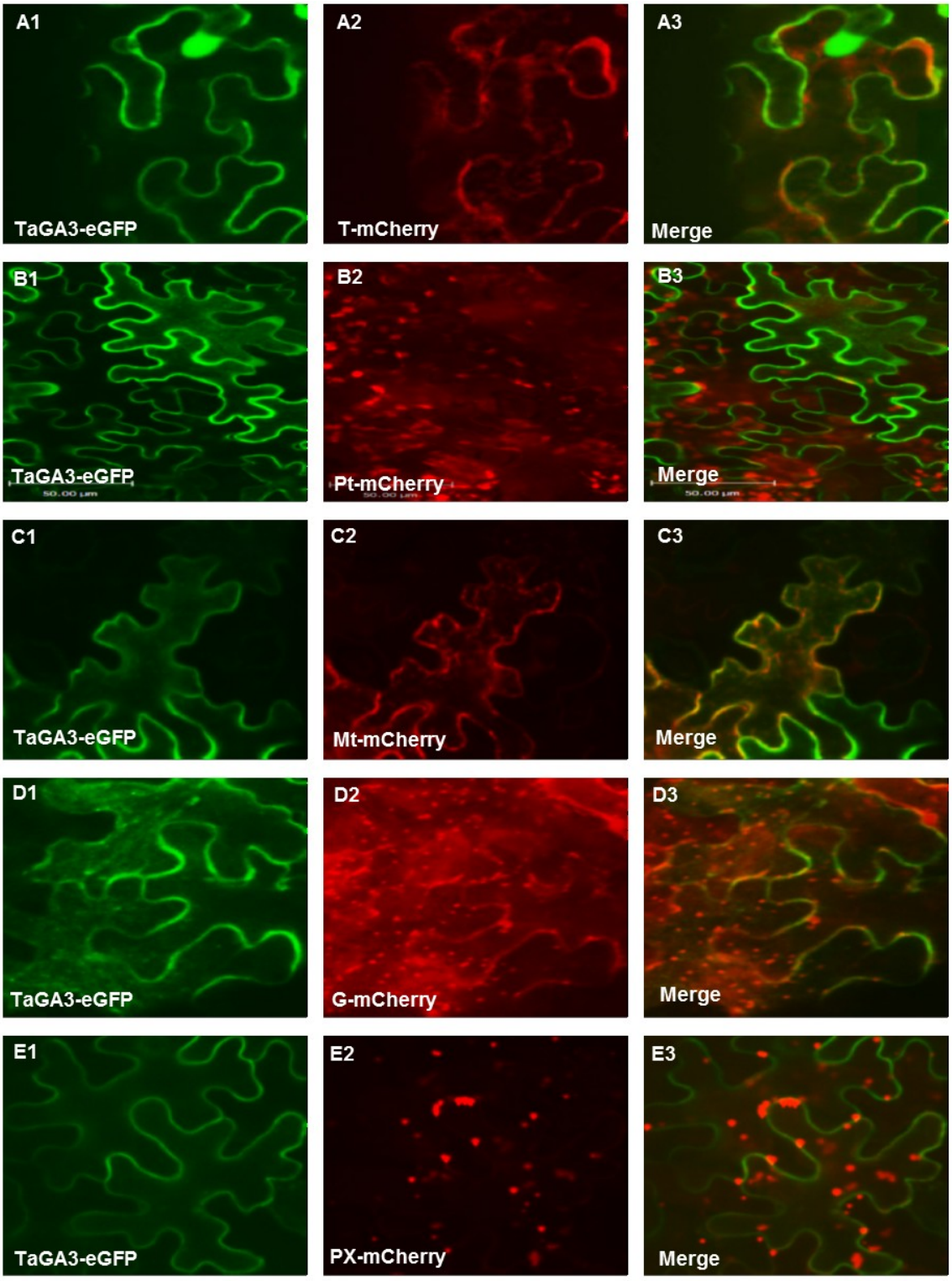
SUPPLEMENTAL FIGURES



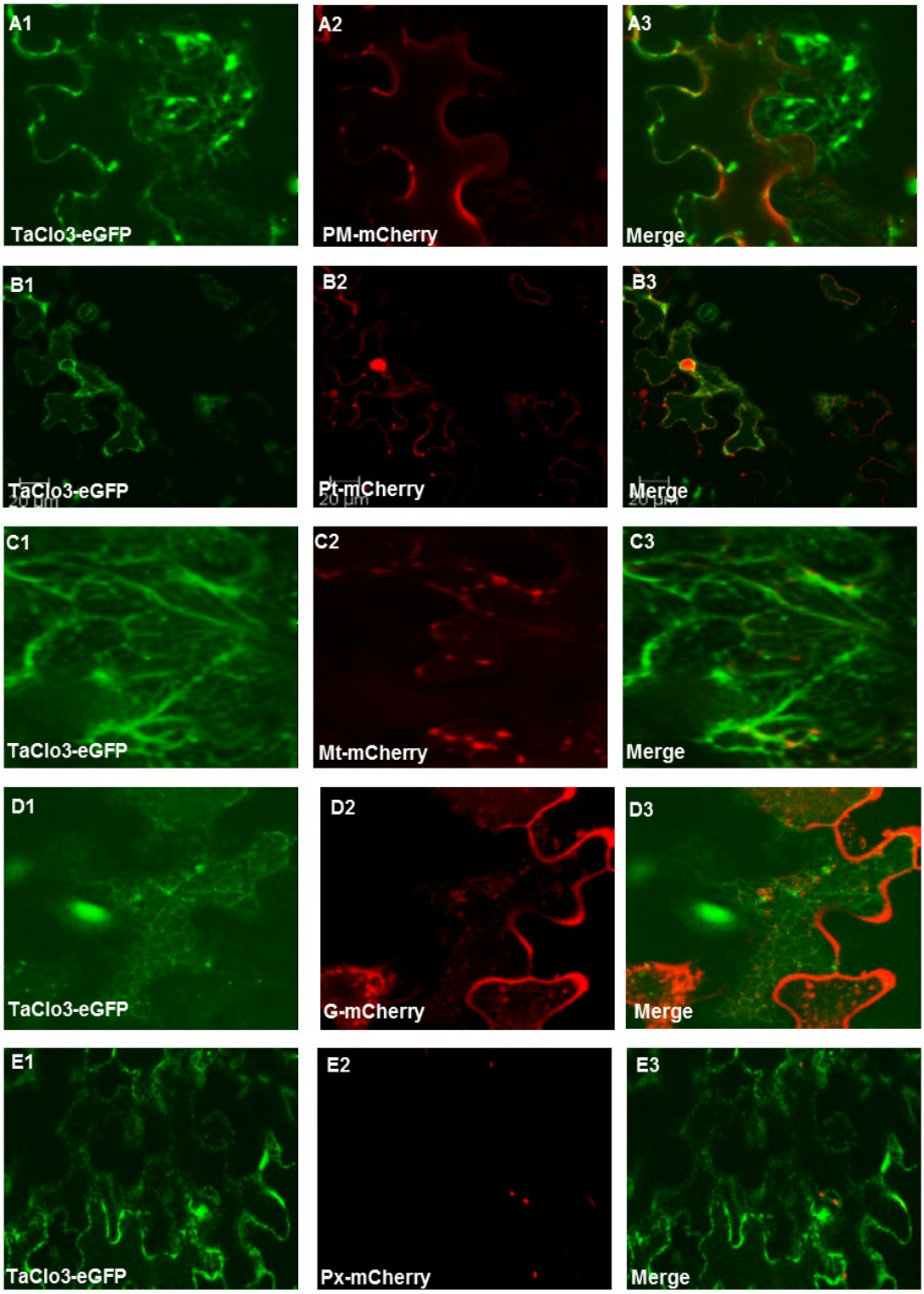
SupplementalFig.S1 TheEF-hand motif of the TaClo3 compared with other calcium binding proteins.
The conservation of thirty amino acids at several positions of *Triticum aestivum* Clo3, *Oryza sativa* Clo (BAD45228), *Hordeum vulgare* Clo (CAB71337) and Arabidopsis Clo3/RD20 (AAB80656) indicates the consensus sequence for the EF-hand motif. The calcium binding EF-hand domain is underlined according to Takahashi *et al.,* (2000). Ca^{2+} is coordinated by ligands within the 12-residue loop. The X, Y, Z,y, x and Z are six putative amino acids that are bound with the calcium ion. n = hydrophobic residue, G = glycine.



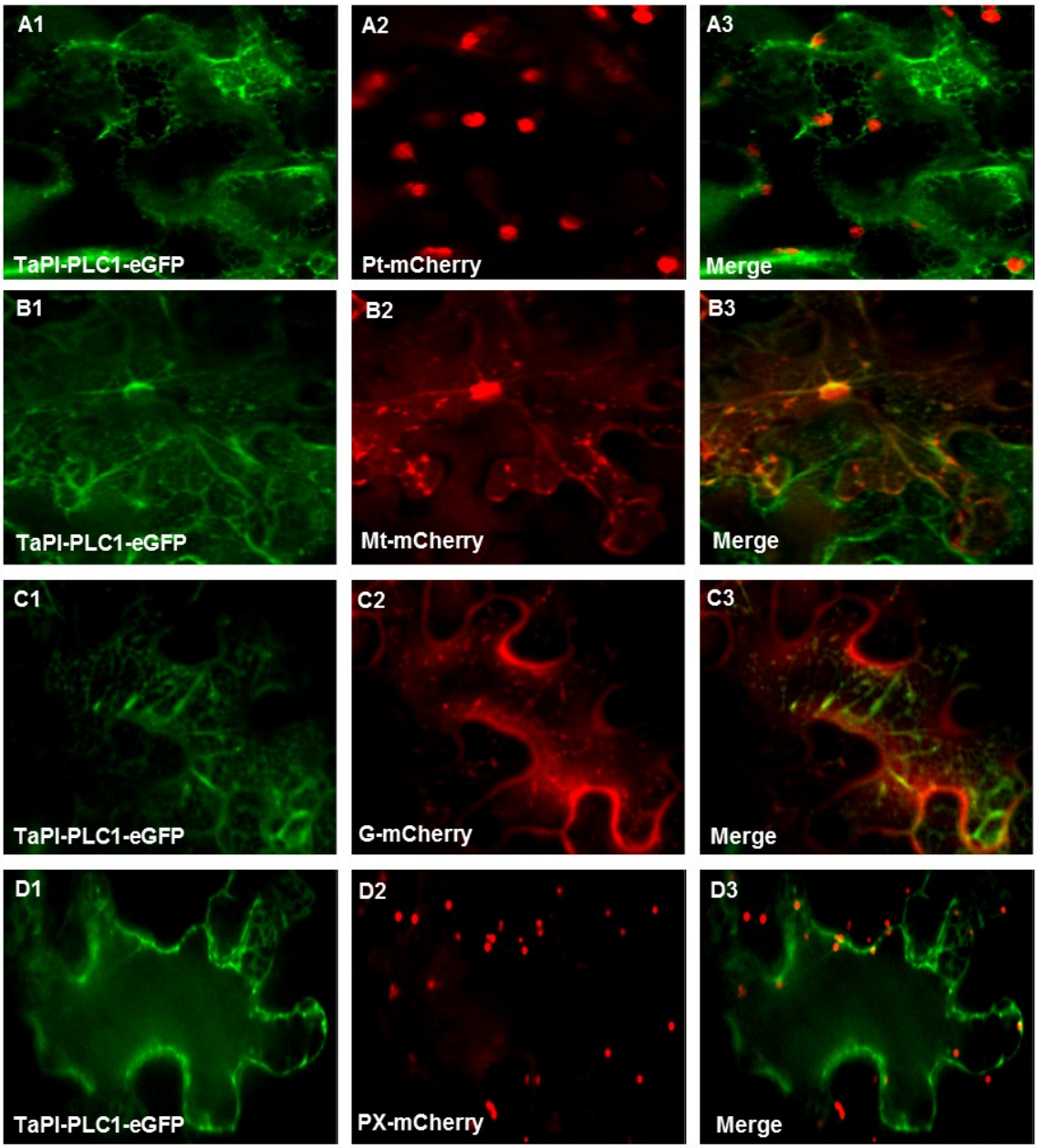
Supplemental Fig. S2 Co-localization of TaGA3-CS-eGFP with mCherry organelle markers in tobacco leaf epidermal cells.
A: PM; B: ER; C: tonoplast; D: plastid; E: mitochondria; F: golgi; G: peroxisome; Scale bar = 50 μm.



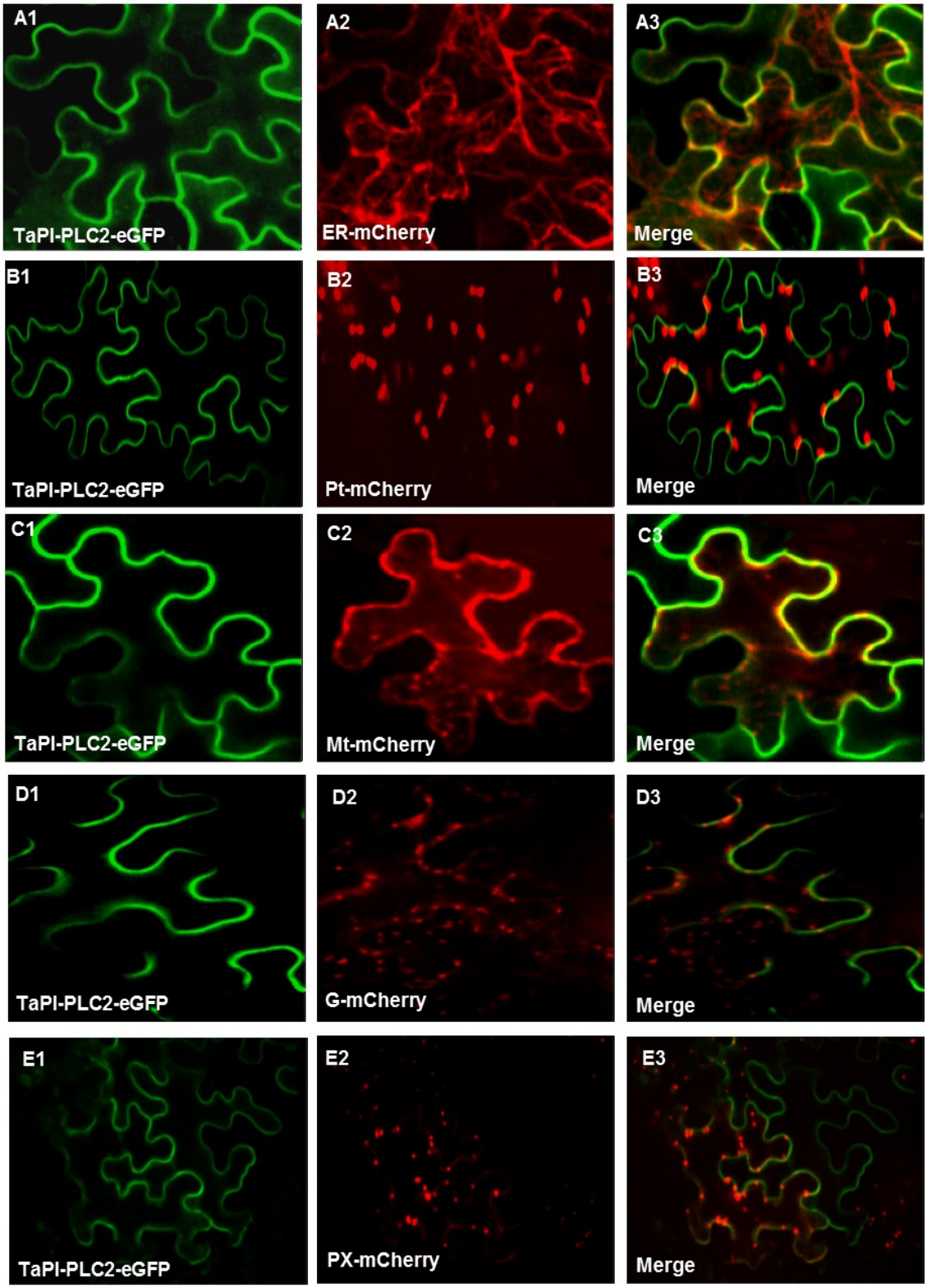
Supplemental Fig. S3 Co-localization of TaGA3-eGFP with mCherry organelle markers in tobacco leaf epidermal cells.
A:tonoplast; B:plastids; C:mitochondria; D:golgi; E:peroxisome; Scale bar = 50 μm.



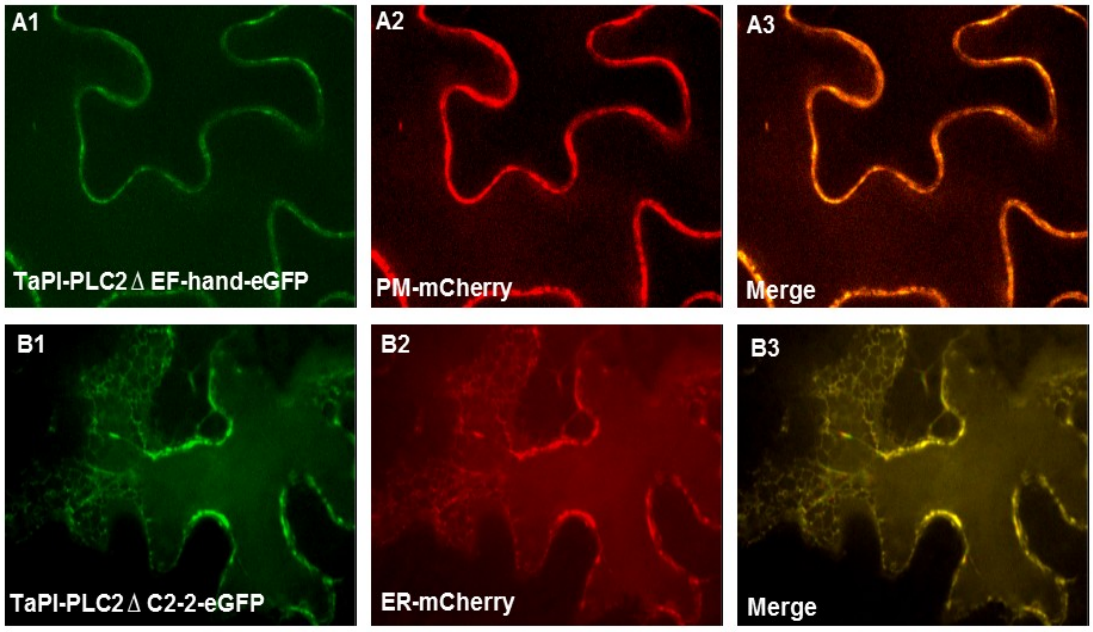
Supplemental Fig. S4 Co-localization of TaClo3-eGFP with mCherry organelle markers in tobacco leaf epidermal cells.
A:plasma membrane; B:plastids; C:mitochondria; D: golgi; E:peroxisome; Scale bar = 50 μM.



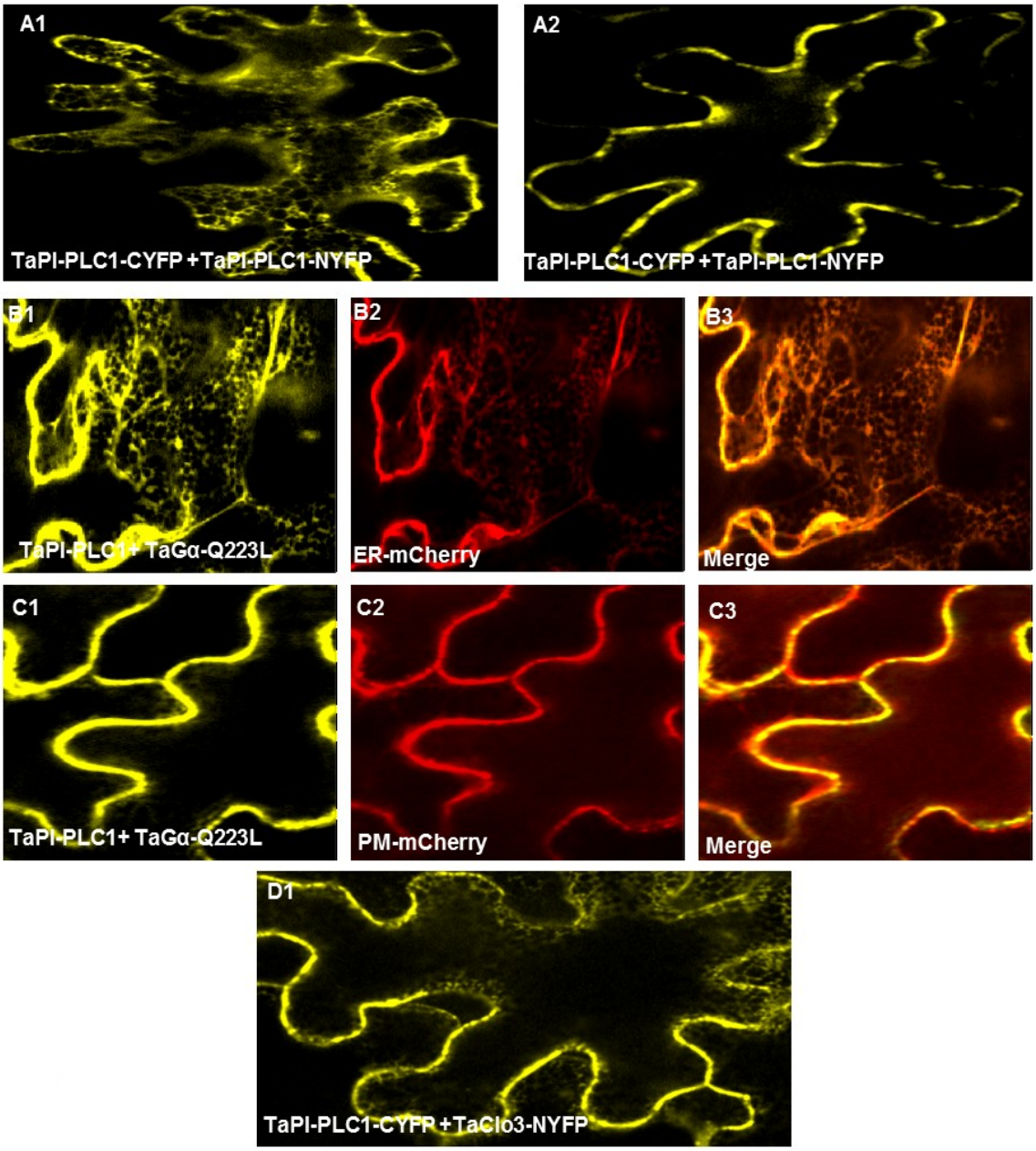
Supplemental Fig. S5 Co-localization of TaPI-PLC1-eGFP with mCherry organelle markers in tobacco leaf epidermal cells.
A:plastids; B:mitochondria; C:golgi; D: peroxisome; Scale bar = 24 μM.



Supplemental Fig. S6 Co-localization of TaPI-PLC2-eGFP with mCherry organelle markers in tobacco leaf epidermal cells.
A:ER; B:plastids; C: mitochondria; D:golgi; E: peroxisome; Scale bar = 24 μM.



Supplemental Fig. S7 Subcellular localization of TaPI-PLC2 truncations in tobacco leaf epidermal cells.
A: co-localization of TaPI-PLC2ΔEF-hand-eGFP with ER; B: co-localization of TaPI-PLC2ΔC2-2-eGFP with ER Scale bar = 24 μM.



Supplemental Fig. S8 BiFC interactions in tobacco leaf epidermal cells. A:TaPI-PLC1-CYFP and TaPI-PLC1 -NYFP dimerization on ER and PM; B:co-localization of TaGA3-Q223L-CYFP and TaPI-PLC1-NYFP interaction to ER; C:co-localization of TaGA3-Q223L-CYFP and TaPI-PLC1-NYFP to PM; Twenty cells of TaGA3-Q223L-CYFP and TaPI-PLC1-NYFP BiFC interaction have higher YFP flourecent compered to TaGA3-CYFP and TaPI-PLC1-NYFP BiFC interaction; D:the localization of the interaction between the two TaPI-PLC1-CYFP and TaClo3-NYFP apparent by ER network structure seen in the upper focal planes of the cell. Scale bar = 24 μM

SUPPLEMENTAL TABLES

Supplemental Table S1: Oligo nucleotide primers used in this study.

Primer Name	Sequence
M13 (For.)	5'GTAAAACGACGGCCAGT3'
TaClo3-internal (For.)	5'GGTGGCTGCTCTTCTGTTTC3'
TaGA3 (For.)	5'TGCATTTGATGTTAAAGGAAGGT3'
PCMVSPORT6 (For.)	5'CCATAGAAGACACCGGGA 3'
Partial-TaPI-PLC1(Rev.)	5'CAGGTGAAGGGAGCTCTT3'
AttB1 pCMVSPORT6 (For.)	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGCCTCCGGACTCTAGC3'
AttB2 Partial TaPI-PLC1-1 (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTCGGGTGTAAGGTGGTCT3'
AttB2 partial TaPI_PLC1-2 (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCAGTCGCTGCTGAGCTGGT3'
AttB-P1 TaGA3 (For.)	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGCTCCTCCTG CAGCA3'
AttB-P2 TaGA3 (For.)	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAAGGAGGATGT GCTCCATG3'
AttB-P3 TaGA3 (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTGGGTACGTCCCCGTTCCCTCCCT3'
TaGA3-CS (For.)	5'ATGGGCTCCTCCTGCAGCA3'
TaGA3-CS (Rev.)	5'CCGGCTTGCTGCTCTGGA3'
TaGA3-internal (For.)	5'GGAGACGAAGGAACTGTTCGACTG3'
TaGA3 (Rev.)	5'CGTCCCCGTTCCCTCCCT3'
Q223LTaGA3 (Rev.)	5'CCTCCTCTCATTCCTTAGACCTCCTAC3'
Q223LTaGA3 (For.)	5'GTAGGAGGTCTAAGGAATGAGAGG3'
AttB1 TaClo3 (For.)	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGATCCGGCG ACAAT3'
AttB1 TaClo3 (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCATTGCACTATGATGAGA AAAGGCCC3'
AttB1 TaPI_PLC1 (For.)	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGCACCTACAAGTGC
AttB2 TaPI_PLC1 (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTG>GGTCCACAAACTCAAAGCGCATG
AttB1 TaPI_PLC2 (For.)	5'GGGGACAAGTTTGTACAAAAAAGCAGGCATGAGCACGTACAGGGT3'
AttB2 TaPI_PLC2 (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCGAAAACTCGAAGCGCAT
AttB1 TaPI-PLC2ΔEF-hand (For.)	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGTTCACCATGACAT3'
AttB2 TaPI-PLC2ΔC2-2 (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTGTCTTCACTGGTAGTTT3'
AttB2 TaPI-PLC1-stop (Rev.)	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACACAAACTCAAAGC GCAT3'
SeqL A (For.)	5' TCGCGTTAACGCTAGCATGGATCTC 3'
SeqL B (Rev.)	5' GTAACATCAGAGATTTTGAGACAC 3'
BamH1 TaGA3 (For.)	5' CGCGGATCCGCGATGGGCTCCTCCTGCAGCAGACCT 3'
Sal1TaGA3 (Rev.)	5'TTCCGCGGCCGCTATGGCCGACGTCGACGCGTTTACGTCCCCGTTC CCTCC CT 3'

Supplemental Table S2: Organelle markers as red fluorescent-protein fusions.

Organelle	Construct Name	Targeting Protein
Plasma Membrane	PM-rk CD3-1007	The full length of AtPIP2A, a plasma membrane aquaporin
Tonoplast	T-rk CD3-975	C-terminus of c-TIP, an aquaporin of the vacuolar membranefused to the fluorescent protein
Endoplasmic Reticulum	ER-rk CD3-959	contains the signal peptide of wall-associated kinase2,fluorescent at the N-terminus of protein and ER retention signal, His- Asp -Glu-leu, at the C- terminus.
Plastids	Pt- rk CD3-999	The first 79 aa of the small subunit of tobacco rubisco.
Mitochondria	Mt-rk CD3-991	The first 29 aa of yeast cytochrome C oxidase IV.
Golgi	G-rk CD3-967	The first49 aa of GmMan1, soybean α-1,2 mannosidae 1.
Peroxisome	Px-rk CD3-983	Peroxisomal targeting signal1, Ser-Lys-Leu , at the C- terminus of the fluorescent protein.

PM, Plasma Membrane; **T,** Tonoplast; **ER**, Endoplasmic Reticulum; **Pt**, Plasitds; **Mt**, Mitochondria; **G**, Golgi; **PX**, Peroxisome; **r**, mCherry fluorescent protein; **k**, kanamycin resistance.

Chapter 3. Characterization of the caleosin gene family in the Triticeae

3.1. Outline and contribution of colleagues

Chapter 3 of the thesis presents the characterization of the caleosin gene family members in the Triticeae. This study was conducted to characterize the size and diversity of thecaleosin gene family in wheat and related species. A total of 34 gene family members that belong to eleven paralogous groups of caleosins were identified in the hexaploid bread wheat, *T. aestivum.* Each group was represented by 3 homeologous copies of the gene located on corresponding homeologous chromosomes, except the caleosin 10, which has four gene copies. Ten gene family members were identified in diploid barley, *Hordeum vulgare*, and in rye, *Secalecereale,* seven in *Brachypodium distachyon*, and six in rice, *Oryza sativa*. The analysis of gene expression was assayed by RNA-Seq analysis of 454 sequence sets and members of the gene family were found to have diverse patterns of gene expression in the nine tissues that were sampled in rye and in triticale, the hybrid hexaploid species derived from wheat and rye. Dr. P. Gulick carried out gene sequence compilation and verification with the first author U. M. Pham. I have identified all caleosin gene family members in rye and studied their expression in rye and triticale tissues using rye and triticale 454-sequence sets. Dr. P. Gulick, U. M. Pham and me, H. B. Khalil, contributed to writing and revision of the manuscript. This paper has been submitted for publication in BMC Genomics (Pham *et al.,* submitted paper)

3.2. Abstract

Background: The caleosin genes encode proteins with a single conserved EF hand calcium-binding domain and comprise small gene families found in a wide range of plant species. Some members of the gene family have been shown to be upregulated by environmental stresses including low water availability and high salinity. Caleosin 3 from wheat has been shown to interact with the α-subunit of the heterotrimeric G proteins, and to act as a GTPase activating protein (GAP). This study was conducted to characterize the size and diversity of the gene family in wheat and related species and to characterize the differential tissue-specific expression of members of the gene family.

Results: A total of 34 gene family members that belong to eleven paralogous groups of caleosins were identified in the hexaploid bread wheat, *T. aestivum.* Each group was represented by 3 homeologous copies of the gene located on corresponding homeologous chromosomes, except for caleosin 10, which has four gene copies. Ten gene family members were identified in diploid barley, *Hordeum vulgare*, and in rye, *Secalecereale,* seven in *Brachypodium distachyon*, and six in rice, *Oryza sativa*. Gene expression was assayed by RNA-Seq analysis of 454 sequence sets and members of the gene family were found to have diverse patterns of gene expression in the nine tissues that were sampled in rye and in triticale, the hybrid hexaploid species derived from wheat and rye.

Conclusions: The caleosin gene family had a greater degree of expansion in the Triticeae than in the other monocot species, *Brachypodium* or rice. Phylogenetic analysis indicates that the gene family has had a marked expansion after the separation of the Triticeae lineage from *Brachypodium.* The prior implication of one

member of the gene family in the stress response and heterotrimeric G protein signaling, points to the potential importance of the caleosin gene family. The complexity of the family and differential expression in various tissues opens the possibilities that caleosin family members may play diverse roles in signaling and development that warrants further investigation.

3.3. Background

Caleosins are calcium-binding proteins encoded by small gene families in plants, and some members of the gene family have been shown to play an important role in signaling and in the response to stress. Ta-*Clo3* encoded by a member of this gene family in wheat (*Triticum aestivum),* was shown to have GAP activity with the heterotrimeric G protein subunit, Gα [1]. The caleosins do not have significant sequence similarity with the Regulators of G protein Signaling (RGS) in other species and appear to be a new class of proteins that act as heterotrimeric G protein GAPs. The wheat Clo3 was also shown to interact with phosphoinositide-specific phospholipase C1 (PI-PLC1) *in vivo* and *in vitro*, and the interaction between Clo3, Gα and PI-PLC1 was found to be competitive [1]. A homolog of Ta-*Clo3* in Arabidopsis, At-*Clo3*, also known as Response to Drought 20 (RD20), has been shown to be strongly induced by drought, abscisic acid and high salinity and was experimentally shown to bind calcium [2]. The *clo3/rd20* mutants in Arabidopsis showed enhanced sensitivity to drought conditions and *RD20* was implicated in the control of stomata aperture, reduction of growth, and increased transpirational efficiency [3]. Plants' responses to water deficit are known to activate signal transduction cascades that increase the level of secondary messengers, including calcium, thus some members of the caleosin gene family appear to play a critical role in water deficit signaling and to link calcium regulation to G protein signaling. Analysis of caleosinsin barley also suggests a role in lipid trafficking and membrane expansion [4]. The caleosin-assembled oil bodies have been proposed as useful components of a nano-carrier for therapeutic purposes, and have been specifically used as drug carriers, targeting cancer cells[5]. It is unknown if the role of caleosins in the stress response is related to their role in lipid bodies, or if they are simply different

functions carried out by different members of the gene families. The caleosins gene family has seven members in Arabidopsis and five members in the rice genome. In order to facilitate the analysis of the members of this gene family and investigate the diverse roles these proteins may play in signaling, we report a description of the whole gene family in hexaploid wheat and diploid rye based on analysis of high-throughput cDNA sequencing data sets and compare these to the other diploid species including barley (*Hordeum vulgare)*, *Brachypodium distachyon*, rice, and Arabidopsis.

3.4. Methods

Caleosin contigs assembly

Triticum aestivum

The calcium binding protein Ta-Clo3/J900 from wheat was used for BLASTp [6] searches in the GenBank NR databases for related caleosin gene sequences in Arabidopsis and rice. The complete set of Arabidopsis and rice caleosin amino acid sequences, as well as that of the Ta-Clo3 were then used to search, with tBLASTn, the GenBank EST database forrelated sequences from*T. aestivum*. The EST sequences were assembled into contigs using CAP3 [7] under high stringency parameters of a minimum sequence identity of 98%; minimum overlap length of 40 nt; gap penalty, 6; match value, 2; mismatch penalty (-5). Open reading frames and translation of the contigs were carried out with the ExPASy translate tool [8] and confirmed by BLASTx with the GenBank NR database comparison to related sequences, and among the derived *T. aestivum* caleosin sequences as the data set was developed. Contigs were manually edited to obtain full length cDNA sequences by first identifying partial length contigs by BLASTx to the NR database and then identify in additional ESTs with 100% identity and a minimum overlap of 120 nt in the *T. aestivum* EST database at NCBI. These were selected for extending the 5' or 3' end of the initial contig. CAP3 was used to assemble all ESTs together with partial sequences to generate full-length contigs. After an initial set of contigs was completed, the process was reiterated to identify additional ESTs and additional homeologous members of the gene family that were not represented in the initial contig set. The *T. aestivum* caleosin contigs were used to search the Wheat Survey Sequences (WSS) of the International Wheat Genome Sequencing Consortium [9] in order to identify the chromosome or chromosome arm on which the gene was located. One cDNA, *Clo4-B*, not represented in the *T. aestivum* GenBank EST database, was

assembled independently from 454 (Roche) cDNA sequences of triticale and from matches of the homeologous Clo4-A to the WSS genomic wheat sequence database.

Secale cereale

Eleven paralogous wheat caleosin genes were used to search for orthologs in five 454-cDNA rye libraries obtained from anther, pistil, crown, roots, and stem, using BLASTn. All 454-cDNAs that matched to wheat caleosin genes with a minimum overlap of 100 nt were selected as caleosin homologs regardless of their BLASTn scores or percent identities. The selected candidates of rye 454-cDNAs were assembled using CAP3 with the same parameters as described above for *T. aestivum*, except the minimum overlap length was set at 35 nt. Using these high stringency assembly parameters, the CAP3 assembly yielded 45 contigs. The preliminary set of contigs was searched with the eleven wheat paralogs using BLASTn to eliminate duplicates and to select contigs with the longest contig length and highest sequence similarity. Contigs assembled with a low depth of coverage were individually compared to the rye 454-cDNAs to confirm the accuracy of their assemblies.

Hordeum vulgare

The 11 paralogous caleosin gene sequences of wheat were used to search in the GeneBank databases for homologous genes in barley (*H. vulgare)*, a close relative of wheat. Gene sequences were identified for eight of these genes in the nr database of full length nucleotide sequences. The sequence of Hv-*Clo9* [GenBank:AK375872.1] had a frame-shift error that became apparent by comparison to the wheat orthologs; the sequences were corrected by comparison to sequences in the GenBank barley EST database. An additional barley caleosin, Hv-*Clo8*, was assembled from EST sequences in the GenBank EST database. Another contig, Hv-*Clo11* was identified in

the second generation sequence database for barley [10] by a BLASTn search of the assembly 1 morex rcba database.In the latter case, the tentative barley transcript was manually assembled from the output of the BLASTn alignment of the barley subject sequence detected by the wheat Ta-*Clo11* caleosin query sequence.

Brachypodium distachyon

T. aestivum caleosin protein sequences were used to search the *B. distachyon* database [11]by tBLASTn. The complete caleosin gene sequence and coding sequences were acquired in FASTA format and were translated with the ExPASy Translate tool [8].In cases where the original annotation appears to have miss-identified the exon/intron junctions or start codons, the flanking sequence for the annotated gene region was reanalyzed and annotated manually by searching for extended ORFs and sequence similarity to the caleosin protein sequences from *T. aestivum.*

Caleosin genes conserved domains and family phylogenetic tree

The most conserved region of the gene family was determined by using NCBI Batch Conserved Domain Search Tool [12] for all contigs, and the result was confirmed with conservation scores calculated by JasView. The calcium binding domain EF-hand motif was identified by alignment of all contigs with EF-hand motifs (1XO5, 1NYA) and calcium binding proteins (1TIZ, 1NKF, 3OX6) obtained from the Protein Data Bank [13]. The results were verified based on the EF-hand motif in Arabidopsis, described by Takahashi and coworkers [2]. The structural and functional features of caleosin genes were examined using InterPro Scan Sequence Search [14],and the Simple Modular Architecture Research Tool (SMART) [15]in genomic mode. These two programs were used in parallel to support the result from the NCBI Search Tool and to verify the presence of the EF-hand motif.

A phylogenetic tree was constructed using wheat caleosin nucleotide sequences aligned using Molecular Evolutionary Genetics Analysis (MEGA), version 5.1 [17]. The maximum likelihood method was employed based on the Jukes-Cantor model [16]. Initial tree(s) for the heuristic search were obtained automatically as follows: When the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise the BIONJ method with MCL distance matrix was used. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.2379)). The analysis involved 35 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 510 positions in the final dataset.

Multiple sequence alignment and phylogenetic tree construction for caleosins from six species were performed using MEGA, version 5.1 [17]. Sequence comparisons were based on the 174 amino acid caleosin domain of protein sequences from eleven paralogous *Clo* genes (one representative of each homeologous gene group) of hexaploid *T. aestivum*, and the gene sequences from the diploid species: *H. vulgare*, *B. distachyon* and *S. cereale* as well as those of *O. sativa* and *A. thaliana*. The amino acid sequences from the six species were aligned using MUSCLE. Phylogenetic trees were constructed using the maximum likelihood method based on the Whelan and Goldman (WAG) model [18]. This method uses standard statistical techniques for inferring probability distributions to assign probabilities to possible phylogenetic trees. The WAG model for amino acids was employed [18], an empirical model of globular protein evolution. Initial tree(s) for the heuristic search were obtained automatically as described above. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter =

1.0411)). The analysis involved 54 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 93 amino acid positions included in the final dataset.

Caleosin expression analysis

The level of expression of 34 caleosin gene family members was estimated in *T. aestivum* using the EST database at GenBank. The relative level of expression and tissue specific expression for each of the caleosin gene family members in rye and in triticale, a synthetic hybrid between *Triticum durum* and rye, was determined by analysis of Roche 454-cDNA sequence libraries. A description of the analysis is presented as a flowchart (Figure S1). Extensive quality control analysis of triticale and rye 454-cDNAs was performed in two steps. Poor quality ends which were removed by deleting continuous nucleotides with a phred score of less than 15 and internal nucleotides with a phred score of less than 20 were masked with N's using the tools available through the Galaxy server from Pennsylvnia State University and Emory University [19]. All *T. aestivum* GenBank ESTs as well as the high quality transcripts obtained from triticale and rye tissues were aligned to their associated wheat, rye, and triticale caleosin reference sequences using the BWA-SW algorithm of BWA aligner [20] using the default parameters except, the mismatching penalty was set at 10 and the z-best heuristics were set to 100 to reduce the false read aligning due to the high similarity of three members of each homeologous group from the hexaploid triticale. The reads that were uniquely mapped to each homeolog were selected and counted. The expression of rye and triticale cDNA reads in 454 libraries was normalized to readsper kilobase of gene length per million reads to correct for the biases from differences in the gene lengths and to normalize the expression among libraries of different sizes.

3.4. Results

Caleosin genes in *T. aestivum*

CAP3 assembly parameters for the minimum perecent identity to allow sequences to be included in the same contig ranging from 80% (default) to 99% per cent identity were evaluated for the assembly of gene family members from hexaploid wheat. The optimal value was determined to be 98% identity. Assembly at a lower minimum percent identity resulted in contigs that included sequences from different homeologous gene copies. Assembly at 99% minimum identity resulted in more numerous and shorter contigs with more independent contigs for the same gene.

Among the species surveyed, hexaploid *T. aestivum* had the largest caleosin gene family with 11 gene family members per haploid genome. In total, 34 full length caleosin-like cDNA sequences were identified in this species (Figure S2; Table S1). All but three (*Clo4-B*, *Clo10-2-D*, *Clo11-A*) of these transcript sequences were compiled from the wheat EST database at NCBI.These three additional sequences were obtained from the WSS derived from second generation sequences from genomic DNA (*Clo4-B*, *Clo10-2-D*, *Clo11-A*). *Clo10-2-D* had a single supporting EST sequence in the GenBank EST database, and *Clo11-A* was supported by a single read from the triticale 454-cDNA data set. *Clo4-B* was identified in the WSS genomic data base and the full coding region was verified by several reads from the triticale 454-cDNA data set, though *Clo4-B* from triticale has a 3 base pair deletion relative to *Clo4-B* from *T. aestivum*. Six cDNA sequences were confirmed by sequencing individual cDNA clones. Pair-wise BLASTn and ClustalW2 [21] comparisons among the sequences identified 11 paralogous sets of three genes corresponding to

homeologous groupings of genes derived from the three ancestral genomes of wheat (Figure S2).

The accuracy of each contig sequence was judged to be excellent, as the contigs generally had a minimum depth of at least three ESTs and many had a depth of four to ten sequences. There was also a very good agreement between the contigs assembled from EST sequences, sequences of individual cDNA clones and the WSS assemblies, which indicates the high accuracy of the WSS database. Genes within homeologous groups had high nucleotide sequence similarity, ranging from 96% to 97% nucleotide sequence identity within the coding region. This high degree of similarity is common among homeologous copies of genes in wheat [22]. The WSS sequences are derived from shotgun 454 sequencing of chromosome arm specific BAC libraries, thus sequences have a chromosomal assignment, but not a map location along the chromosomes. In nearly all cases, homeologous copies of genes were located on the same homeologous chromosome arm. One exception is *Clo9*. The A and D genome copies of *Clo9* are located on the short arm and long arm of chromosomes 4A and 4D, which are considered homologous, the B genome copy is located on the short arm of 4B which is not homologous to the other two chromosome arms [23]. Ten of the 11 paralogous groups had three homeologous copies, but caleosin 10 had four gene copies, one on each of the long arms of chromosome 2A and 2B and two copies identified on chromosome 2DL. These two copies of *Clo10* on 2DL had 93% nt sequence identity within the coding region, which is somewhat lower than the sequence identity among the homeologous copies on chromosomes 2A and 2B and one of the D copies, *Clo10-2D*.

The degree of similarity among the paralogous sequences in wheat spanned a wide range; the most similar paralogs were *Clo10* and *Clo11* which shared 89% nucleotide sequence identity and 95% amino acid sequence similarity within the conserved 174 EF-hand domain. The most dissimilar members of the wheat caleosins gene family were *Clo10* and *Clo4* which had 37% amino acid sequence identity and 57% similarity within the conserved 174 amino acid EF-hand domain. The size of the proteins encoded by members of the gene family ranged from 198 to 302 amino acids. All members of the gene family are characterized by the presence of a single EF-hand calcium binding region of approximately 174 amino acids and a predicted transmembrane domain of 20 amino acids.

Caleosin genes in *Secale cereale* and *Hordeum vulgare*

Ten full-length caleosin cDNA sequences in *S. cereale* (rye) were assembled from the Roche 454-cDNA sequence set. Only Caleosin 2 was not identified in our rye sequence set, though a 239 nt fragment of a gene with 97% sequence similarity to wheat Caleosin 2, is present in the whole genome shotgun sequence for rye (https://www.gabipd.org/ database/cgi-bin/Blast.pl.cgi) [24]. This indicats that all 11 *Clo* genes are present in rye.The rye sequences showed between 90% and 99% sequence identity with their homologs in wheat within the coding region of each sequence. This high degree of similarity is common between wheat, barley, and rye, which are all members of the Triticeae tribe, as seen in the phylogenetic diagram in Figure 1. Ten full length cDNA sequences of caleosin genes were identified in *H. vulgare*. Seven of these were obtained from the non-redundant database at GenBank, one sequence was assembled from ESTs, and one sequence, Hv-*Clo11*, was tentatively derived from second generation sequencing of the barley genomic sequence.

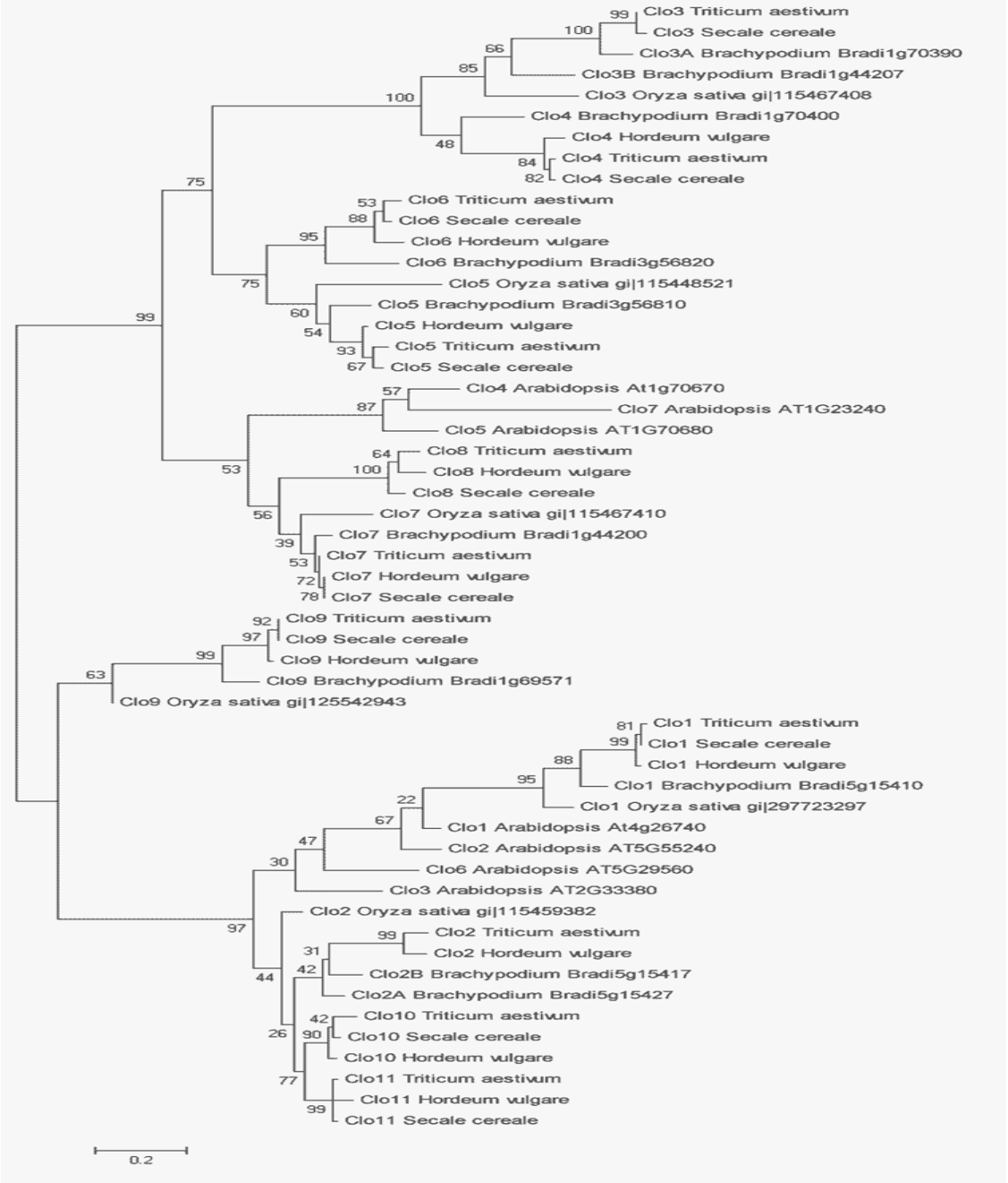


Figure 1.Molecular phylogenetic analysis of the caleosin gene families of *T. aestivum*, *H. vulgare*, *B. distachyon*, *S. cereal*, *O. sativa*, and *A. thaliana* amino acid sequences. The evolutionary history was inferred by using the maximum likelihood method based on the Whelan and Goldman model [18]. The tree with the highest log likelihood (-3638.1820) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Brachypodium gene identifiers are taken from http://brachypodium.org/; rice IDs are from genBank and Arabidopsis IDs are from TAIR, http://www.arabidopsis.org/.

The latter sequence could not be confirmed independently, but sequence differences between the other nine FL sequences from barley and the assembly1-morex rcba, suggest Hv-*Clo11* derived from the viroblast high through-put database may contain some inaccuracies. A *Clo3* ortholog was not identified in the barley sequence databases.

Caleosin genes in *Brachypodium distachyon*

Ten caleosin genes were identified in *B. distachyon,* a species for which the complete genome sequence is available. Two of the sequences were modified by extending the ORF relative to the annotated sequence that is available for the *Brachypodium* genome. All contigs were verified as full-length by comparison to the wheat sequences, and open reading frames and protein sequences were obtained.

Conserved structural elements

The conserved domain in caleosins was identified by Batch CD Search Tool on NCBI and InterPro Scan for each paralogous group from *T. aestivum*, *H. vulgare*, and *B. distachyon*. Results were verified by comparing with the conservation score calculated by aligning multiple sequences with ClustalW2 [21]. The conserved EF-hand calcium binding domain of 174 amino acids is common to all caleosins. This EF-hand motif is composed of 36 amino acids with a calcium chelation loop and calcium ligating residues. The DGSLFE box which is the putative Casein Kinase Phosphorylation site is also highly conserved [2]. The alignment of all peptide sequences from *T. aestivum*, *H. vulgare*, and *B. distachyon* using ClustalW revealed the presence of other less conserved motifs, including the GS loop, transcription termination factor, and a trans-membrane domain.

Relative level of expression of caleosin gene family members in *Triticum aestivum*

The relative level of expression of 34 caleosin gene family members was characterized in *T. aestivum* by calculating the relative abundance of ESTs for each gene in the GenBank EST database using RNA-seq analysis. A total of 483 ESTs were aligned to the three homeologous copies of the eleven *T. aestivum* caleosin paralogs. The abundance of ESTs for the different paralogous groups in the *T. aestivum* database varied widely from a high of 136 representative ESTs for the *Clo1* group, to a low of 7 sequences for the *Clo3* and *Clo10* groups (Table 2); only *Clo11-A* had no corresponding EST sequence in the GenBank EST database. The ratio of expression among homeologous gene copies for each paralog showed an average of approximately three-fold between the most highly expressed copy and the least expressed gene. A few paralogous sets showed a greater degree of differential expression, most notably due to very low expression levels for single members of the group; for example, *Clo4-B* had only a single EST whereas *Clo4-D* had 19 ESTs; *Clo11-A* had no ESTs whereas *Clo11-B* had 7 hits in the EST database (Table 2).

Tissue-specificity of rye and triticale caleosin paralogs

The abundance of ESTs for ten rye caleosin genes as well as twenty-two genes from eleven paralogous groups assigned to the A and B subgenomes was investigated to assess their relative expression level in 454-cDNA libraries obtained from different rye and triticale tissues (Table S2). The data were normalized to take into account the sizes of the different tissue specific data sets and gene lengths, and are expressed as reads

Table 1. Caleosin genes from five species.[a]

	T. aestivum				Homologues			
Gene	Chromosome	Nucleotide	aa	Accession No.[b]	*B. distachyon*	*H. vulgare*	*O. sativa (Japonica)*	*S. cereale*[b]
	2AL	1600	302		Bradi6g15410	gi|345389412	gi|297123291	
Clb1	2DL	1178	302					
	2BL	1154	302					
	2DL	877	245		Bradi6g15427	gi|345389416		
Clb2	2AL	963	244		Bradi6g15417			
	2BL	986	245					
	6DS	793	220		Bradi1g70890			
Clb3	6AS	907	217					
	6BS	818	215					
	3DL	1084	217		Bradi1g70400		gi|115467408	
Clb4	3AL	1089	217		Bradi1g44207			
	3BL	1075	217					
	6AL	1063	229		Bradi3g56810	gi|6900307		
Clb5	6BL	1051	224					
	6DL	1150	222					
	6AL	827	216		Bradi3g56820	gi|151420803	gi|115467410	
Clb6	6BL	735	205					
	6DL	815	214					
	7AS	827	198		Bradi1g44200	gi|151419867	gi|115448521	
Clb7	7BS	917	213					
	7DS	968	212					
	7DS	939	215			gi|326515641		
Clb8	7BS	962	213					
	7AS	1102	214					
	4BS	1129	235		Bradi1g69571	gi|151426143		
Clb9	4DL	1164	233					
	4AS	994	232					
	2DL	897	243			gi|326490092	gi|115459382	
Clb10	2BL	866	244					
	2AL	907	234					
	2DL	836	244					
	2BL	1062	244			VirtBlast		
Clb11	2AL	907	234					
	2DL	932	242					

per kb of gene length per million 454-cDNA reads (RPKM), and presented in Figures 2 and 3, Additional file 3 (Figure S3), Additional file 4 (Figure S4), and Table 2A and 2B.

Caleosin gene family members were expressed in all tissues sampled, and most tissues showed the expression of several paralogous members of the gene family. In rye, expression of six or more caleosin gene members was detected in four of five tissues that were sampled (anther, crown, root and stem) (Figure 2); however, expression of only three caleosin members (*Clo1, Clo5, and Clo11*) was detected in rye pistil tissue, with the expression of *Clo5* being dominant and *Clo1* and *Clo11* being detected at very low levels (Figure 2). Most caleosin transcripts were detected in multiple tissues; however, *Clo6* and *Clo10* were only detected in anther tissue, both at relatively high levels, and *Clo1* expression was detected only in pistil and root tissues and at very low levels (Figure 2 and Figure S3), demonstrating high tissue specificity for these caleosin members. In contrast, *Clo5* was detected in all five tissues (Figure S3).

Triticale showed a similar diversity of expression of all gene family members. Eight tissue types was sampled, and expression of eight or more caleosin gene members were detected in most of these tissues; however, expression of only four caleosin genes was detected in triticale ovary (*Clo4, Clo5, Clo7*, and *Clo9*), and pollen tissues (*Clo2*, *Clo4, Clo5*, and *Clo6*), and expression of only three caleosin genes (*Clo4, Clo7*, and *Clo8*) was detected in stigma tissue, with *Clo7* expression being dominant (Figure 3).Expression of caleosins in the seed was dominated by relatively high levels of *Clo1* and *Clo2*, and expression in pollen tissue was dominated by very high levels of *Clo6* (Figure 3). Most caleosin transcripts were detected in four

or more different triticale tissues. *Clo4* and *Clo5* displayed a very broad tissue expression pattern, as they were detected in all of the tissues investigated (Figure S4).

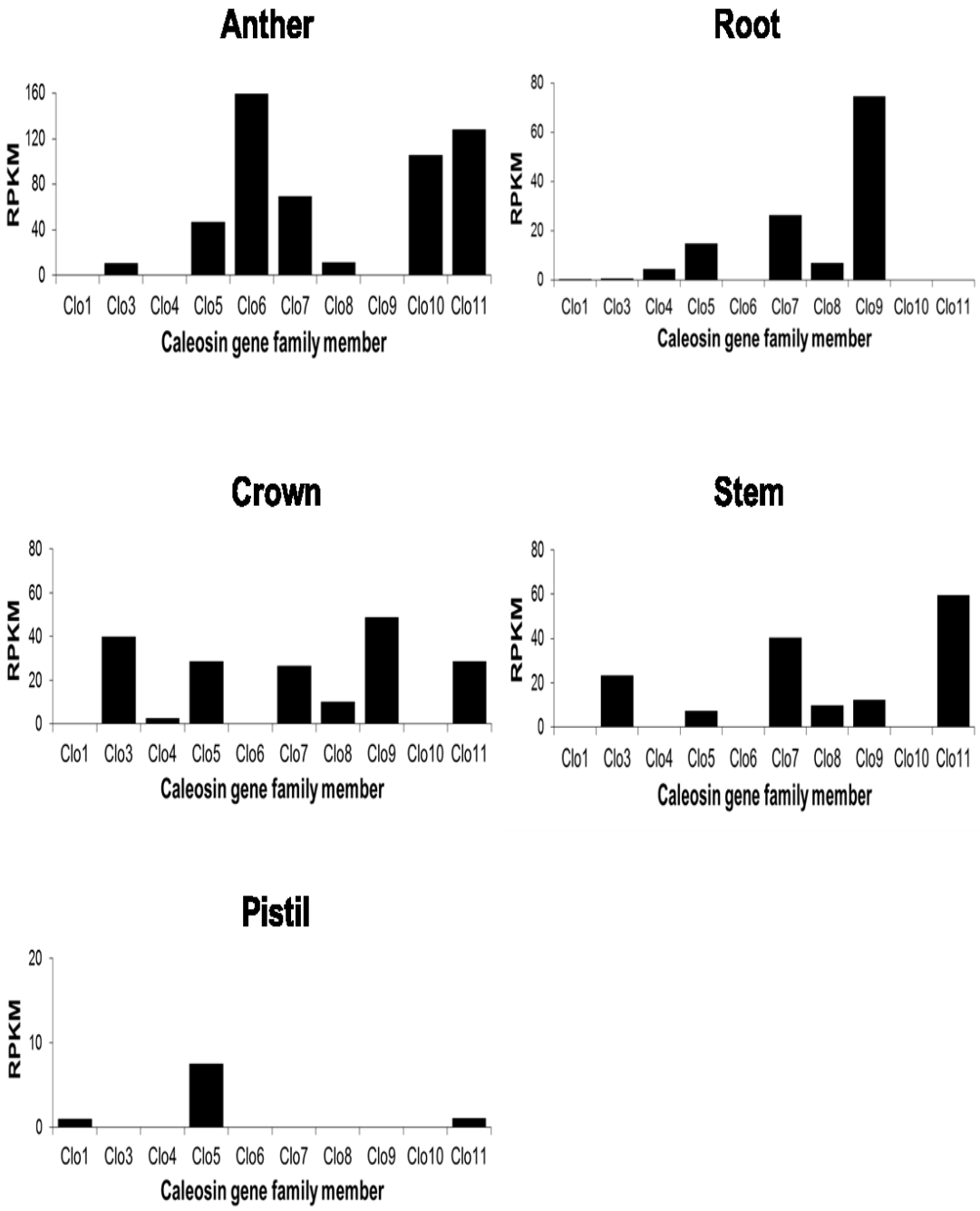


Figure 2. The relative level of expression of ten caleosin gene family members measured in a panel of rye tissues. The expression of caleosin gene family members was estimated in 454 cDNA sequence libraries from anther, crown, pistil, root and stem rye tissues using RNA-seq analysis. The aligned 454-cDNAs to each caleosin member were counted, then normalized based on gene lengths and library depths using the RPKM method.

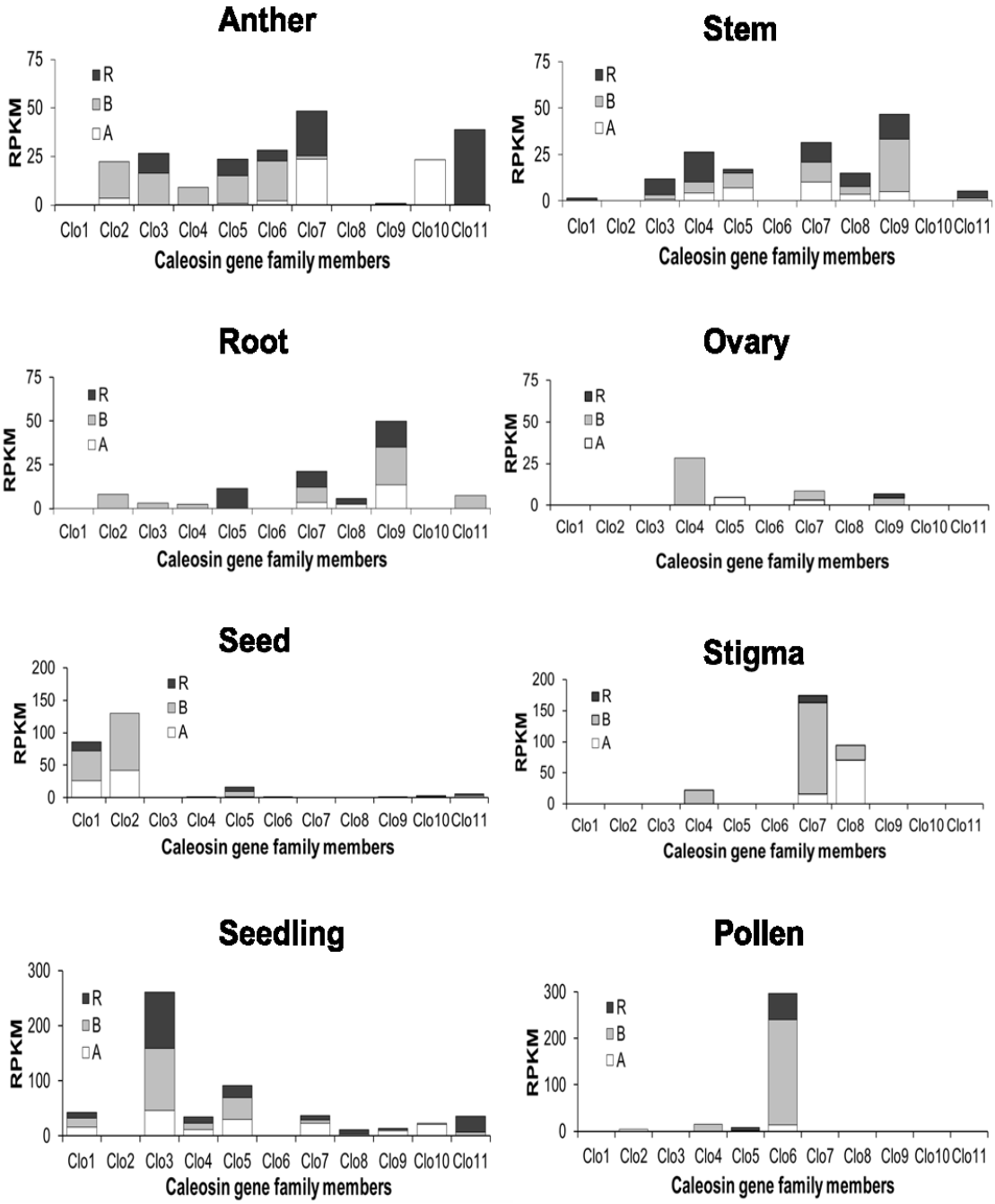


Figure 3. The relative level of expression of eleven caleosin paralogs measured in a panel of triticale tissues. The expression of thirty-two caleosin gene family members was individually measured in 454 cDNA sequence libraries from anther, ovary, pollen, root, seed, seedling, stem and stigma triticale tissues using RNA-seq analysis. The aligned 454-cDNAs to each caleosin member were counted, then normalized based on gene lengths and library depths using RPKM method. The expression of each paralog is subdivided into expression of each of the three homeologs, visualized as black bars for the R subgenome, grey for the B subgenome, and white for the A subgenome. *Clo2* represented the expression of only the A and B homeologs.

Conversely, *Clo3* and *Clo6* showed very high expression levels specifically in seedling and pollen tissues, respectively, and *Clo1* was observed to be primarily expressed in seed and seedlings (Figure S4). Therefore, there is tissue specificity for certain caleosin members, as was observed with expression in rye tissues. Overall, these results demonstrate a very diverse pattern of tissue-specific expression of the caleosin genes. (Figure 4A).*Clo4* was detected at a low level in rye roots, but *Clo4-R* was not detected in triticale (Figure 4A).

The effect of polyploidization on the expression of R subgenome caleosins

Three tissue types were sampled for both rye, and triticale which included root, stem and anther. This facilitated the comparison of the R genome homeologs in both the rye and triticale genetic background, and examination of the effect of polyploidization on R gene copies. In roots, the pattern of gene expression of rye genes that were highly expressed in rye was maintained in triticale. *Clo9*, which was a highly expressed gene in both species, showed expression of all three homeologs at near equal levels in triticale (Figure 4A). Whereas *Clo7*, also highly expressed in the roots of both species, had relatively high levels of expression of the R gene copy in triticale, but much lower expression levels of the A gene copy compared to the other two homeologs (Figure 4A). For genes with low levels of expression only one member of the homeologous group was detected in most cases; due to the common disparity of expression of homeologs as seen in *T. aestivum*, this may be simply a reflection of small sample size. Transcripts of the wheat homeologs of *Clo5* were not detected in the roots of triticale, though the rye homeolog was detected at a moderate level, similar to that observed in rye roots (Figure 4A)

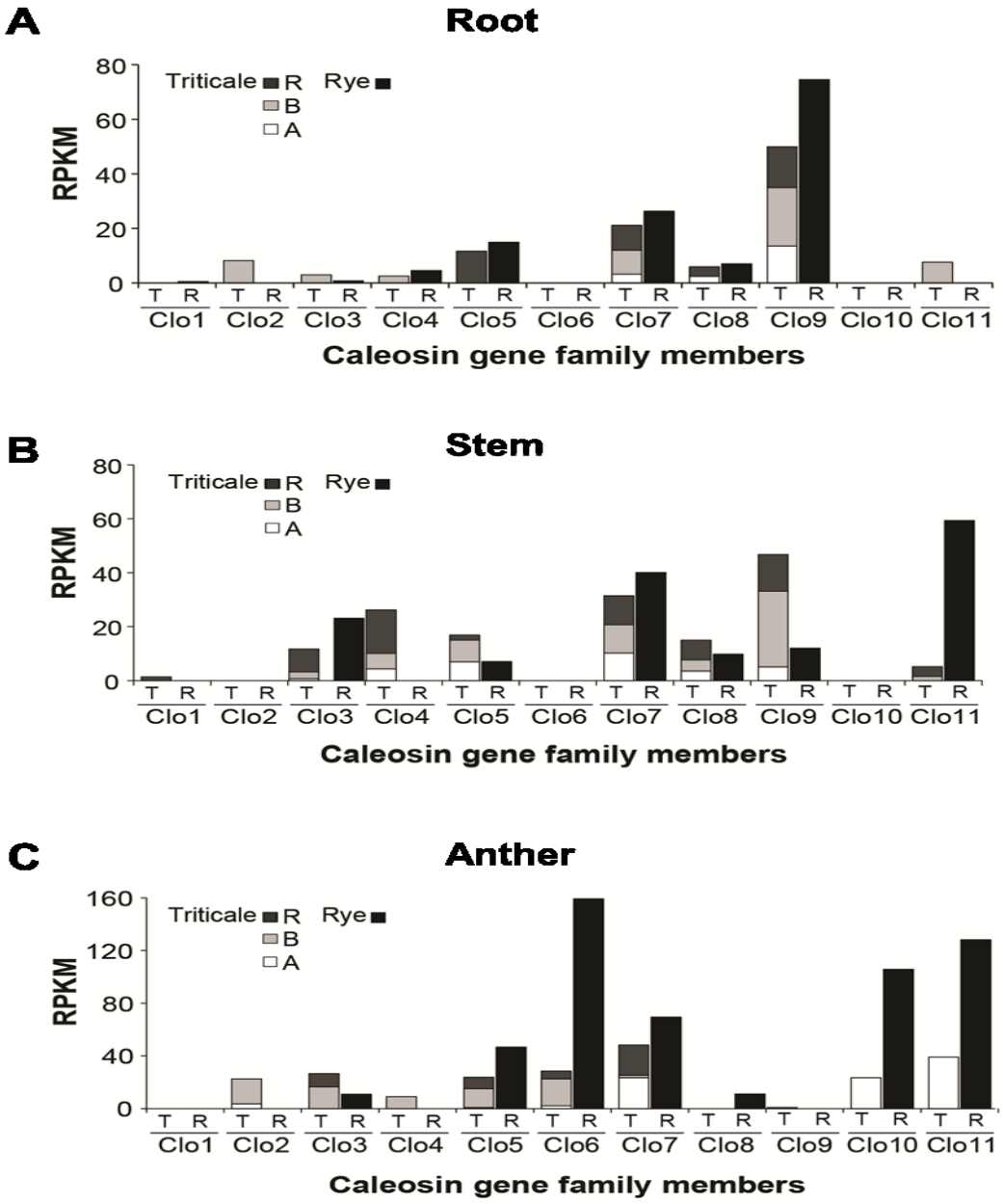


Figure 4.A comparison of Clo gene expression in rye and triticale, to assay the changes in gene expression as a result of polyploidization. The expression of caleosin gene family members measured using RNA-seq analysis in the root **(A)**, stem **(B)**, and anther **(C)** of rye and triticale, respectively. The aligned 454-cDNAs to each caleosin member were counted, then normalized based on gene lengths and library depths using RPKM method. The expression of each paralog is subdivided into expression of each of the three homeologs for triticale (T), visualized as black bars for the R subgenome, grey for the B subgenome, and white for the A subgenome. The expression of caleosin family members in rye (R) is represented by solid black bars.

Stem tissue, in general, had similar patterns of expression in rye and triticale and all three homeologs were expressed (Figure 4B). The most striking exception was *Clo11-R*, which was the most highly expressed caleosin in rye stems, but was expressed at low levels in triticale. *Clo11-B* was expressed at very low levels and *Clo11-A* was not detected (Figure 4B). *Clo4-R* was not detected in rye stem tissue but was expressed in moderate levels in triticale stems, thus *Clo11-R* appears to have been suppressed and *Clo4-R* appears to have been activated by the polyploidization event.

In anther tissue, the relative level of expression of all caleosin genes was lower in triticale than in rye. Though the pattern of expression among gene family members was similar in the two species (Figure 4C) several of the genes showed marked differences. *Clo11-R* was highly expressed in rye, but expressed at lower levels in triticale in the stem and anther, and the A and B wheat homeolog were not detected at all (Figure 4C). *Clo10-R* was highly expressed in rye, and completely suppressed in triticale, indicating a loss of expression in the R genome of triticale in anthers (Figure 4C). In triticale, *Clo7-A* and *Clo7-R* had relatively high expression, but *Clo7-B* was represented by a single read (Figure 4C). *Clo8-R* was detected in rye anthers but not in triticale anthers; however, the level of expression was relatively low in rye, thus the loss of expression in triticale could not be assessed with confidence.

3.6.Discussion

The caleosin gene family

The 11 paralogous groups of caleosin genes present in *T. aestivum* represent the largest gene family for caleosins among the seven species included in this analysis. Though the entire wheat genome has not been completely sequenced, the description of the caleosin gene family appears to be complete or nearly so, since orthologs of all caleosins detected in rye, barley and Brachypodium were also found in wheat. In addition, three homeologous copies for each paralogous group were detected in wheat as well as a fourth member of the *Clo10* group. All 32 of the wheat caleosin genes identified in the EST databases at GenBank were also identified in the WSS and two additional genes, *Clo4-B* and *Clo11-A*, that were not represented in the *T. aestivum* EST database at GenBank were identified in the WSS. This indicates that the depth of the WSS is very robust. The wheat caleosin genes were also represented in the whole genome sequencing database (Genbank Whole-genome shotgun contigs database) [25], though the sequences were fragmentary and did not cover the full length CDS. Brachypodium, which is the most closely related monocot species with a completed and annotated genome sequence available, has seven caleosin genes and rice has six caleosin genes. All caleosin genes in rice and Brachypodium have orthologous sequences in wheat, as judged by sequence similarity, and seen in the phylogram in Figure 1. Since these more distantly related species have gene family members which share branches throughout the tree with sequences from wheat, it seems that the representation of caleosins from wheat is likely complete. The larger number of caleosin genes in wheat, supported by sequences from rye and barley, and the representation of genes from *Oryza* on the major sub-branches in the phylogram indicate that the Triticeae had five gene duplications after the evolutionary separation

from the *Oryza,* and before the separation of the three wheat progenitor species from each other. These duplications in the Triticeae lineage after the separation from rice are the *Clo3* and *Clo4* duplication, the *Clo5* and *Clo6* duplication, the *Clo7* and *Clo8* duplication, and the *Clo2* and *Clo10* duplication, and subsequently the *Clo10* and *Clo11* duplication. The localization of these putative gene duplications on the same chromosomes supports this notion, since gene duplications often occur as tandem duplications [26], though further investigation would be required to demonstrate that the duplications were in tandem. Hv-*Clo3* was not identified in the barley data sets in GenBank; this may represent gene loss, loss of gene expression or may be due to incomplete transcriptome or genome sequence from these species in the current databases. The lack of a full length sequence for rye *Clo2*, but the presence of a gene sequence fragment for it in a high through-put sequence database is indicative of the advanced but still incomplete state of the sequencing for these species. The length of the branches in Figure 1, are scaled according to amino acid differences between sequences, thus providing an estimate of evolutionary distance. The duplication of *Clo10* and *Clo11* appears to have happened more recently than the other duplications, and the presence of *Clo10* and *Clo11* only in wheat, rye and barley provides supporting evidence for recent duplication. A similar evolutionary pattern is also observed for *Clo5* and *Clo6*, though that duplication appears to be older than the *Clo10*, *Clo11* duplication.

In Brachypodium, Bd-*Clo1* and two copies of *Clo2* (Bd-*Clo2-A*, and Bd-*Clo2-B*) [http://brachypodium.org; Bradi5g15410, Bradi5g15417, and Bradi5g15427] are identified as tandem duplications. In addition, Bd-*Clo3-A* [Bradi1g70390] and Bd-*Clo4* [Bradi1g70400] are also adjacent to each other. However *Clo3-A* and *Clo3-B* which share high sequence similarity and are both on chromosome 1, are not closely

linked. Brachypodium provides evidence for both recent and old tandem duplications for caleosins as well as relatively recent non-tandem duplications. Arabidopsis' caleosin genes are largely clustered together in the phylogenetic tree suggesting that the gene duplication events occurred independently in the monocotyledonous and dicotyledonous branches of the tree.

Variation of caleosin gene expression in plant tissues

The diverse expression profile of the caleosin genes in rye and triticale tissues suggests that these calcium binding proteins likely play a broad role during plant development. Whereas some caleosins , such as *Clo6*, exhibited more restricted, tissue-specific patterns of expression, others such as *Clo5*, were expressed in virtually all rye and triticale tissues sampled. It is, therefore, conceivable that some caleosins may have a more general, 'housekeeping' role in most plant tissues and cell types, whereas other caleosins may have tissue- and cell-type specific roles in signaling and regulation during plant development. Evidence that *Clo3* in wheat acts as a GAP for the α subunit of the heterotrimeric G protein [1], raises the possibility that members of the gene family may play key roles in signaling. The expression of other calcium-binding proteins in plants that function in all cell types or that have more restricted, tissue-specific functions, have been previously reported. For example, calmodulin, the predominant Ca^{2+} sensor, plays a critical role in decoding Ca^{2+} signatures into proper cellular responses in numerous tissue types and cellular compartments in eukaryotes [27]. Six members of the calmodulin-like protein gene family are expressed in a developmentally controlled pattern during nodulation in the roots of *Medicago truncatula [28],* and a kinesin-like calmodulin-binding protein, was found to be selectively expressed in the flowers, roots, and leaves of Arabidopsis [29].

Although *T. aestivumClo3* was found to be expressed in several triticale and rye tissues, only a single read was detected in the roots of both species, and there was no expression in several triticale tissues, including seeds. Though these results are in agreement with previous work [2], demonstrating that the expression of At-*Clo3*, the ortholog of Ta-*Clo3*, was undetectable in Arabidopsis root and mature seed using northern analysis, we have detected expression in Arabidopsis in response to abscisic acid treatment (unpublished observations) using transgenic plants with promoter:Gus gene constructs. This underscores the difficulties of tissue-specific expression analysis, since expression can be regulated developmentally, as well as by other factors such as environmental conditions and hormonal responses, and may not be identified in the limited tissue samples that are represented in the cDNA sequence databases.

Interestingly, the gene expression results demonstrate the effect of polyploidization on the expression of R subgenome caleosins. The effect of polyploidization was clearly observed in the case of several caleosins, such as the expression of the rye homeologs *Clo4-R*, *Clo8-R* and *Clo10-R*, in triticale. The rye *Clo4-R* homeolog appeared to be completely suppressed in the roots, and rye *Clo8-R* and *Clo10-R* homeologs were found to be suppressed in anthers, although expression of these caleosins was relatively high in the corresponding rye tissues. Tissue-specific silencing of homeologs from one of the genomes of polyploid species has been reported in other allopolyploids including *Tragopogon miscellus*[30], and *Gossypium hirsutum* [31], though the mechanisms that lead to suppression are somewhat speculative at this point. Since these genes are expressed in other triticale tissues, gene deletion is clearly not the explanation for suppression, and mechanisms related

to chromatin remodelling or the incompatibility of signaling and regulation pathways of parental genomes in newly derived polyploids warrants investigation.

3.7. Conclusion

An apparent full set of wheat caleosin gene sequences was acquired as full length cDNAs, the open reading frames were identified, and the peptide sequences were obtained. The gene sequences were confirmed with the WSS database. One member of the caleosin gene family, *Clo3*, has previously been identified as a stress inducible gene encoding a Ca^{2+} binding protein that acts as a negative regulator of the α subunit of the heterotrimeric G protein GA3 [1]. The identification and full description of the gene family for caleosins can be a significant step in further investigating the role of members of this gene family in signaling and regulation. The very diverse pattern of tissue-specific expression indicates a potential for a very broad role in signaling and regulation throughout plant development.

REFERENCES

1. Khalil HB, Wang Z, Wright JA, Ralevski A, Donayo AO, and Gulick PJ: Heterotrimeric Gα subunit from wheat (*Triticum aestivum*), GA3, interacts with the calcium-binding protein, Clo3, and the phosphoinositide-specific phospholipase C, PI-PLC1. *Plant Mol Biol*2011, 77: 145-158.
2. Takahashi S, Takeshi K, Yamaguchi-Shinozaki K, and Shinozaki K: An Arabidopsis Gene Encoding a Ca2+-Binding Protein is induced by Abscisic Acid during Dehydration. *Plant Cell Physiol* 2000, 41:898-903.
3. Aubert Y, Vile D, Pervent M, Aldon D, Ranty B, Simonneau T, Vavasseur A, Galaud JP: RD20, a Stress-Inducible caleosin, Participates in Stomatal Control, Transpiration and Drought Tolerance in *Arabidopsis thaliana*.*Plant Cell Physiol* 2010, 51:1975-1987.
4. Liu H, Hedley P, Cardle L, Wright KM, Hein I, Marshall D, Waugh R: Characterisation and functional analysis of two barley caleosins expressed during barley caryopsis development. *Planta* 2005, 221:513–522.
5. Chiang CJ, Che, CJ, Lin LJ, Chang CH, Chao YP: Selective delivery of cargo entities to tumor cells by nanoscale artificial oil bodies. *J Agric Food Chem* 2010, 58:11695-702.
6. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ: Basic local alignment search tool. *J Mo. Biol* 1990, 215:403-410.
7. Huang X, and Madan A: CAP3: A DNA sequence assembly program. *Genome Res* 1999, 9:868-77.
8. ExPASy Translate Tool [www.web.expasy.org/translate/]
9. International Wheat Genome Sequencing Consortium, Sept. 1, 2012[http://urgi.versailles.inra.fr/srs83/displayTool.do?toolName=BlastN]
10. International Barley Sequencing Consortium [http://webblast.ipk-gatersleben.de/barley/viroblast.php] Sept 1, 2012
11. Brachypodium database []
12. NCBI Batch Conserved Domain Search tool

[http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi]

13. RCSB Protein Data Bank [http://www.rcsb.org/pdb/home/home.do]
14. InterPro Scan Sequence Search. European Bioinformatics Institute, 2011. [http://www.ebi.ac.uk/Tools/pfa/iprscan/]
15. Simple Modular Architecture Research Tool (SMART) [http://smart.embl-heidelberg.de/]
16. Steel MA, Fu YX: Classifying and counting linear phylogenetic invariants for the Jukes-Cantor model. *Journal of Computational Biology* 1995, 2:39-47.
17. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S: MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011, 28:2731-2739.

18. Whelan S, Goldman N: A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001, 18:691-9.

19. Goecks J, Nekrutenko A, and Taylor J: Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010, 11:R86.

20. Li H and Durbin R: Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009, 25:1754-1760.

21. ClustalW2- Multiple Sequence Alignment [www.ebi.ac.uk/Tools/msa/clustalw2/]

22. Farajalla R and Gulick PJ: The alpha-tubulin gene family in wheat (*Triticum aestivum L.*) and differential gene expression during cold acclimation. *Genome* 2007, 50(5):502-10.

23. Miftahudin RK, Ma XF, Mahmoud AA, Layton J, Milla MA, Chikmawati T, Ramalingam J, Feril O, Pathan MS, Momirovic GS, Kim S, Chema K, Fang P, Haule L, Struxness H, Birkes J, Yaghoubian C, Skinner R, McAllister J, Nguyen V, Qi LL, Echalier B, Gill BS, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorák J, Dilbirligi M, Gill KS, Peng JH, Lapitan NL, Bermudez-Kandianis CE, Sorrells ME, Hossain KG, Kalavacharla V, Kianian SF, Lazo GR, Chao S, Anderson OD, Gonzalez-Hernandez J, Conley EJ, Anderson JA, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Nguyen HT, Gustafson JP: Analysis of expressed sequence tag loci on wheat chromosome group 4. *Genetics* 2004, 168:651-663.

24. Riaño-Pachón DM, Nagel A, Neigenfind J, Wagner R, Basekow R, Weber E, Mueller-Roeber B, Diehl S and Kersten B: GabiPD: the GABI primary database – a plant integrative 'omics' database. *Nucleic Acids Research* 2009, 37(Database issue): D954-9.

25. Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N: Analysis of the bread wheat genome using whole-genome shotgun sequencing.*Nature* 2012, 491(7426):705-10.

26. The Arabidopsis Genome Initiative [http://www.genomenewsnetwork.org/articles/04_00/what_makes_chart.shtml]

27. Kim MC, Chung WS, Yun DJ, Cho MJ: *Calcium and calmodulin-mediated regulation of gene expression in plants. Mol Plant 2009, 2: 13–21.*

28. Liu J, Miller SS , Graham M, Bucciarelli B, CM, Sherrier DJ, Samac DA, Ivashuta S, Fedorova M, Matsumoto P, Gantt JS, and Vance CP: Recruitment of Novel Calcium-Binding Proteins for Root Nodule Symbiosis in *Medicago truncatula*. *Plant Physiol* 2006, 141: 167–177.

29. Reddy AS, Narasimhulu SB, Safadi F, and Golovkin M*:* A plant kinesin heavy chain-like protein is a calmodulin-binding protein. *Plant J* 1996, 10: 9–21.

30. Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE: Gene loss and silencing in *Tragopogon miscellus*(Asteraceae): comparison of natural and synthetic allotetraploids.*Heredity* 2009,103: 73–81.

31. Adams KL, Cronn R, Percifield R, and Wendel JF: Genes duplicatedby polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Nat AcadSci* 2003, 100:4649–4654.

SUPPLEMENTAL FIGURES

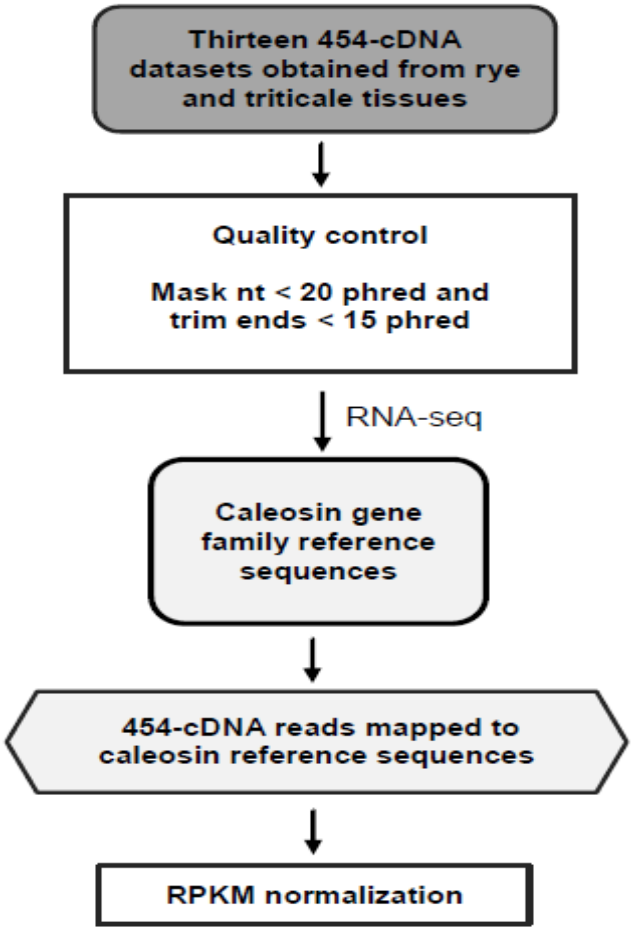


Figure S1. RNA-seq workflow used to measure the abundance of caleosin gene family members in *T. aestivum* GenBank EST database and in thirteen rye and triticale 454-cDNA libraries expressed in different tissues. Quality control of 454-cDNA reads was performed using the tools that have been simplified by free browser-based access through Galaxy server [19]. This step involved converting 454-sff format to fastq format, trimming low quality reads using FASTQ quality trimmer, and masking low quality score nucleotides using FASTQ masker. High quality rye and triticale 454-cDNA reads obtained from different tissues were aligned to caleosin gene family FL-cDNAs using BWA-SW aligner. The expression of each gene was normalized using RPKM normalization method. RNA-seq was used to calculate the relative frequency of *T. aestivum* ESTs in the GenBank database.

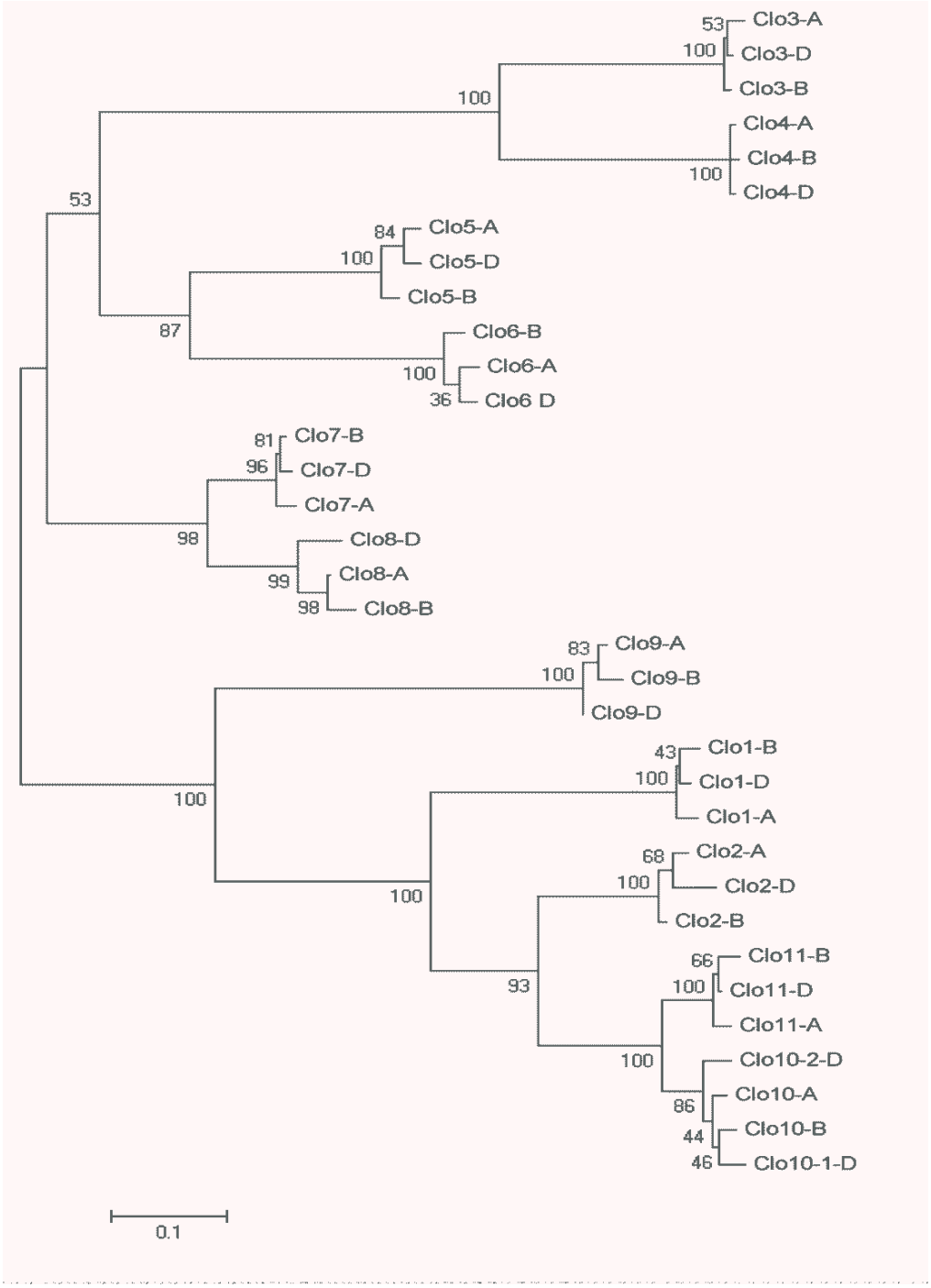


Figure S2. Molecular phylogenetic analysis by maximum likelihood method of *T. aestivum* caleosin nucleotide sequences. The evolutionary history was inferred by using the maximum likelihood method based on the Jukes-Cantor model [17]. The tree with the highest log likelihood (-6347.7079) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The number of bootstrap replications was set to 100.

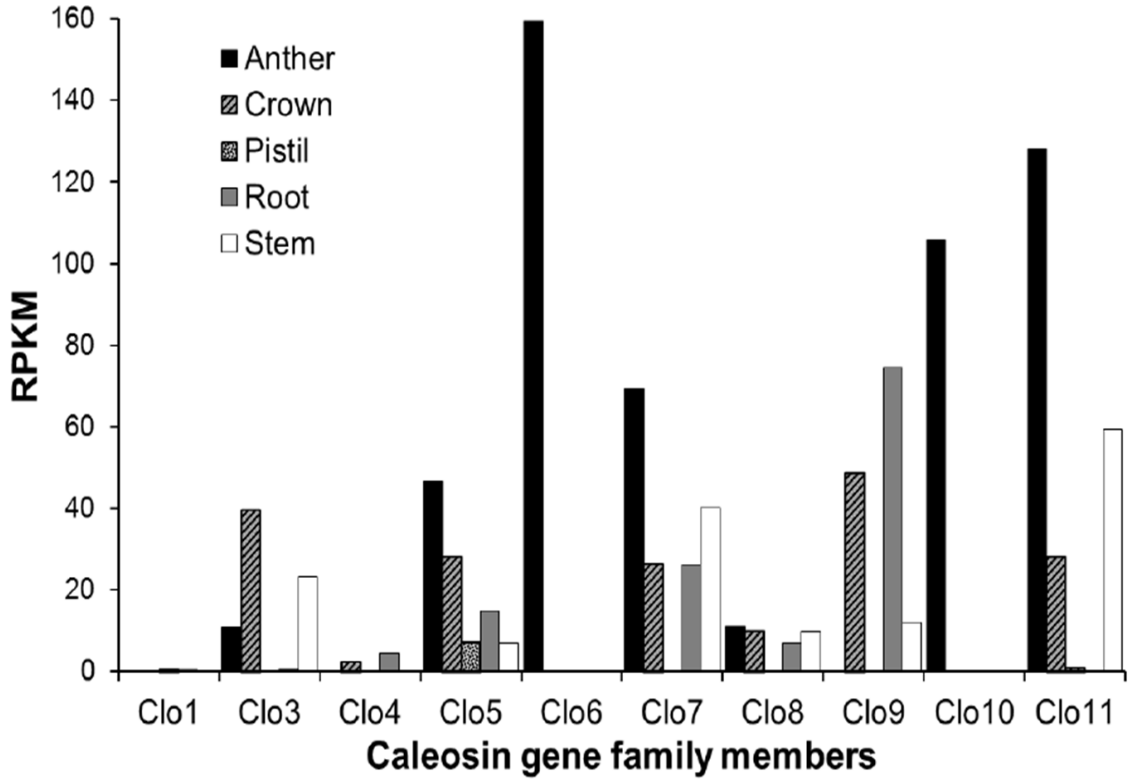


Figure S3. A representation of the expression abundance of ten caleosin gene family members in five different rye tissues. The expression of caleosin gene family members was estimated in anther, crown, pistil, root and stem rye tissues using RNA-seq analysis. The aligned 454-cDNAs to each caleosin member were counted, then normalized based on gene lengths and library depths using the RPKM method.

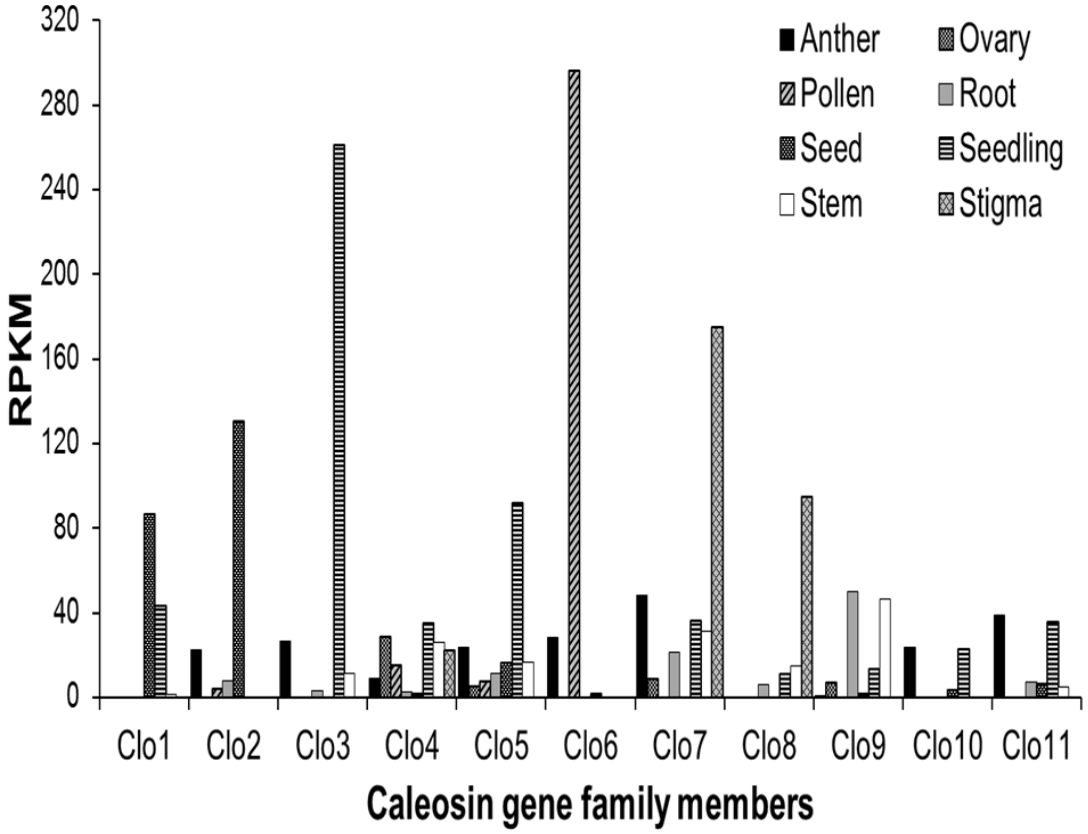


Figure S4. A representation of the expression abundance of eleven caleosin gene family members in eight different triticale tissues. The expression of thirty-two caleosin gene family members was individually measured in anther, ovary, pollen, root, seed, seedling, stem and stigma triticale tissue using RNA-seq analysis. The aligned 454-cDNAs to each caleosin member were counted, then normalized based on gene lengths and library depths using RPKM method. The total expression of the three homeologs was combined to graph the expression of each paralog, except *Clo2* represented the expression of only the A and B homeologs.

SUPPLEMENTAL TABLES

Table S1. Summary of454-cDNA libraries obtained from rye and triticale tissues.

Plant	Tissue	No. of 454-cDNAs	Read Length Average (nt)
Rye	Anther	694529	343
	Crown	379863	392
	Pistil	864151	313
	Root	1664768	365
	Stem	924268	365
Triticale	Anther	1129578	365
	Ovary	393743	329
	Pollen	550758	372
	Root	371416	474
	Seed	611062	430
	Seedling	1550183	350
	Stem	1771253	323
	Stigma	296740	327

Table S2. The impact of polyploidization on triticale caleosin gene family members assigned to the R subgenome.

		Rye Expression		R[a] Expression in Triticale			
Caleosin	Tissue	Reads	RPKM[b]	Reads	RPKM[b]	Expression Change	Fold Change
Clo4	Root	8	4.57	0	0	Suppression	
Clo7	Root	39	26.29	3	9.07	↓ Expression	3
Clo9	Root	137	74.53	6	14.96	↓ Expression	5
Clo1	Stem	0	0	3	1.37	Novel Expression	
Clo3	Stem	17	23.11	12	8.51	↓ Expression	5
Clo4	Stem	0	0	30	16.1	Novel Expression	
Clo5	Stem	6	7.02	3	1.83	↓ Expression	4
Clo11	Stem	61	59.35	7	3.55	↓ Expression	17
Clo5	Anther	30	46.7	9	8.61	↓ Expression	5
Clo6	Anther	101	159.28	6	5.82	↓ Expression	27
Clo8	Anther	6	11.13	0	0	Suppression	
Clo10	Anther	67	105.78	0	0	Suppression	

[a]Rye subgenome

[b] Reads Per Kilobase per Million

Chapter 4.Identification of rye genes silenced and deleted from rye genome in allohexaploid triticale: the nature of rye homeologs prone to be suppressed

4.1. Outline and contribution of colleagues

Chapter 4 of the thesis is a study of the impact of polyploidization on gene expression in triticale. The expression of 23,503 rye cDNA contigs obtained from the Canadian Triticale Biorefinery Initiative (CTBI) project were compared in the same tissues in both rye and triticale. High throughput cDNA sequence comparisons between the diploid rye and the hexaploid triticale detected suppression of expression of approximately 2% of rye genes surveyed in the triticale allopolyploid. The expressions of 23,503 rye cDNA contigs was analyzed in 454-cDNA libraries obtained from anther, root and stem from both triticale and rye as well as in five 454-cDNA data sets created from ovary, pollen, seed, seedling shoot and stigma from triticale. Among these, 112 rye cDNA contigs were found to be totally suppressed in all triticale tissues, although their expression was relatively high in rye tissues. PCR assays showed that 6 out of 10 candidate suppressed genes were deleted from the triticale genome. The comparison of rye genes silenced in triticale and the rye random set, a control set consisting of 200 sequencesto the *T. aestivum*, *T. urartu* and *Ag. tauchi*genomes revealed that the degree of similarity in rye genes silenced in triticale was significantly lower than the global average similarity compared to the tested set of 200 rye genes.I carried out all the bioinformatic and gene ontology analyses. M. Ehdaeivand and I contributed equally to the PCR and RT-PCR confirmation of gene deletion or suppression. Dr. Gulick and I designed, wrote and revised the manuscript. I expect to submit this manuscript for publication at the end of this year.

4.2. Abstract

Background

One of the most important evolutionary processes in plants is polyploidization. Some polyploids are preadapted to the unification of two or more different genomes into one cell by rapid and reproducible gene suppression and elimination. The synthetic triticale**,**Triticosecale, is an intergeneric allohexaploid plant generated from Triticum durum (*Triticum turgidum*) and rye (*Secale cereale*). Molecular investigations reported DNA elimination of repetitive DNA and low-copy sequences from the rye genome in triticale (Ma *et al.,* 2004; Ma and Gustafson, 2006). The mechanisms behind these significant changes in the rye genome in allohexaploid triticale call for new tools to provide a clearer understanding of these genetic changes. The new technologies of high-throughput DNA sequencing offer a particular advantage over other types of analysis in that a large number of genes can be surveyed and the sequence of the transcript provides a clear way to distinguish transcripts from the different homeologous copies of the genes. Transcription profiles were compared using rye reference assemblies to compare the transcription levels of rye genes in triticale and rye tissues to identify the classes of rye genes silenced and deleted genes in triticale.

Results

High-throughput cDNA sequence comparisons between the diploid rye and the hexaploid triticale detected suppression of expression of approximately 2% of rye genes surveyed in triticale. The expression levels of 23,503 rye cDNA reference contigs were analyzed in 454-cDNA libraries obtained from anther, root and stem from both triticale and rye as well as in five 454-cDNA data sets created from ovary, pollen, seed, seedling shoot and stigma from triticale. Among these, 112 rye cDNA

contigs were found to be totally suppressed in all triticale tissues, although their expressions were relatively high in rye tissues. Suppressed rye genes were found to be strikingly less similar to their closest BLASTN matches in the wheat genome than random rye genes, a test set of 200 rye genes. The comparison of rye silenced genes in triticale and the rye control setto the International Wheat Genome Survey Consortium, Wheat Survey Sequence database (IWGSC-WSS), revealed that 89% of the rye genes silenced in triticale do not have a best match in *T. aestivum* with sequence identity higher than 90%, while only 41% of random rye contigs had best matches with such low sequence similarity in wheat. The comparisons to diploid wheat ancestors, *T. urartu* and *Ag.tauschii*draftgenomes, were consistent with the wheat genome survey database comparison observation. PCR assays showed that six out of ten candidate suppressed genes were deleted from the triticale genome.

Conclusion

The strikingly low similarity between suppressed rye sequences and their closest matches in wheat genomes suggests gene deletion or suppression involved in the recognition of non-similar sequences in the allopolyploid genome. The comparison analyses of these genes to wheat genomes indicated a bias for the suppression of rye genes that do not have a close homeolog in the other *Triticum* genomes.

4.3. Background

The cause of the striking alteration of plant genomes after allopolyploidization has been a central question in allopolyploid genome evolution. In general, the initial combination of two or more genomes in one organism is accompanied by considerable genomic reorganization and widespread changes in gene expression relative to that of the parental species. Plants, unlike animals, tolerate the changes subsequent to interspecific genome hybridization and chromosome duplication such that polyploidy is relatively common among plant species. The studies of paleopolyploids indicate the diploidization process involves major genome rearrangements including chromosome loss (Lysak *et al.,* 2006), reduction in chromosome number by various forms of chromosome fusion and rearrangements, gene loss (Madlung*et al.,* 2005), changes of gene expression (Lee and Chen, 2001) and in some cases genome expansion (Renny-Byfield *et al.,* 2013). More recent polyploids such as *Triticum* and *Brassica* species thought to have formed (0.5 MYA) and (5,000-10,000 YA), respectively, maintain polyploid chromosome numbers but have diploid chromosome pairing patterns during meiosis. The genomes maintain synteny, but they nevertheless undergo gene loss (Feldman *et al.,* 1997; Song *et al.,* 1995), suppression (Kashkush *et al.,* 2002), inversion (Brubaker *et al.,* 1999) and translocation events (Maestra and Naranjo, 1999).

Although the mechanisms of gene suppression and elimination are still unknown, several studies have found that these changes occur rapidly and more frequently in one parental genome of an allotetraploid as reported for *Triticum* (Kashkush *et al.,* 2002; Feldman and Levy, 2009), *Tragopogon* (Tate *et al.,* 2009; Koh *et al.,* 2010) and *Gossypium* (Adams *et al.,* 2004). The preferential control of traits by the genes from one parental genome, referred to as genetic diploidization, is

apparently not random in allopolyploids and natural selection for balanced gene dosage effects have a strong impact on this process. Phenotypic comparisons of allotetraploid and allohexaploid wheat, and their diploid parents indicated that genes controlling traits related to domestication such as autogamy, non-brittle spike, free-threshing glumes, and large kernel size are predominately controlled by genes of the A genome and, in contrast, the B and D genomes, preferentially control biotic and abiotic stress-regulated gene expression (reviewed in Feldman and Levy, 2009 and in Feldman *et al.,* 2012).

A significant degree of genome alteration occurs directly after allopolyploidizations. The nuclear amount of both natural and newly generated wheat was found to be 2-10 % less than the sum of the amount of DNA of their parents (Eilam *et al.,* 2008; Eilam *et al.,* 2010). The synthetic allopolyploid triticale has a genome structure similar to hexaploid bread wheat except that it has rye as one of its progenitors (Figure 1). This hexaploid genome was found to have a higher degree of DNA reduction with measurements of DNA loss in the range of 22-30 % (Boyko *et al.,* 1988; Bennet and Leitch, 2011). Hence, understanding what controls DNA loss in allopolyploids would become clearer with further study of wide hybridizations, since deletions occur at high frequencies. The recent allopolyploid, triticale, is a useful model for identifying these rapid changes at both the level of gene expression and genomic restructuring soon after allopolyploidization.

Molecular techniques have been developed to facilitate the global estimation of homeolog suppression in both natural and synthetic allopolyploids. The percentage of suppressed genes varies from one study to another, probably as a result of methodological differences. One of these methods, cDNA-AFLP, a qualitative method employed to study transcriptional changes, detected homeologous gene

silencing in several synthetic allopolyploids. These studies found gene silencing for approximately 5% of genes in allopolyploid cotton (Adams *et al.,* 2004), between 1% and 5% of genes in allotetraploid wheat (Kashkush *et al.,* 2002) and 0.4% of genes in Arabidopsis (Comai *et al.,* 2000).

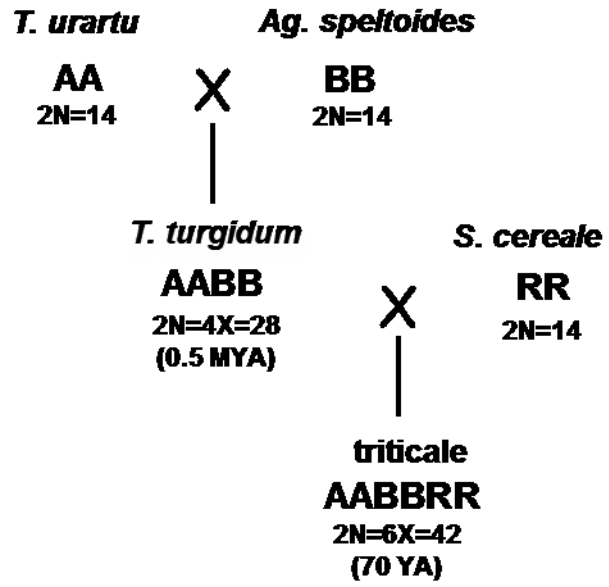


Figure 1. Hexaploid triticale allopolyploid.
T. turgidum was derived in nature from a cross between *T. urartu* and an unknown ancestor thought to be closely related to *Ag. speltoides* which is estimated to have occurred approximately 0.5 million years ago. The first synthetic hybrid triticales were developed approximately seventy years ago by the intergeneric hybridization between *T. turgidum,* allotetraploid durum wheat (cytoplasmic doner parent) and rye, *S. cereale*.

In addition, these studies detected changes of tissue specific gene expression, termedsubfunctionalization (Adams *et al.,* 2004). More recently, comparative gene expression studies by microarray analysis revealed that 19% of the genes analyzed in wheat showed differential expression between homeologous gene copies (Akunova*et al.,* 2010). Microarrays and cDNA-AFLP analyses are highly sensitive tools used in several molecular studies to detect changes of gene expressions in polyploids, however, these methods are technically demanding, especially in the analysis of a large number of bands in AFLP and the variability of fluorescent signals in microarrays. Estimating gene expression using second generation high-throughput cDNA sequencing techniques offers the advantage of increasing the accuracy of transcript identification directly from the sequence rather than by DNA or RNA hybridization, thus next-generation sequencing data help to overcome limitations of these previous techniques. Here, we are investigating the impact of allopolyplodization on the rye coding sequences in the triticale transcriptome at a high level of resolution using second generation Roche 454-cDNA, a high-throughput cDNA sequencing technology. The next generation sequencing data are a particularly important advancement for analysis of polyploids such as wheat or triticale, since homeologous genes have very high sequence similarity and often cannot readily be distinguished by hybridization techniques. A comparison of the transcription level of 23,503 rye reference contig assemblies between triticale and rye tissues can shed light on the classes of rye genes prone to be silenced following allopolyploidization.

4.4. Results and Discussion

Rye silent genes in triticale

A comparison between the hexaploid triticale and diploid rye 454 high-throughput RNA-seq profiling data detected suppression of expression in approximate 2% of genes surveyed in the triticale that originated from its rye background. The expression of genes of a reference set of 23,503 rye cDNA assemblies was analyzed in 454-cDNA libraries obtained from anther, root and stem of both triticale and rye as well as from five triticale data sets created from ovary, pollen, seed, seedling shoot and stigma (Additional File 1: Figure S1). There were 465 genes whose expression was detected in rye but were not found in any of the eight triticale libraries used in the study. Further analysis was narrowed to a smaller subset of genes that had relatively high expression in rye, namely 112 rye genes that were represented by at least 10 transcripts in at least one of the rye tissues but which were completely absent among the 6,674,733 triticale 454 reads.

Rye sequence comparison to *Triticum* and *Aegilops* databases

To investigate the potential relationship between the triticale genes from the rye sub-genome that were silenced in the allopolyploid, the corresponding rye contigs were compared to the genome sequence assemblies from *T. aestivum*. The comparison revealed the striking feature that most of the rye genes silenced in triticale did not have a close homolog in *T. aestivum,* indicating that they apparently do not have a homeologous copy in triticale. The distribution of the percent identity between the 112 silenced rye genes to their closest matches in the A and B genomes was significantly lower than a parallel comparison of a set of 200 randomly selected rye genes (Figure 2A). A comparison of rye silenced genes in triticale to (IWGSC-WSS) database indicated that 89% of rye genes silenced in

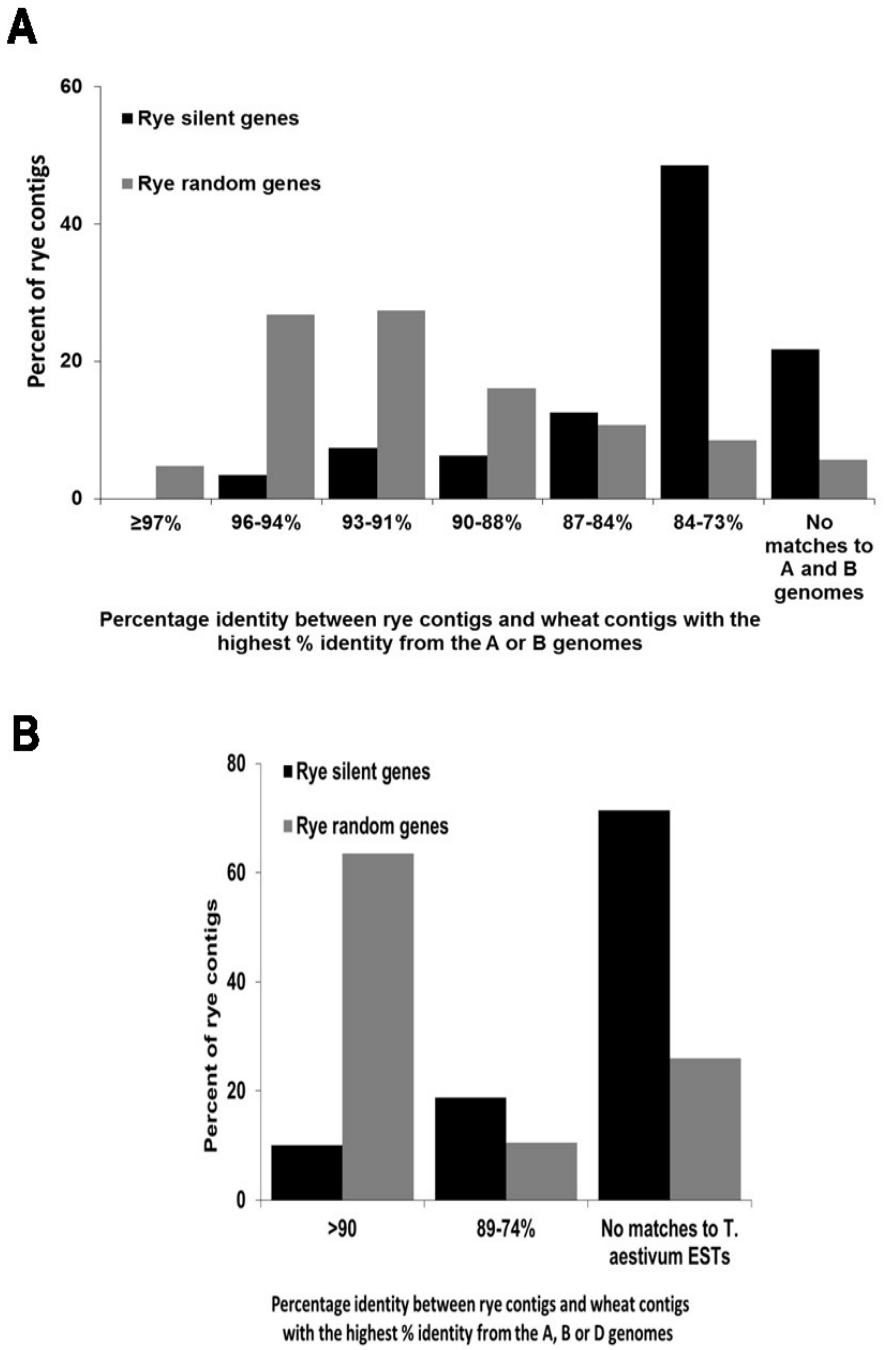


Figure 2. Comparisons of rye genes suppressed in triticale and 200 random selected rye genes to *T. aestivum* databases.

(A) The majority of rye silenced genes in triticale has lower sequence identity to wheat genome sequences (black bars) than the control set consisted of 200 rye random sequences (gray bars). **(B)** About 10 % of Rye silenced genes in triticale has more than 90% sequence identity percentages to *T. aestivum* GeneBank EST database (black bars), while 64% of the control set of 200 rye random sequences had high sequence identity to wheat ESTs (gray bars).

triticale do not have a best match in *T. aestivum* with sequence identity higher than 90%, whereas 59% of random rye contigs have a best hit of 90% or higher in *T. aestivum* (Figure 2A). The averageDNA sequence identity between rye genes suppressed in triticale and their most similar contigs in the A and B genomes in *T. aestivum* was only 81%. This degree of identity was significantly lower than the global average of 91% identity between the set of 200 randomly selected rye genes and their best matches in the A and B genomes of *T. aestivum*. There appears to be a bias for the suppression of rye genes that do not have a close homeolog in the *Triticum* genome. Previous studies of well-characterized gene families in the triticale found sequence identity between homeologous sequences to be in the range of 95-97 %. Ten members of the caleosin gene family in rye ranged from 99% to 90% identical to their orthologs in *T. aestivum* (Pham *et al.,* unpublished). This degree of identity is similar to that between homeologous gene copies from the A, B and D genomes of *T. aestivum* and their orthologs in *Hordeum vulgare*, another member of the Triticeae. Members of the α-tubulin gene family show similarly high levels of similarity between homeologous gene copies within *T. aestivum,* (Ridha-Farajalla and Gulick, 2007). The examples of these two well-studied gene families as well as the data from the 200 random set of rye genes indicate that most of the rye genes silenced in triticale do not appear to have homeologous gene copies from the A and B genomes.

The hypothesis that wheat genes that had lower than 90% identity with rye silent genes are not homeologous copies of the rye genes was further investigated.We compared high scoring hits among the wheat sequences back to the rye gene reference set, the 23,503 rye assemblies, to determine if they had matches that were more similar than those of the rye silent genes. We first compared the rye silent genes to

the *T. aestivum* GenBank EST database, subsequently we compared the highest scoring wheat hits that had lower than 90% identity back to the rye reference set. Though the number of hits in the wheat EST database was relative modest, some 29 wheat ESTs, it showed that 34% of these had matches higher than 95% identity in the rye reference set, 66% had a match with higher than 90% identity and 62% had matches with percent identity higher than the identity of the initial match between the silent rye gene and the wheat EST. Since the wheat EST hit by the silent rye gene has a closer match in the rye reference sequence set, it is apparent that the initial wheat EST hit does not represent the homeolog of the rye silent gene, and by extension the highest hit in the IWGSC-WSS database that is lower than 90% identity, is not a homeolog to the rye gene (Additional File2: Table S1). The threshold for distinguishing homeologs may be higher, near 95%, but it is not possible to define this distinction with the present data. The depth of the *T. aestivum* GenBank EST database is somewhat limited, since most of the silent genes had no match in the database above default parameters, however, this pattern of gene identity supports the hypothesis that most of the rye silent genes whose highest matches have less than 90% identity in the wheat genome do not have a homeologous gene copy in the wheat genome.

These results give rise to an empirical question: Do rye suppressed genes in triticale appear to lack their homeologs because the comparisons were *T. aestivum*-based analyses? This genome has experienced two allopolyploidizations, so these types of genes might be selectively lost following polyploidization and thus were previously eliminated from the *T. aestivum* genome. In addition, the triticale analyzed here is derived from *T. trugidum*, which carries only the A and B genomes. The hexaploid *T. aestivum* was used as the primary basis of sequence comparison since the

data sets available for *T. aestivum* are far larger than those for other *Triticum* species. To investigate this question, the same comparison between rye silenced genes in triticale and the rye random set was performed with the draft genomes of two of the diploid progenitors of *T. aestivum*. The draft genome of *T. urartu*, the A genome donor, includes 499,222 scaffolds (Ling *et al.,* 2013), and *Ag. tausshii*, the D genome donor, include 429,893 scaffolds (Jia *et al.,* 2013); both are available in the GenBank NR database. Although both data sets have less depth than that of *T. aestivum*, the comparisons support the previous observations based on *T. aestivum*; only 19% and 14 % of the rye silenced genes had matches with 90% or higher sequence similarity in A or D genome, respectively, whereas approximately 59% and 63 % of the randomly selected rye control set of genes had matches with high degrees of identity in the A and D genomes (Figure 3A and 3B).

Gene descriptions and ontologies of rye silent sequences in triticale

A comparison of the 112 silenced rye gene assemblies to the GenBank NR databases though the BLAST2GO workstation (Conesa *et al.,* 2005), resulted in 60 contigs with significant similarity with GenBank protein sequences (Table 1) and those without a match in the protein database are tabulated in Additional File 2: Table S2. The ontology of the rye silent genes that found through Blastx search in the GenBank NR database was varied, but most of these code for proteins with catalytic activity and proteins with binding activity, especially nucleotide-binding and ion-binding proteins (Additional File 1: Figure S2). The BLASTX results revealed five rye disease resistance genes with a NB-ARC domain, a novel nucleotide-binding signalling motif shared by plant disease resistance gene products and regulators of cell death in animals,suppressed in all triticale tissues that were sampled. In contrast, the NB-ARC containing domain rye genes, namely R1, R8, R11, R19 and R20, were

expressed in rye stem and their abundance ranged from 48.6 (RPKM) to 13.9 (RPKM). The proteins most similar to R20, R19, R1, R11 and R8, in the GenBank NR database were encoded by disease resistance genes RGA1 (GB: EMT10593.1), and RGA3 (GB: EMT03843.1) from *Ag. tauschii*, and two disease resistance genes. RPP13 from the same *T. urartu* (GB: EMS68441.1) and RPP13 from *Ag. tauschii* (GB:EMT01897.1).

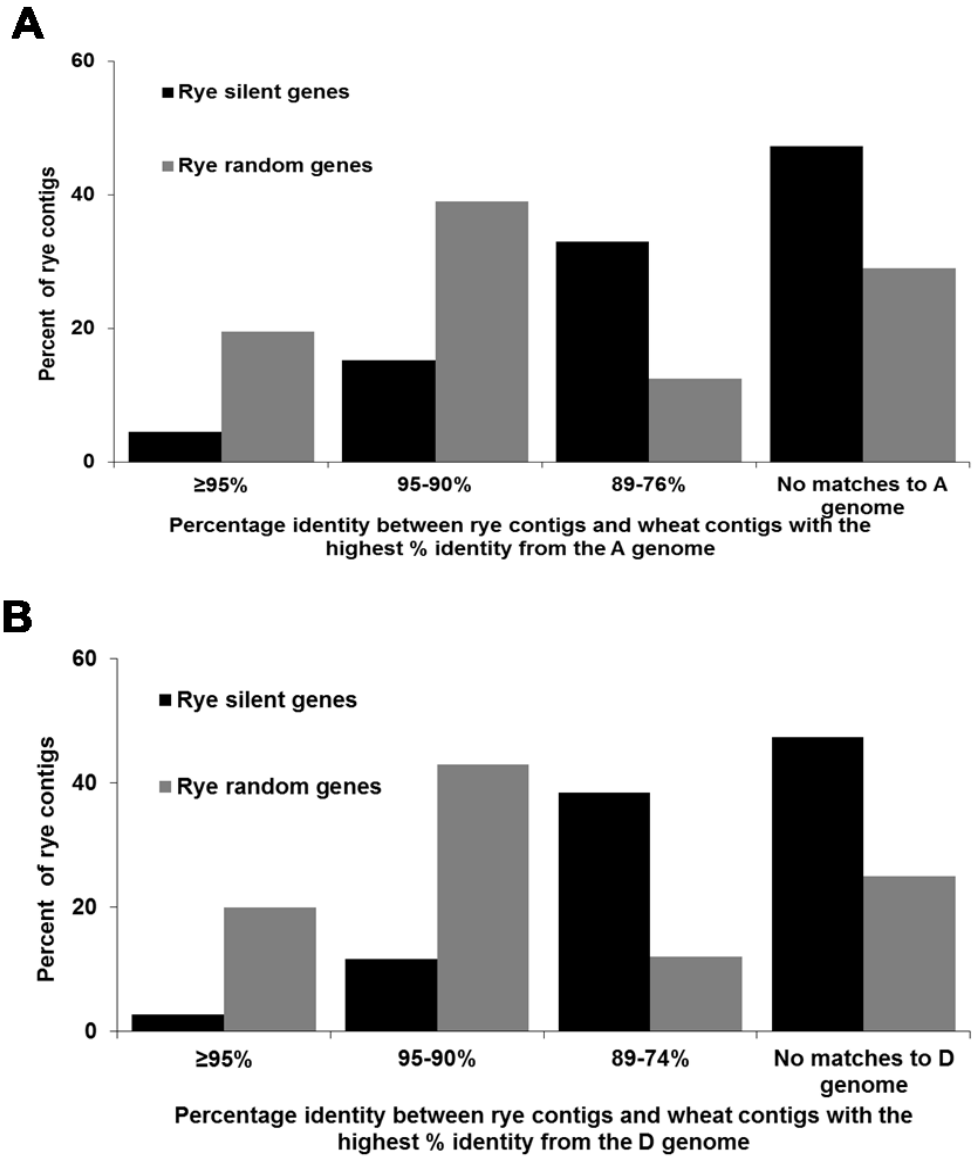


Figure 3. Comparison of rye genes suppressed in triticale to the diploid genomes of *T. urartu* and *T. tauschii*.
Rye silenced genes in triticale (black bars) have lower percent identity than random rye contigs (gray bars) to their highest scoring hits in the genomes of (A) *T. urartu*, the ancestor of A genome in *T. aestivum* and *T. turgidum* and (B) *T. tauschii*, the ancestor of D genome in *T. aestivum*.

Table 1. Annotation of rye genes silenced in triticale[a].

Rye ID	Sequence Description	#[b] Hits	Min[c]. E-value	Mean Similarity	Tissues	Reads	RPKM[d]
R1	Disease resistance protein RPP13	100	2.39E-92	0.7788	Stem	44	48.6
R2	Monothiol glutaredoxin-chloroplastic	1	1.46E-13	0.74	Stem	98	39.4
R3	vacuolar-processing enzyme	100	4.35E-53	0.6262	Stem	127	38.2
R4	Hypothetical protein	1	2.27E-17	0.54	Stem	96	34.7
R5	ULP1 protease C-terminal catalytic domain containing protein	22	1.97E-18	0.4868	Stem	21	25.5
R6	Na^+ H^+ antiporter	36	2.52E-24	0.6236	Stem	27	23.7
R7	Coatomer subunit delta	13	2.69E-13	0.8462	Stem	19	21.7
R8	Disease resistance protein RPM1	100	0	0.6127	Stem	54	21.7
R9	U3 small nucleolar RNA-associated protein 18-like	77	1.79E-93	0.4681	Stem	28	20.7
R10	SKP1-like protein 1A	100	1.2E-163	0.7212	Stem	30	20.5
R11	Disease resistance protein RPP13	100	2.8E-107	0.7074	Stem	16	19.1
R12	Hypothetical protein	3	2.83E-15	0.52	Stem	26	18.4
R13	Hypothetical protein	1	1.3E-16	0.75	Stem	46	17.7
R14	Protein far1-related sequence 5	16	3.12E-23	0.5981	Stem	24	17.3
R15	Protein	100	0	0.5356	Stem	29	17.1
R16	Fact complex subunit SPT16	100	4.3E-129	0.6229	Stem	16	16.8
R17	Predicted protein	2	5.99E-17	0.63	Stem	18	14.3

[a] Annotation based on BLAST2GO with minimum E value of e-10 and a maximum of 100 hits.
[b] Number of hits using a maximum of 100 hits as a threshold
[c] Minmum
[d]Reads Per Kilobase per Million

Continued

Rye ID	Sequence Description	# [b] Hits	Min[c]. E-value	Mean Similarity	Tissues	Reads	RPKM[d]
R18	MDR-like ABC transporter	14	1.94E-17	0.9443	Stem	14	14.2
R19	Disease resistance protein RGA1	100	4.5E-155	0.632	Stem	27	14.0
R20	Powdery mildew resistance protein PM3 variant	100	9.79E-92	0.4844	Stem	15	13.9
R21	Malonyl-coenzyme A:anthocyanin 3-o-glucoside-6 -o-malonyltransferase	7	2.78E-29	0.6171	Stem	17	13.7
R22	Hypothetical protein	7	1.43E-41	0.5757	Stem	15	12.4
R23	DNA excision repair protein ERCC-6-like	100	1.5E-103	0.5359	Stem	21	12.1
R24	Hypothetical protein	5	2.96E-15	0.844	Stem	15	10.6
R25	Protein far1-related sequence 5-like	100	0	0.5394	Stem	16	8.8
R26	CCR4-not transcription complex subunit 1-like	2	1.57E-23	0.66	Stem	10	8.6
R27	Protein	100	0	0.4725	Stem	11	7.5
R28	Hypothetical protein	6	6.95E-23	0.7183	anther and stem	17 and 13	15.1 and 8.3
R29	Anthranilate n-benzoyltransferase protein 1-like	100	0	0.5905	Root	26	84.0
R30	Protein	100	6.68E-59	0.6094	Root	17	69.2
R31	Chitinase	100	1.8E-103	0.825	Root	14	68.0
R32	Jasmonate-induced protein	100	2.54E-65	0.5075	Root	15	66.9

[a] Annotation based on BLAST2GO with minimum E value of e-10 and a maximum of 100 hits.
[b] Number of hits using a maximum of 100 hits as a threshold
[c] Minmum
[d]Reads Per Kilobase per Million

Continued

Rye ID	Seqequence Description	#[b] Hits	Min[c]. E-value	Mean Similarity	Tissues	Reads	RPKM[d]
R33	Lecithin-cholesterol acyltransferase	100	1.3E-172	0.6018	Root	13	51.0
R34	Esterase	100	0	0.6574	Root	12	44.9
R35	Hypothetical protein	1	3.98E-16	0.62	Root	19	42.5
R36	Hypothetical protein	2	3.2E-54	0.565	Root	16	41.2
R37	Cytochrome P450	100	0	0.6591	Root	11	34.7
R38	Fusion protein	18	0	0.6222	Root	13	20.8
R39	Hypothetical protein	1	9.51E-67	0.49	Anther	364	316.4
R40	AF111710_1 DNAJ-like protein	13	8.86E-87	0.5962	Anther	180	225.3
R41	F-box domain containing protein	76	1.17E-96	0.4496	Anther	154	182.2
R42	Hypothetical protein	25	0	0.516	Anther	104	96.0
R43	F-box domain containing protein	43	1.7E-86	0.533	Anther	62	84.7
R44	Hypothetical protein	7	3.35E-20	0.5429	Anther	80	49.0
R45	Hypothetical protein	1	2.33E-70	0.8	Anther	42	43.5
R46	Retrotransposon unclassified	84	2.92E-42	0.5408	Anther	29	41.6
R47	ENT-copalyl diphosphate synthase chloroplastic	13	3.85E-27	0.8531	Anther	27	41.1
R48	Protein	100	1.33E-77	0.6174	Anther	25	35.3
R49	60s ribosomal protein l7	7	1.86E-13	0.6343	Anther	24	30.0
R50	AC098566_17 4 protein	10	7.49E-30	0.589	Anther	22	24.4

[a] Annotation based on BLAST2GO with minimum E value of e-10 and a maximum of 100 hits.
[b] Number of hits using a maximum of 100 hits as a threshold
[c] Minmum
[d]Reads Per Kilobase per Million

Continued

Rye ID	Sequence Description	#[b] Hits	Min[c]. E-value	Mean Similarity	Tissues	Reads	RPKM[d]
R51	U-box domain-containing protein 35-like	100	2.44E-42	0.5816	Anther	15	17.6
R52	Protein	100	2.79E-43	0.6457	Anther	12	17.6
R53	Two-component response regulator ARR11	4	2.02E-82	0.68	Anther	16	14.5
R54	Cysteine proteinase RD21a-like	11	1.41E-12	0.9336	Anther	18	14.0
R55	F-box domain containing protein	42	7.97E-20	0.549	Anther	12	13.8
R56	Os10g0520400	18	1E-106	0.5867	Anther	11	13.8
R57	protein far1-related sequence 5	25	8.57E-35	0.6664	Anther	13	13.8
R58	UPF0481 protein at3g47200-like	100	3.4E-147	0.6089	Anther	13	13.1
R59	Hypothetical protein	1	2.15E-11	0.81	Anther	10	12.9
R60	Horma domain-containing protein	65	1.01E-57	0.6422	Anther	10	10.5

[a] Annotation based on BLAST2GO with minimum E value of e-10 and a maximum of 100 hits.
[b] Number of hits using a maximum of 100 hits as a threshold
[c] Minmum
[d]Reads Per Kilobase per Million

Plant resistance genes (R genes) have been reported to be eliminated by alloploy ploidization. Genomic analyses in Arabidopsis, cotton and soybean indicated that these genes, especially Nucleotide Binding-LeucineRich Repeat (NB-LRR) genes, were preferentially lost following polyploidyzation (Cannon *et al.,* 2004; Nobuta *et al.,* 2005; Zhang *et al.,* 2011).

Rye silent genes likely to be deleted from triticale genome

To determine if the cause of the suppression of gene expression of rye genes in the triticale background could be gene deletion, a PCR assay was performed using primer sets derived from ten rye genes found to be suppressed in the survey of the 454-expression profile. PCR amplification was carried out on the genomic DNA of two cultivars of triticale (Pika and AC Certa). Control assays were done with two rye cultivars (Prima and Muskateer) and two cultivars of *T. aestivum* (winter and spring cultivars of Anza). Six out of ten rye candidate suppressed genes were found to be absent from the triticale genomes, although they were present in the rye cultivars (Figure 4); wheat cultivars showed no amplification products, as expected, since closely matching sequences were not found in wheat. The rye genes, R9, R11, R15, R16, R32 and R40, were deleted from both triticale cultivars.

RT-PCR analysis was performed on roots and stems collected from the same triticale and rye cultivars to test for the expression of two candidate genes that were detected in the genomic DNA screen. The analysis did not detect any expression from R8 in the cDNA generated from the stems of two-week old triticale plants, although the expression of the same gene was found in

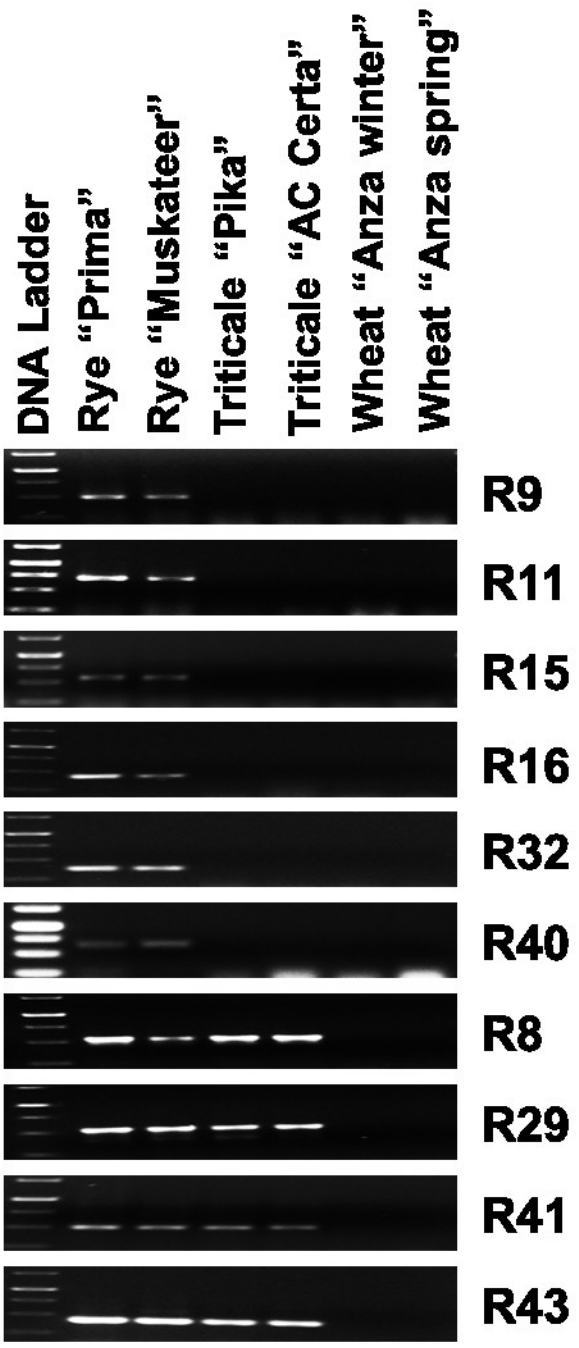


Figure 4. PCR amplification of rye silent genes in rye, triticale and wheat.
A selected set of ten rye reference sequences found to be completely suppressed from all triticale tissues were used to screen by PCR amplification in the genomic DNA of triticale for evidence of gene deletion using genomic DNA of two cultivars of rye, triticale and wheat plants. The molecular size ladder has DNA fragments of molecular sizes of 1500bp, 1000bp, 750bp, 500bp and 250bp.

-

the same rye tissue. However, the expression of R29 was detected at very low levels in triticale and rye roots (Figure 5). The other two silent genes that were detected in the genomic DNA of triticale, R41 and R43, that were expressed in rye anthers were not deleted in triticale anther were not testedby RT-PCR since isolating RNA from anther is technically demanding.

What could be the mechanism behinds genetic alteration of allopolyploids?

The combination of diverged genomes in newly formed allopolyploids can result in dramatic changes in the genome structure and in the transcriptome. These changes occur under extreme selection for the formation of stable fertile hybrids, and changes in genomes formed in allopolyploids likely increase fitness (reviewed in Feldman *et al.,* 2012). Both triticale and wheat have high degrees of plasticity due to their natural capabilities to overcome these dramatic changes (Ma *et al.,* 2004; Feldman *et al.,* 2009). The structural genomic changes might play a vital role in chromosome pairing during meiosis for restoring the full fertility of the plant after the extreme genetic shock facing the new hybrid. Although this process occurs naturally, genome changes do not appear to be random (Ozkan *et al.,* 2001; Kashkush *et al.,* 2002). In the current study, most of the rye silenced genes in triticale apparently have low sequence similarity to genes in the other genomes of triticale. Of the ten rye genes that appeared to be silenced, six were found to be deleted from the rye genome in triticale. It appears that there are molecular mechanisms for self-recognition and that dissimilar sequences are somehow eliminated from the rye genome in triticale. Though numerous deletions were detected in Triticum and triticale hybrids by AFLP analysis, this analysis did not differentiate between regions that had high or low sequence similarity between the two parental species, nor did the analysis distinguish between coding and non-coding regions (Shaked *et al.,* 2001; Ma *et al.,* 2004).

Similarly, the analysis of newly formed triticale (Ma *et al.,* 2004), allotetraploid wheat (Kashkush *et al.,* 2002) and allohexaploid

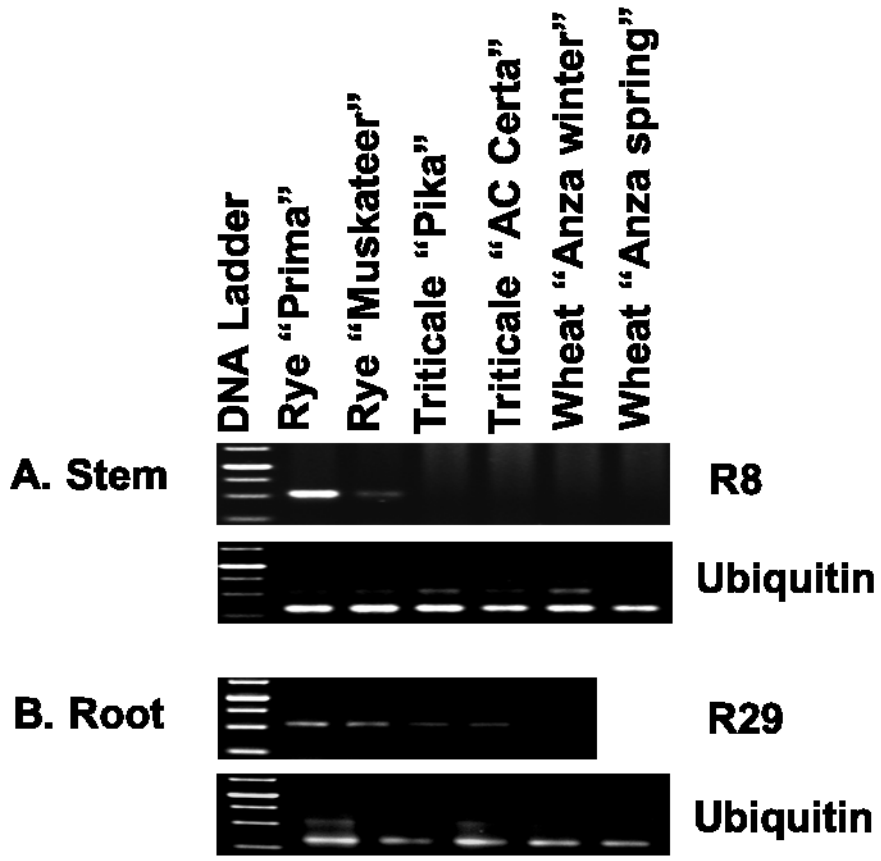


Figure 5. Validation of the suppression of expression of two rye genes in triticale allopolyploid.
Two rye coding sequences found to be silenced in triticale but found to be present in the genomic DNA of triticale were rescreened by RT-PCR on the cDNA generated from RNAs of two-week old rye, triticale and wheat seedling. Ubiquitin gene was used as a control for the expression in all plants. **(A)** RT-PCR revealed that R8 from rye genome was suppressed in triticale stem. **B.** Low expression from R29 from rye genome was detected in triticale root. The ladder has DNA fragments with molecular sizes of 1500bp, 1000bp, 750bp, 500bp and 250bp.

wheat (Feldman *et al.,* 1997; Ozkan *et al.,* 2001) by Restriction Fragment Length Polymorphism analysis revealed coding regions were were deleted from the A, B, D or R genomes, but the probes cross hybridized to several genomes and would not detect sequences in one sub-genome that did not share high sequence similarity with another sub-genome. The analysis of the well characterized hardness locus, (*Ha*) locus, gives insight into the susceptibility of particular regions of the genome for elimination. The locus which regulates seed hardness in wheat, indicated that regions of the genome may be particularly liable to deletion (Chantret*et al.,* 2005; Li *et al.,* 2008). Though selection in agriculture is clearly a strong driving force for the preservation of deletions leading to hard seeds, a comparison of allotetraploid and allohexaploid wheat showed that the alleles arose several times independently. What is particularly revealing is that the breakpoints for deletion were very similar, but not identical; such a pattern indicated a bias or targeting of this region for deletion, and the authors suggested that they may be related to transposon activation and illegitimate recombination (Chantret*et al.,* 2005). Although great efforts have been made to detect the genetic changes and epigenetic modifications subsequent to genome hybridization and doubling, understanding gene regulation mechanisms by merging two or more genomes was not an easy task for plant researchers. These investigations have implicated mechanisms including intergenomic recombination (Gaeta and Pires, 2010), transposon activation (Zhao *et al.,* 1998; Parisod*et al.,* 2009) and double-strand break repair (Wicker *et al.,* 2011). Studies in yeast showed that the presence of unpaired regions of DNA within homologous sequences triggers mismatch repair proteins to correct sequence DNA differences (Kirkpatrick and Petes, 1997; Kearney *et al.,* 2001). This repair system involves DNA strand invasion between the heterologous sequences with deletion of unpaired sequences. The

example of heterologous sequence elimination through deleting unpaired loops was based on pairing between homologous chromosomes. The critical question here: Could this process occur between homeologs? Though homeologous recombination is strongly suppressed in polyploid wheat by the *Ph1* locus, an invasion of the A genome by sequences from the B genome was identified in the tetraploid wheat using genomic *in situ* hybridization (Belyayev *et al.,* 2000; reviewed in Levy and Feldman, 2002). Comai (2000) reported that homeolog paring can lead to chromosomal deletion resulting in the breakdown in the post-replicative mismatch repair system. The excessive increase in the potential for mismatches from strand invasion between homeologous chromosomes could lead to saturation and disfunction of the mismatch repair proteins that normally have roles in blocking homeolog recombination. The deleted genes detected in this analysis of high-throughput sequencing offer important candidate genes for further analysis. The comparison of BAC clones for these genes from rye and triticale would be informative to deeply understand the nature and extent of these deletions, especially if they were relatively small and flanking regions could be identified in clones from both rye and triticale. Further genomic studies in rye and triticale are necessary to detect the recombination motifs at the sites of deletion for deciphering the mechanisms of genome rearrangement and evolution.

4.5. Methods

Rye, triticale and wheat growth condition

Seeds of rye (*Secale cereale*, cv. Muskateer and Prima), triticale, 2n=6x, AABBRR, (Triticosecale, cv. AC Certa and Pika), as well as *T. aestivum* winter and spring Anza cultivars were germinated in 20 cm pots containing equal volumes of peat moss, vermiculite and black earth and, grown under 16 h light and 8h dark at 22°C. After fifteen days, the seedling shoots and the roots of the two cultivars of each species were collected individually, frozen immediately in liquid nitrogen, and stored at -80°C.

Rye silent reference cDNA assemblies in triticale tissues

A rye gene reference set of 23,503 cDNA assemblies with a minimum length of 1000 nt were assembled from rye 454-cDNAs in the laboratory of Dr. André Laroche (Agriculture and Agri-Food Canada, Lethbridge) and was used to study their expression in triticale and rye tissue sets. A total of 8,674,186 cDNAs, 6,674,733 from triticale and 1,999,453 from rye derived from eleven triticale and rye tissue specific 454-cDNA libraries developed in the Canadian Triticale Biorefinery Initiative project of the Functional Genomics & Proteomics of Triticale, (Agriculture and Agri-Food Canada, Lethbridge). Three triticale and rye libraries were generated from stem, root and anther tissues; the other five triticale libraries were created from ovary, pollen, seedling, seed and stigma. Quality control analyses of triticale and rye 454-cDNAs were performed in two consecutive steps: poor quality ends were removed by deleting continuous nucleotides with phred scores less than 15, then internal nucleotides with phred scores less than 20 were masked with N's using the tools available by free browser-based access through Galaxy server from Penn. State University and Emory University (Goecks *et al.,* 2010). The high quality 454-cDNAs

obtained from each triticale and rye tissue were aligned to rye reference assemblies with the BWA-SW algorithm aligner (Li and Durbin, 2009) using the default parameters except, mismatching penalty was set at 10 and the z-best heuristics set to 100. The transcripts uniquely mapped to each rye reference sequence were selected and counted. The expression of each rye contig in the reference assemblies was normalized based on the depth of each library and the length of each rye reference sequence using the reads per kilobase per million reads (RPKM) normalization method. Initially, all rye contigs were compared to the triticale reads to detect rye genes that were not expressed in triticale. For more detailed analysis, more highly expressed rye reference sequences with a minimum level of expression of at least ten transcripts in any rye library and suppressed in all triticale libraries were selected for further analysis.

Identifying most similar *Triticum* and *Aegilops* sequences corresponding to rye silenced genes in triticale tissues

The genes in*T. aestivum,* fromthe A and B genomes, as well as in *T. urartu* and *T. tauschii*, A and D genomes, that were most similar to the rye silenced genes in triticale and a control set of 200 random rye reference sequences were identified in the IWGSC-WSS survey sequence repository (http://urgi.versailles.inra.fr/srs83/displayTool.do?toolName=BlastN), *T. aestivaum* GenBank EST database (Release, May 4, 2012), GenBank *T. urartu* and *T. tauschii* genome scaffolds (GB: AOTI00000000 and GB: AOCO010000000, respectively) through BLASTN search. The most similar A, B and D gene copies in all the databases that have at least 100 nt alignment block overlaps were selected. The percent identities between the most similar A, B and D hits to rye sequences were calculated based on the total length of the alignment blocks of each hit.

Gene ontologies for rye-specific silenced reference sequences

The selected set of genes that were highly expressed in rye and were not found to be expressed in the eight triticale tissues was further characterized by their ontologies. They were compared to GenBank databases using the BLAST2GO workstation (Conesa *et al.,* 2005). Functional annotations were taken by sequence comparison to the GenBank non-redundant protein database using BLASTX with a threshold E-value of $1e^{-10}$.

Screening for rye gene deletion

Ten silent rye genes were selected for further characterization in triticale. Ten pairs of rye gene-specific primers (Additional File 2: Table S3) were employed to screen genomic DNA for deletions using genomic DNA from two triticale cultivars. Both rye and wheat cultivars were used as positive and negative controls for the presence of DNA sequences. The genomic DNAs were extracted from one week old seedlings using a CTAB protocol (Doyleand Doyle, 1987). PCR amplification using genomic DNA was performed with Taq polymerase under the following conditions: 95°C for 4 min., followed by 40 cycles of 30 sec. at 94°C, 40 sec. at a temperature between 45° and 60°C depending on the specific primers used, 1 min. at 72°C; these followed by 12 min. at 72°C.

Rye-specific silenced reference validation using RT-PCR

To validate rye gene silencing from the rye sub-genome in triticale, RT-PCR was performed by amplifying a selected set of rye coding sequences based on the cDNAs. Total RNA was extracted from the roots and shoots of seedlings of rye, triticale and wheat cultivars using TRIzol reagent (Invitrogen) according to manufacturer's instructions, then the RT reaction was performed by Superscript™ First-strand Synthesis System for RT-PCR (Invitrogen). Twenty-five µl of reaction

mixtures; 5 µl 10X RT buffer, 10 µl 25 mM $MgCl_2$, 2.5 µl0.1 M DTT, 1 µl RNAse inhibitor and 4 µl DEPC-treated water; were added to 1 µg RNA/ 50 µM oligo dT primer tubes, incubated at room temperature for 2 min. One µl of enzyme SuperScript^{TM} II RT (50 units/µl) was added to each tube, mixed and incubated at room temperature for 10 min. The reverse transcription was performed at 42^{o}C for 50 min. and terminated at 70^{o}C for 15 min. The same rye primers used for testing gene deletion were employed to validate the suppression of expression in triticale. PCR reactions were carried out using rye, triticale and wheat first strand cDNAs. PCR amplifications with Taq polymerase were performed under the following conditions: 95°C for 2 min., followed by 35 cycles of 30 sec. at 94°C, 40 sec. at 50-60°C, 1 min. at 72°C; these were followed by 12 min. at 72°C (Different temperatures were used for the annealing step in different reactions, depending on the melting temperature of the specific primers).

REFERENCES

Adams KL, Percifield R and Wendel JF (2004). Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid.*Genetics*, 168: 2217–2226

Akunova AR, Matniyazov RT, Liang HQ and Akhunov ED (2010). Homoeolog-specific transcriptional bias in allopolyploid wheat.*BMC Genomic*, 11: 505

Belyayev A, Raskina A, Korol A and Nevo E (2002). Coevolution of A and B genomes in allotetraploid *Triticum* dicoccoides.*Genome,* 43: 1021–1026

Bennett MD and Leitch IJ (2011). Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annal. Bot,* 107: 467–590

Boyko EV, Badaev NS, Maximov NG and Zelenin AV (1988). Regularities of genome formation and organization in cereals: I. DNA quantitative changes in the process of allopolyploidization. *Genetika*, 24: 89–97

Brubaker CL, Paterson AH and Wendel JF (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors.*Genome,* 42: 184–203

Cannon SB, Mitra A, Baumgarten A, Young ND and May G (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*.*BMC Plant Biol,* 4: 10

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Cattolico L, Beckert M, Aubourg S, Weissenbach M, Caboche M, Bernard M, Leroy P and Chalhoubb B (2005). Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell,* 17: 1033–1045

Comai L (2000). Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol Biol,* 43: 387–399

Conesa A, Götz S, García-Gómez J, Terol J, Talón M and Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21: 3674–3676

Doyle JJ and Doyle JL (1987).A rapid DNA isolation procedure for small quantities of fresh leaf tissue.*Phytochemistry,* 19: 11–15

Eilam T, Anikster Y, Millet E, Manisterski J and Feldman M (2008). Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum*.*Genome*, 51: 616–627

Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M (2010). Genome size in diploids, allopolyploids, and autopolyploids of Mediterranean triticeae.*J Botany*, 210doi: 10. 1155/ 2010/341380

Feldman M and Levy AA (2009). Genome evolution in allopolyploid wheat: a revolutionary reprogramming followed by gradual changes. *J Genet & Genomics,* 36: 511–518

Feldman M, Levy AA, Fahima T and Korol A (2012). Genomic asymmetry in allopolyploid plants: wheat as a model. *J Exp Bot,* 14: 5045–5059

Feldman M, Liu B, Segal G, Abbo S, Levy AA and Vega JM (1997). Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics,* 147: 1381–1387

Gaeta RT and Chris Pires J (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet.*New Phytol,* 186: 18–28

Goecks J, Nekrutenko A and Taylor J (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol,* 11:86

Jizeng J., Shancen Z., Xiuying K., Yingrui L., Guangyao Z., Weiming H., Rudi A., Matthias P., Yong T. and Xueyong Z. (2013).*Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, 496: 91–95

Kashkush K, Feldman M and LevyAA (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid.*Genetics,* 160: 1651–1659

Kearney H M, Kirkpatrick DT, Gerton JL and Petes TD (2001). Meiotic recombination involving heterozygous large insertions in *Saccharomyces cerevisiae*: formation and repair of large, unpaired DNA loops. *Genetics,* 158**:** 1457–1476

Kirkpatrick DT, and Petes TD (1997). Repair of DNA loops involves DNA mismatch and nucleotide excision repair proteins. *Nature,* 387**:** 929–931

Koh J, Soltis PS and Soltis DE (2010). Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus (Asteraceae)*. *BMC Genomics,* 11: 97

Lee H-S and Chen ZJ (2001). Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc Natl Acad Sci* (USA) 98: 6753–6758

Levy AA and Feldman M (2002). The impact of polyploidy on grass genome evolution.*Plant Physiol,* 130: 1587–1593

Li H and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754–1760

Li W, Li H, and Gill BS (2008). Recurrent Deletions of Puroindoline Genes at the Grain Hardness Locus in Four Independent Lineages of Polyploid Wheat.*Plant Physiol.,* 146: 200–212.

Ling H-Q, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature, 496: 87–90

Lysak M, Berr A, Pecinka A, Schmidt R, McBreen K and Schubert I (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related *Brassicaceae* species. *Proc Natl Acad Sci* (USA),103: 5224–5229

Ma X-F and Gustafson JP (2008). Allopolyploidization-accommodated genomic sequence changes in triticale. *Ann Bot (Lond),*101: 825–832

Ma X-F, P Fang and JP Gustafson (2004). Polyploidization-induced genome variation in triticale. *Genome,*47: 839–848

Ma X-F and Gustafson JP (2006). Timing and rate of genome variation in triticale following allopolyploidization. *Genome,* 49: 950–958

Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R and Comai L (2005). Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.*, 41: 221–230

Maestra B and Naranjo T (1999). Structural chromosome differentiation between *Triticum timopheevii* and *T. turgidum* and *T. aestivum*.*Theor Appl Genet,* 98: 744–750

Nobuta K, Ashfield T, Kim S and Innes RW (2005). Diversification of non-TIR class NB-LRR genes in relation to whole-genome duplication events in Arabidopsis. *Mol Plant Microbe Interact,* 18: 103–109

Ozkan H, Levy AA and Feldman M (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops–Triticum*) group. *Plant Cell,* 13: 1735–1747

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MAand Ainouche ML (2009). Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina. New Phytologist*, 184*:* 1003–1015

Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien MA and Leitch AR (2013). Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *Plant J*,74: 829–839

Ridha-Farajalla R and Gulick PJ (2007). The alpha-tubulin gene family in wheat (*Triticum aestivum L.*) and differential gene expression during cold acclimation.*Genome*, 50:502–510

Shaked H, Kashkush K, Ozkan H, Feldman M and Levy AA (2001). Sequence eliminations and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant cell,* 13:1749–1759

Song K, Lu P, Tang K and Osborn TC (1995). Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evaluation.*Proc Natl Acad Sci* (USA)92: 7719–7723

Tate JA, Joshi P, Soltis KA, Soltis PS and Soltis DE (2009). On the road to diploidization?Homoeolog loss in independently formed populations of the allopolyploid Tragopogon miscellus (*Asteraceae*).*BMC Plant Biology* , 9: 80

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Šimková H, Kubaláková M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Doležel J, Keller B and Stein N (2011). Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell*, 23: 1706–1719

Zhang X, Feng Y, Cheng H, Tian D, Yang S and Chen JQ (2011). Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication.*Mol Genet Genomics 285:79-90*

Zhao XP, Si Y, Hanson RE, Crane CF, Price JH, Stelly DM, Wendel JF and Paterson AH (1998). Dispersed repetitive DNA spread to new genomes since polyploid formation in cotton.*Genome Res,* 8: 479–492

SUPPLEMENTAL FIGURES

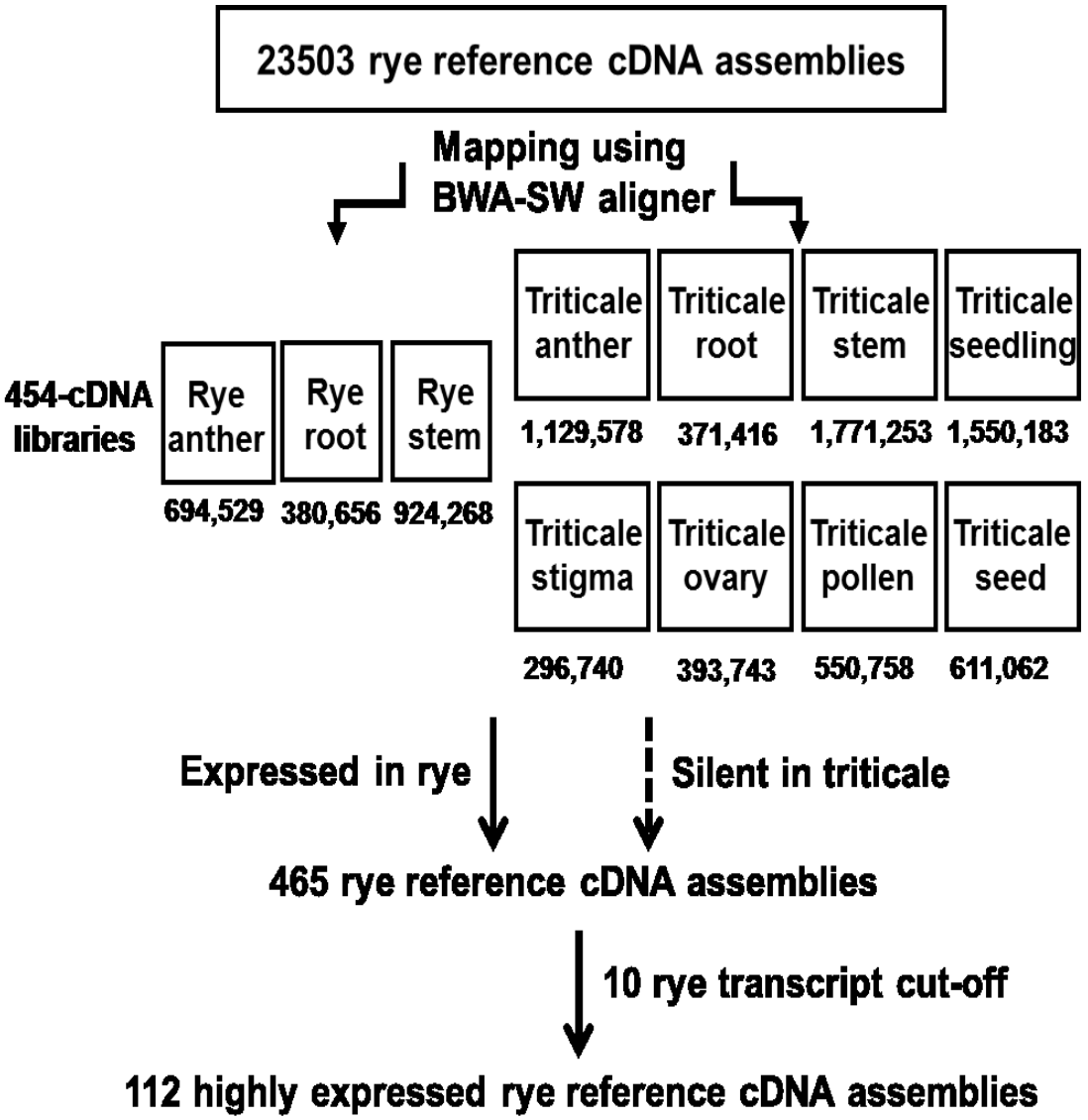


Figure S1. Flowchart of the expression analysis used to detect genes from rye genome silenced in triticale.

RNA-seq workflow was performed to characterize the expression of 23,503 rye reference contig assemblies. Rye and triticale transcripts were aligned to their corresponding rye reference assemblies using BWA-SW algorithm of BWA. About 465 reference sequences were found to be expressed in at least one rye tissue and were not found in any of the eight triticale libraries used in the study. Out of these, hundred-twelve rye reference sequences were found to have at least ten transcripts were aligned to each sequence but to be totally suppressed in all triticale tissues. The depth of each library, the number of transcripts of each library, is written underneath each tissue.

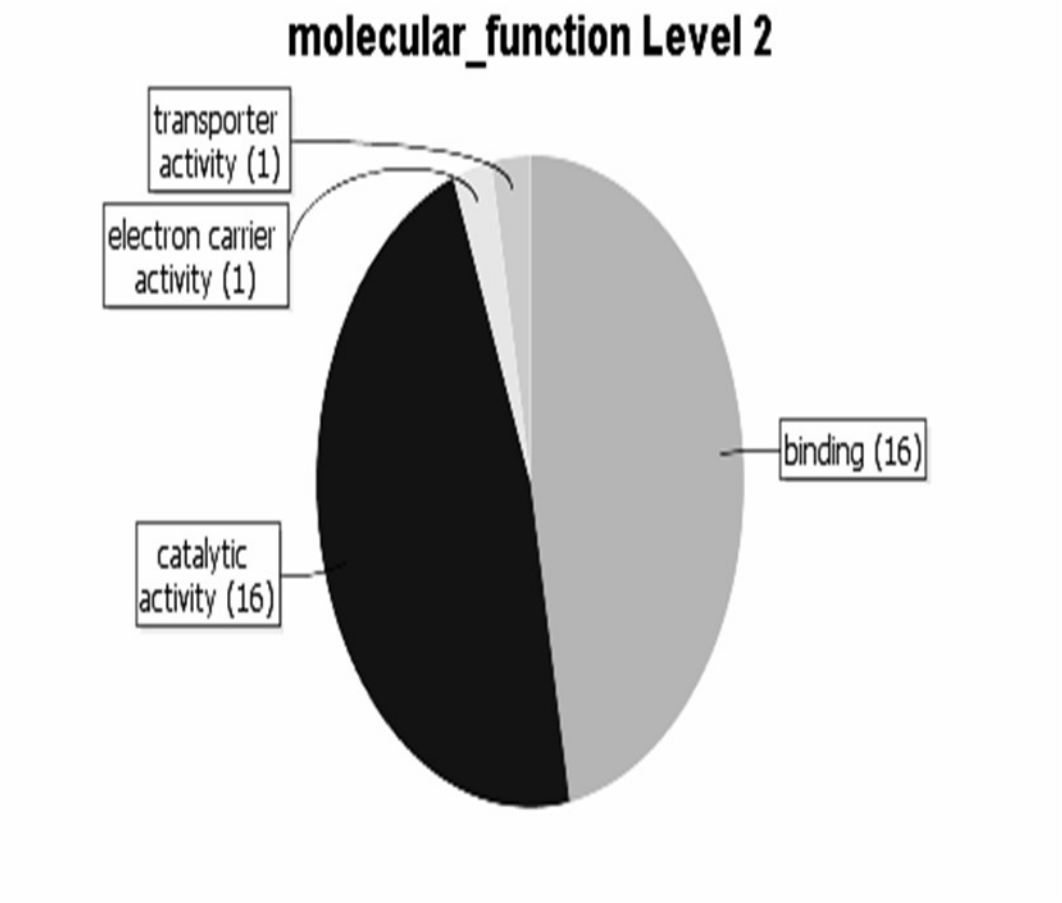


Figure S2. GO-annotation classification of rye silenced genes in triticale.
The annotation of rye genes silenced in triticale was carried out using BLAST2GO. Level 2 category is shown for the molecular function of rye suppressed genes in triticale. Most of GO-annotated genes have catalytic and binding activity.

SUPPLEMENTAL TABLES

Table S1. The percentage identity of rye silenced genes in triticale with their highest scoring hits in the A and B genomes of *T. aestivum* IWGSC database as well as *T. aestivum* GenBank EST database and the most similar rye reference sequences similar to the *T. aestivum* GenBank ESTs that was the highest scoring hits by a rye gene silenced in allohexaploid triticale.

Query		Hit in IWGSC (A genome)			Hit in IWGSC (B genome)			Hit in *Ta*EST database		Query	Hit in Rye Reference sequences	
Rye ID	Rye silenced genes in triticale	Genome A	Contig ID	% Ident	Genome B	Contig ID	% Ident	*Ta*EST ID	% Ident	*Ta*EST	Rye Reference sequences	% Ident
R37	ContigMNoTE9089	A	6005686	94	B	7004512	94	CD881662.1	97	CD881662.1	ContigMNoTE9089	97
R42	ContigMNoTE7032	.	.	.	B	10867584	89	CJ566861.1	95	CJ566861.1	ContigMNoTE7032	95
R24	ContigMNoTE9632	.	.	.	B	10627325	82	GH731363.1	94	GH731363.1	ContigMNoTE11695	98
R54	mergeContigMNoTE8707	.	.	.	.	.	.	BI479046.1	94	BI479046.1	ContigMNoTE6027	96
R68	NoContigMNoTE12664	A	3495600	89	.	.	.	CJ548424.1	93	CJ548424.1	NoContigMNoTE 12664	93
R13	mergeContigMNoTE6789	.	.	.	.	.	.	BE606866.1	93	BE606866.1	mergeContigMNoTE 22075	86
R109	NoContigMNoTE22310	A	3408707	92	B	10623432	91	BU099470.1	92	BU099470.1	NoContigMNoTE 22310	92
R47	ContigMNoTE20990	A	4557653	93	B	6740037	81	CJ812323.1	92	CJ812323.1	ContigMNoTE2139	94
R7	ContigMNoTE27920	.	.	.	B	6736309	77	CJ803068.1	91	CJ803068.1	ContigMNoTE5258	98
R1	ContigMNoTE31637	.	.	.	.	.	.	CJ630249.1	90	CJ630249.1	ContigMNoTE31637	90
R102	ContigMNoTE16232	.	.	.	B	10661817	87	CJ824141.1	89	CJ824141.1	mergeContigMNoTE 4290	96
R20	ContigMNoTE19873	.	.	.	B	3478741	87	CJ582237.1	88	CJ582237.1	ContigMNoTE20140	88
R70	ContigMM588	.	.	.	B	3429689	84	BJ268895.1	87	BJ268895.1	ContigMM588	87
R17	ContigMNoTE19203	.	.	.	B	3000895	79	CA599987.1	86	CA599987.1	ContigMNoTE9977	86
R79	ContigMNoTE19363	.	.	.	B	1372344	84	CD891022.1	86	CD891022.1	ContigMNoTE1439	94
R3	ContigMNoTE446	A	4447452	84	B	10457438	84	CJ711382.1	86	CJ711382.1	ContigMNoTE6952	95
R52	ContigMM200	A	1507015	89	.	.	.	BF484696.1	85	BF484696.1	ContigMNoTE12886	98
R28	NoContigMNoTE6421	A	7130705	85	B	4957931	86	BJ318710.1	85	BJ318710.1	mergeContigMNoTE 10402	87

Continued

Query		Hit in IWGSC (A genome)			Hit in IWGSC (B genome)			Hit in *Ta*EST database		Query	Hit in Rye Reference sequences	
Rye ID	Rye silenced genes in triticale	Genome A	Contig ID	% Ident	Genome B	Contig ID	% Ident	*Ta*EST ID	% Ident	*Ta*EST	Rye Reference sequences	% Ident
R106	ContigMNoTE18823	A	4180399	91	.	.	.	CJ705476.1	84	CJ705476.1	ContigMNoTE413	96
R103	ContigMNoTE17572	A	4180399	95	B	3144835	74	CJ705476.1	84	CJ705476.1	ContigMNoTE413	96
R49	ContigMNoTE13263	A	2697716	76	B	10856664	74	BJ296610.1	84	BJ296610.1	ContigMNoTE19068	98
R78	NoContigMNoTE12633	A	4288213	73	B	10562796	85	CJ843198.1	83	CJ843198.1	NoContigMNoTE 12633	83
R32	ContigMNoTE16436	.	.	.	B	2979005	75	CJ631389.1	81	CJ631389.1	ContigMNoTE16436	81
R31	CL1250Contig1_TC369488	.	.	.	.	.	.	CJ826272.1	81	CJ826272.1	ContigMNoTE16178	95
R67	ContigMNoTE18890	.	.	.	B	901182	83	CJ826299.1	80	CJ826299.1	ContigMNoTE35223	94
R19	ContigMNoTE3513	A	3975375	74	B	3429689	87	CN008026.1	80	CN008026.1	ContigMNoTE8854	89
R30	ContigMNoTE14440	A	4481614	75	B	6684371	76	BJ278960.1	80	BJ278960.1	ContigMNoTE40574	83
R8	ContigMNoTE1541	A	5806460	71	.	.	.	CJ953498.1	77	CJ953498.1	mergeContigMNoTE 15500	92
R34	ContigMNoTE12647	.	.	.	B	2292527	92	CJ667350.1	74	CJ667350.1	ContigMNoTE9697	96

Table S2. Rye silenced genes in triticale that did not have any similar match to GeneBank NR database based on BLAST2GO with minimum E-value of e-10.

Rye ID	Sequence Length	Tissues	Reads	RPKM[a]
R61	1116	stem and anther	11 and 35	12.4 and 55.3
R62	1765	Stem	56	40.1
R63	1403	Stem	28	25.2
R64	1340	Stem	25	23.6
R65	1329	Stem	22	20.9
R66	2574	Stem	42	20.6
R67	1090	Stem	17	19.7
R68	1449	Stem	22	19.2
R69	2349	Stem	31	16.7
R70	2069	Stem	26	15.9
R71	1771	Stem	20	14.3
R72	1370	Stem	15	13.8
R73	1551	Stem	16	13.0
R74	1004	Stem	10	12.6
R75	1635	Stem	16	12.4
R76	1350	Stem	13	12.2
R77	1678	Stem	16	12.0
R78	1452	Stem	13	11.3
R79	1723	Stem	15	11.0
R80	1171	Stem	10	10.8
R81	1194	Stem	10	10.6
R82	1480	Stem	12	10.2
R83	1249	Stem	10	10.1
R84	1517	Stem	12	10.0
R85	1947	Stem	15	9.7
R86	1473	Stem	11	9.4
R87	1349	Stem	10	9.4
R88	1445	Stem	10	8.7
R89	1601	Stem	11	8.7
R90	2045	Stem	12	7.4
R91	1689	stem and root	10 and 18	7.5 and 57.9
R92	3213	Anther	174	270.6
R93	2723	Anther	165	266.3
R94	2400	Anther	334	245.5
R95	2104	Anther	106	164.4
R96	1873	Anther	144	93.3
R97	1817	Anther	66	75.5

[a]Reads Per Kilobase per Million

Continued

Rye ID	Sequence Length	Tissues	Reads	RPKM[a]
R98	1541	Anther	73	70.9
R99	1532	Anther	36	54.9
R100	1487	Anther	31	54.5
R101	1333	Anther	60	50.3
R102	1227	Anther	90	49.4
R103	1157	Anther	30	48.9
R104	1137	Anther	27	38.8
R105	1134	Anther	36	33.9
R106	1093	Anther	16	27.7
R107	1083	anther	22	26.1
R108	1078	anther	15	25.9
R109	1043	anther	15	25.4
R110	1022	anther	15	19.8
R111	1017	anther	15	17.3
R112	1004	anther	10	16.4

[a]Reads Per Kilobase per Million

Table S3. Primers used to study the deletion of rye genes from rye genome in triticale.

Rye ID	Tissue	Forward Primer	Reverse Primer
R8	stem	5′ AGAATGGATGCTCTTACTTGTCAACC 3′	5′ TTACAAACCCAGAATTAATGTGATCTC 3′
R9	stem	5′ CTCATGTGATGTGGTGTCTGCCAACCA 3′	5′ TAGTCTCCAAACGTCACTACCCACCCA 3′
R11	stem	5′ CAACATGGAAGATGCTGTGGACT 3′	5′ TTGTTCGGTGCTTCGATCTTGTC 3′
R15	stem	5′ ATCACTGTCGAGCCGCCGGTAT 3′	5′ TAGTACTCCTTAATTACTCTGAATCGTGGA 3′
R16	stem	5′ CGTACGCAACACAAAGCATATGAA 3′	5′ GACTAGCTTGTCATGGTTCAAATCA 3′
R29	root	5′ TCCTATGTGCTGGCGGACGAG 3′	5′ CGTCGGCACACGTAACACAGAAATTG 3′
R32	root	5′ CAGCAAGACCCTCAACGCGCT 3′	5′ TGGGATGGATGAGTAGAGTAGGCAG 3′
R40	anther	5′ CGCGCTTGGTTATGCGACTT 3′	5′ GGCCATATAACAAGTCTGCTTTAAGTTT 3′
R41	anther	5′ TTGGTTGGACACTGGCACACGAAGC 3′	5′ GAGTACGTACAGGAGCATGTCAAGCCAA 3′
R43	anther	5′ CGGGACGGATGGCTCTGAATAT 3′	5′ CGTTCATGAAGGAACTACGAACATGAT 3′