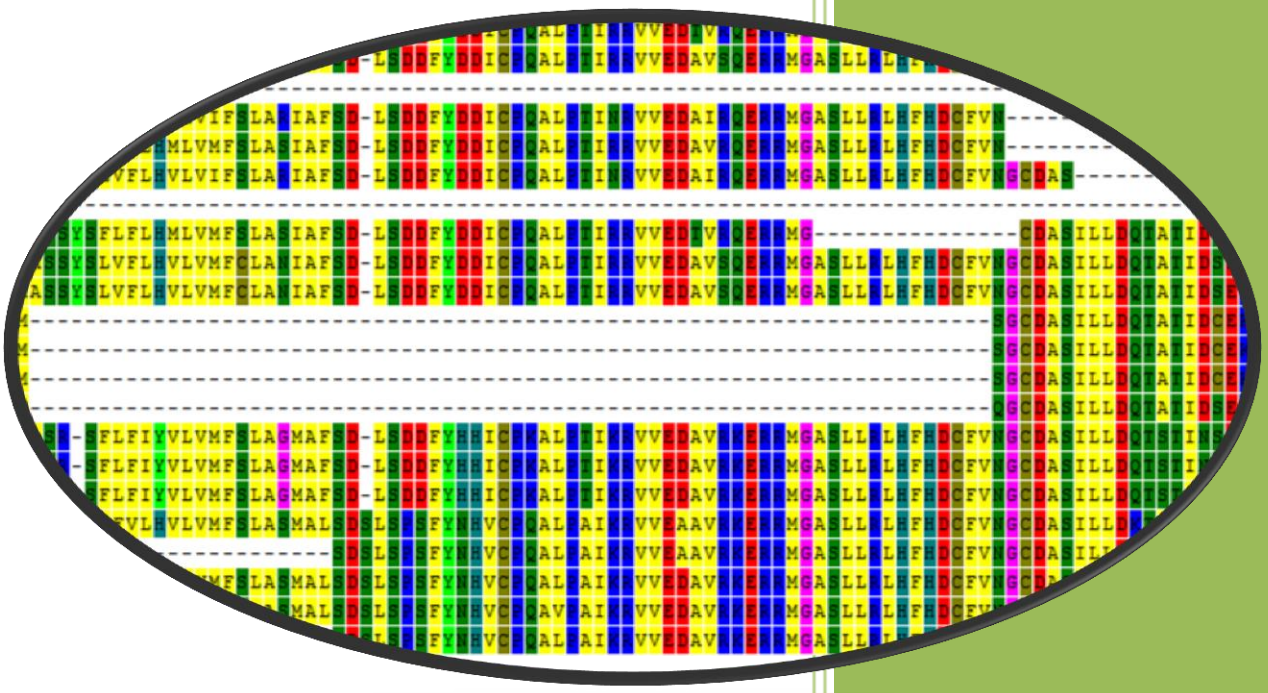
2015

Identifying the gene involved in the shape of potato tubers



Jean Custers
Laboratory of Plant Breeding
Wageningen University

Identifying the gene involved in the shape of potato tubers

Jean Custers
920330-166-080

Master Thesis (MSc) in Plant Breeding
PBR 80436

Supervised by:
Dr. Christian Bachem
Dr. Ir. Herman van Eck
Johan Willemsen MSc
Dr. Jose Abelenda

March 2015,
Laboratory of Plant Breeding, Wageningen University and Research Institute,
Wageningen, The Netherlands.

Abstract

Potato tuber shape is a morphological trait which displays continuous variation. Tuber shape ranges from compressed, round, oval and long. A large proportion of the phenotypic variation in tuber shape is determined by the *Ro*-locus. The *Ro*-locus contains 15 annotated candidate and 8 of these candidate genes are peroxidases. In this study it is tried to identify the gene influencing tuber shape by gene expression experiments and by apoplastic peroxidase activity measurements. Prior to the identification of this gene, several *in silico* experiments have been performed in order to get a better understanding about the *Ro*-locus at a functional genomic level. Also the ability to study tuber shape *in vitro* has been studied in this thesis. Small indications for a general enhancement of expression in the *Ro*-locus have been observed in round genotypes. Also a trend of higher apoplastic peroxidase activity has been found in the round genotypes. After the *in silico* experiments it became clear that the genome assembly of the reference sequence is of a relatively low quality. Therefore, it is not exactly known how many peroxidases are located in the *Ro*-locus. Round and long genotypes treated with tuber inducing conditions showed differences in their time to tuber formation. No differences in tuber shape have been observed *in vitro.* No hard conclusions could be drawn about the ability to study tuber shape *in vitro*, because the phenotypes of the genotypes used in the *in vitro* tuber formation experiments did not correlate with earlier acquired phenotypic data.

Key words: Potato, *Solanum tuberosum*, *Ro*-locus, Tuber shape, Peroxidase, *In vitro* tuber formation, Gene expression, Functional genomics

Table of Contents

1. Introduction

1.1 Potato

Potato (*Solanum tuberosum*) is one of the most important food crops in the world. The production in 2013 was estimated at 368 million tonnes, which makes potato the fourth highest produced directly consumable food crop in the world (FAOSTAT, 2015). The family of *Solonaceae* contains many economically important crops like tomato, pepper and tobacco (PGSC, 2011). Cultivated *Solanum tuberosum* has its centre of origin in the Andean region, where genetic diversity for this species is the highest and many wild relatives have been discovered (PGSC, 2011). Most potato cultivars are outbreeding autotetraploid (2n = 4x = 48) cultivars with highly heterozygous genomes, which suffer from acute inbreeding depression (PGSC, 2011). Taking into account the growing importance of the crop in countries like China and India (FAOSTAT, 2015) and the high nutritional value of the crop (PGSC, 2011), it is almost certain potato will play an important role in guaranteeing food security worldwide.

Potatoes are consumed freshly and are processed in many different ways. Especially for the processing of potatoes the shape of the tubers is very important. In general cultivars producing long potatoes are preferred to produce French fries, whereas potato cultivars producing round tubers are preferred to produce crisps (Van Eck *et al.*, 1994). The genetic basis of a morphological trait such as tuber shape is less investigated, because in the past research mostly focussed on the genetic basis of agronomical traits (van Eck, 2007). Tuber shape seems a continuous trait due to the range of different tuber shapes like round, oval and long which are described for cultivated potato species. For wild potato relatives even a broader spectrum of tuber shapes is observed (The European Cultivated Potato Database, 2015). An example of this huge variation is shown in Figure 1. The loss of the large tuber shape diversity as is observed *Solanum tuberosum* species could be caused by a domestication event, where people selected for uniform shape to satisfy the demands of both consumers and processors. However this loss of genetic diversity for tuber shape could also be caused by selection on another trait, which is genetically linked to tuber shape (Uitdewilligen *et al.,* 2012).

Figure 1: Illustration of the huge variation in tuber shape and tuber colour (USDA, 2015)

1.2 Development of a Potato Tuber

In order to understand and study the trait tuber shape, it is necessary to understand the development of a potato tuber. Potato tubers are formed from thickened stolons, which will develop into the storage organs. Stolons are in fact stems, bearing small leaves and axial buds (Ewing and Struik, 1992). Therefore Potato belongs to the class of plants bearing stem tubers, just like yams and cyclamen.

Several environmental factors influence the induction of tuberization. A well-studied environmental factor influencing tuberization is photoperiod. When a potato plant is exposed to long day conditions, stolons often tend to grow upward to form new shoots. However when a plant is exposed to short day conditions, the stolon will cease elongation and start to swell below-ground (Jackson, 1999). Other factors known to promote induction of tuberization are low N_2-availability, low temperatures and high sucrose levels (Jackson, 1999).

Under high tuber-inducing conditions or on diseased/injured plants it is even possible for plants to develop tubers on above-ground parts of the plant (Jackson, 1999). Most often plants above-ground formed tubers will develop from axillary nodes, but tubers can also form from flowers under very specific circumstances (Ewing and Struik, 1992). By combining the knowledge of the specific tuber-inducing circumstances and the knowledge about tuber-induction on all different kinds of stem-derived meristems, it has been made possible to induce tuber growth *in vitro.* For *in vitro* tuber-induction also some knowledge about the hormonal-regulation of tuber induction is required, but this will be explained later this chapter.

The development of a tuber can be divided into the stages tuber initiation, tuber growth and tuber filling, before a tuber becomes dormant. Kloosterman et al., (2005) also shows a differential gene-expression in the different developmental stages of the tubers. Somewhere during the development of the tuber, the genes influencing tuber shape are expressed. When this expression really takes place is not known yet.

1.3 *In Vitro* Tuber induction

Tuber induction is regulated by both biotic and abiotic factors. By optimizing both abiotic and biotic factors it is possible to induce tuber-formation *in vitro.* However what the optimal conditions are to induce tuber formation depends strongly on the genotypes which have to be induced. Tuber induction is regulated by several hormones, with all their specific effects. Generally, hormones are divided in a group of stimulating hormones and a group of inhibiting hormones.

A well-studied hormone which is inhibiting tuber induction is Gibberellin (GA). GA's are known to enhance stem organ elongation, therefore GA's are also known to enhance potato stolon elongation (Vreugdenhil and Struik, 1989). Factors stimulating stolon growth are

unfavourable for induction of tuber formation because the induction of tuber formation is related to ceasing stolon growth (Aksenova et al., 2012).

A well-known hormone which stimulates tuber formation is Cytokinin (Ck). Cks are known to stimulate tuber formation in many tuberiferious plant species (Aksenova et al., 2012). When a stolon starts to from a tuber, endogenous Ck concentrations increased substantially in the stolons (Obata-Sasamoto and Suzuki, 1979). A higher rate of tuber formation has been established, by treating potato explants with Cks in the first 3-4 days. This indicates that Cks stimulates tuber formation early in the development (Aksenova et al., 2012).

By growing explants under the right conditions like High Ck concentrations and high sucrose concentrations, Aksenova et al. (2000) managed to induce potato explants to form tubers *in vitro*. Depending on the genotypes, tuber inducing conditions have to be fine-tuned. Unfortunately, it is not known yet whether the trait tuber shape can be studied *in vitro.*

1.4 Loci Involved in Tuber Shape Regulation (*Ro*-locus)

Tuber shape is commonly regarded as a continuous trait, however different studies on diploid potatoes concluded a monogenic inheritance of tuber shape (de Jongh, 1972)(Taylor, 1978)(Masson, 1985). The locus describing this monogenic inheritance of tuber shape is called the *Ro*-locus. Round tubers are dominant over long tubers, segregating according to Mendel's laws (Masson, 1985). The broad-sense heritability of the *Ro*-locus was estimated on h^2=0.8 and the *Ro*-locus explained 75% of the genetic variation found for tuber shape(Van Eck *et al.*, 1994). Twenty years later both Willemsen (2014) and D'Hoop et al. (2014) concluded even a higher broad-sense heritability of h^2=0.9. This difference could be related to the fact that more replicates were used by Willemsen (2014), which resulted into a decrease of variance between samples. Both Willemsen (2014) and Van Eck et al. (1994) concluded the presence of several alleles with different effects on tuber shape for the *Ro*-locus.

Van Eck *et al*. (1994) managed to locate the *Ro*-locus on chromosome 10 by RFLP linkage mapping in a CxE population. Vos *et al.* (data unpublished) showed by an association study significant associations for tuber shape with super-scaffold DMB385, locating the *Ro*-locus in a region of approximately 3.1 Mb. Witteveen (2013) and Willemsen (2014) managed to fine-map the locus to an interval of 280 kb located on chromosome 10 on DMB546 by HRM-analysis. The locus still contains 15 annotated candidate genes, which can be divided in 4 categories: genes of unknown function, non-specific lipid-transfer protein, aspartate aminotransferase and peroxidases. Willemsen (2014) is suggesting that peroxidases are the most logical candidates to be influencing tuber shape. For the other three categories no evidence of involvement in tuber shape regulation can be found. The *Ro*-locus is mainly build up from highly repetitive sequences like transposons and peroxidases.

Another locus influencing tuber shape is described by Prashar *et al.* (2014). This locus is located on chromosome 2 of *Solanum tuberosum*. This locus may be the same locus as described by Sliwka *et al.* (2008). A QTL study in CxE did only reveal a significant association on chromosome 10, which was confirmed in different crosses (Eck, 1994). For this study a CxE population will be used, because this population is segregating for tuber shape and has a relatively similar genetic background. The focus will be on the QTL on chromosome 10 in this study.

1.5 Peroxidases

Peroxidases can be considered as bifunctional enzymes which are able to oxidize various substrates in the presence of H_2O_2 via the peroxidative cycle, but are also able to produce reactive oxygen species (Passardi *et al*., 2004). Peroxidases are involved in several physiological processes like lignin formation, auxin catabolism, defence against pathogens, cell elongation and cross-linking of cell wall components (Passardi *et al*., 2004). In the context of tuber shape cell elongation and cross-linking of cell wall components are interesting because these two processes can determine the shape of cells. The role of peroxidases in cell wall elongation is shown in Figure 2. Lin *et al.* (2014) is even suggesting that peroxidases play a role in turnip shape formation. Peroxidases could have a similar role in determining shape of an organ in the tubers of potato plants.

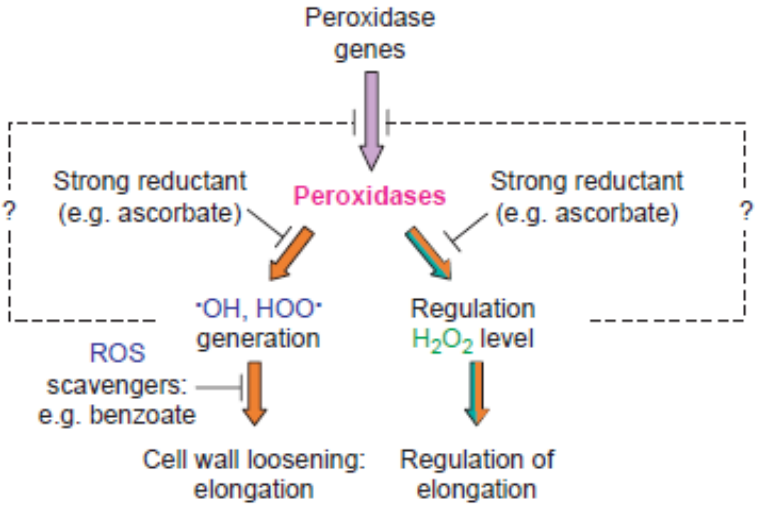


Figure 2: Strongly simplified illustration of the roles of peroxidases in cell wall elongation. (Passardi *et al.*, 2004)

In general three classes of peroxidases can be distinguished, which influence different metabolic processes: 1) Class I peroxidases are intracellular peroxidases like ascorbate peroxidases, which is an hydrogen peroxide scavaging enzyme that protects cells from damage caused H_2O_2 (Asada, 1992) 2) Class II consists of fungal secretory peroxidases like like lignin peroxidases, which is involved in the degradation of lignin. 3) Class III plant secretory peroxidases, which can be involved in cell wall biosynthesis(Welinder 1992). In this study we will mainly focus on class I and class III peroxidases, because class II peroxidases are not found in plants.

Peroxidases are characterized by 4 peroxidase domains (Dellanoy *et al*., 2003)(Passardi et al. 2004a). The first domain is characterized by the amino-acid sequence FHDC and is responsible for distal binding of heme. This is also the catalytic domain (Dellanoy *et al*., 2003)(Passardi et al. 2004a). The second domain is characterized by the amino-acid

sequence GAHT, this domain is responsible for the proximal binding of heme. The last two domains are characterized by the amino-acid sequences VSCAD and GEIR, but the function of these domains is not known yet (Dellanoy *et al*., 2003)(Passardi et al. 2004a).

In *Solanum tuberosum* hardly any research is done to the role of peroxidases. However, in tomatoes (*Solanum lycopersicum*), a close relative of *Solanum tuberosum*, several studies are done to the role of peroxidases in fruit development. It is suggested that peroxidases are involved in the cessation of growth, because a higher activity of peroxidases was observed just before the cessation of growth of the tomato fruit (Thompson *et al.*, 1998). No relation with the shape of tomato fruits has been proposed.

1.6 Goals

The main goal of this study is to characterize the genes/category of genes which are influencing potato tuber shape. This is done by molecular techniques like qPCR, but also biochemical analysis like enzyme activity measurements are used. Measurements are done on dormant tubers which originate from a CxE cross. It is also tried to get a better overview of the highly repetitive region where the *Ro*-locus is located on the basis of sequence-data. This is done by taking a close view at the reference genome of potato cultivar DM *in silico.* The main focus of this study is to get a better understanding about the effects of the different peroxidases on the shape of tubers. The other candidate genes located in the *Ro*-locus will not be neglected in this study.

2. Material & Methods

2.1 Searching for unannotated peroxidases

In order to find unannotated peroxidases transcripts of all the annotated peroxidases shown in Table 1 were aligned using BLAST to the 280 kb long QTL (Altschul *et al.* 1990)(NCBI, 2015). Transcripts used for the blast were extracted from the potato genome reference sequence V4.03 (Sharma *et al.,* 2013)(Hirsch *et al.,* 2014). Default settings were used for this blast alignment. Blast hits per transcript were processed to .gff files and analysed in IGV_2.3.34. Regions were peroxidase transcripts had a blast hit were analysed with the gene prediction program Fgenesh, selecting *Solanum lycopersicum* as a model-species(Solovyev *et al*. 2006). When a gene was predicted, this gene was analysed with the NCBI-domain finder in order to find specific peroxidase domains (Marchler-Bauer *et al*., 2015). When these domains were found, the gene was considered as a newly found peroxidase.

Table 1: All annotated candidate genes, their description and RNA-seq data located within the 280 kb region (between marker Asp6678 and Per20801) (Willemsen, 2014)

Locus	Description	RNA-seq support
PGSC0003DMG400006678	Aspartate aminotransferase	✓
PGSC0003DMG400006679	Peroxidase 2	✓
PGSC0003DMG400006680	Peroxidase 1	✓
PGSC0003DMG400006681	Cationic peroxidase 1	✓
PGSC0003DMG400020800	Cationic peroxidase 1	✓
PGSC0003DMG400045482	Conserved gene of unknown function	×
PGSC0003DMG400040544	Gene of unknown function	×
PGSC0003DMG400020799	Cationic peroxidase 1	✓
PGSC0003DMG400040954	Non-specific lipid-transfer protein 2	✓
PGSC0003DMG400020798	Cationic peroxidase 1	✓
PGSC0003DMG400035649	Gene of unknown function	×
PGSC0003DMG400020797	Gene of unknown function	✓
PGSC0003DMG400039458	Polyprotein	×
PGSC0003DMG400020795	Cationic peroxidase 1	✓
PGSC0003DMG400020801	Peroxidase 2	✓

2.2 *In silico* localization of the peroxidases

To gather more information where to extract peroxidases for the peroxidase activity measurements, *in silico* predictions of the subcellular location of the peroxidase proteins have been made. Genomic sequences from all annotated peroxidases located in the QTL shown in Table 1 were extracted from the potato genome reference sequence V4.03(Sharma *et al.,* 2013) (Hirsch *et al.,* 2014). From these sequences, genes were predicted by the gene prediction program Fgenesh, selecting *Solanum lycopersicum* as a model species. New gene predictions were made to compare the transcripts predicted by Fgenesh and the transcripts

predicted by "PGSC Representative Gene Models" (Sharma *et al.,* 2013) (Hirsch *et al.,* 2014)(Solovyev *et al*. 2006). The gene predictions by Fgenesh are shown in Appendix 6.4 Both transcript predictions were analysed by TargetP 1.1, which predicts the subcellular location of eukaryotic proteins (Emanuelsson *et al*., 2007).

2.3 Peroxidase domain study

To investigate the functionality of the peroxidase located in the Ro-locus, peroxidase domains of the different transcripts have been studied. Both the transcripts predicted by Fgenesh and PGSC were translated to amino acid sequences in MEGA 6.0 (Sharma *et al.,* 2013) (Hirsch *et al.,* 2014)(Tamura *et al*., 2013)(Solovyev *et al*. 2006). All amino acid sequences of the different transcripts were aligned using ClustalW, which is available in the MEGA 6.0 software(Larkin *et al*. 2007). After aligning of the amino acid sequences in ClustalW further improvement on the alignments was done manually. The alignment of all peroxidase proteins is shown in Appendix 6.6. By aligning the amino acid sequences we were able to detect the conserved domains of the peroxidases. These domains were compared with the functional domains described by Dellanoy *et al*. (2003) and Passardi *et al*. (2004a).

2.4 Peroxidase localization by tissue-prints

A preliminary test was done to demonstrate the localization of peroxidases in potato tubers. Both round and elongated tubers from the CxE population, were cut in two parts longitudinally. To localize peroxidase activity a tissue-print was tried according to the following protocol. One half of every tuber was pushed on a piece of nitrocellulose membrane. This nitrocellulose membrane (BioRad 45um thick) was soaked in a 1:1 (v/v) mixture of luminol/enhancer and stable peroxide buffer coming from the kit "SuperSignal West Femto Chemiluminescent Substrate" (Thermo Scientific). This soaked membrane was directly placed under a camera called the G:BOX (Syngene) to take a picture. The lens was opened for 2 minutes in order to capture enough light emitted by the reaction. Pictures were analysed using the software Genesnap.

2.5 Primer development for qPCR

Primer design

qPCR primers were designed for all genes having RNA-seq support located in the QTL region using MEGA 6.0 and Primer3Plus software(Tamura *et al*., 2013)(bioinformatics wur). Both the transcripts predicted by Fgenesh and transcript obtained from the PGSC Genome browser(*Solanum tuberosum* group Phureja) were compared and aligned in MEGA 6.0 per gene(Tamuru *et al.* 2013).

Table 2: PCR protocol used to test primers on a developmental series of *S. tuberosum* RH cDNA

Temperature (°C)	Time	Cycles
94	5'	
94 60 72	30" 30" 30"	40x
72	7'	
10	infinite	

Primers were designed in transcript specific regions of the consensus sequence of both transcripts. 3'UTR of the transcripts were most specific for the transcripts, so most of the primers are located in the 3'UTR. All non-peroxidase candidate primer sets were aligned using BLAST against the PGSC *Solanum tuberosum* group Phureja DM1-3 Transcripts (v3.4) using the blastn nucleotide search, to test the specificity of the primers *in silico* (Altschul *et al.* 1990)(NCBI, 2015). All peroxidase primer sets were tested for specificity *in silico* by aligning the primers, specific for one transcript, with all the other peroxidase transcripts located in the QTL region using CLustalW software(Larkin *et al*. 2007).

Primer testing

Testing of the primers was done by PCR on a developmental series of *Solanum tuberosum* RH89-039-16 (RH), cDNA. How this cDNA was obtained will be explained in chapter 2.6. An GeneAmp® PCR-system 2700 (Applied Biosystems) was used following the PCR protocol shown in Table 2. The reaction mix consisted of 5 ul IQ™ SYBR® Green Supermix, 0.2 ul Primer Forward (10 uM), 0.2 ul Primer Reverse (10 uM), 3.6 ul Milli-Q and 1 ul cDNA (+/- 50 ng/ul).

PCR-products were stained with loading dye and separated by electrophoresis on a 2% agarose using 0.5 x TBE buffer for 30 minutes at 100V. A 1 kb+ ladder was used as a reference. Gels were photographed using an OCTOPUS Fll. Images were analysed to identify the specificity of the primers. When a primer set did not amplify, the whole procedure was repeated to design a new primer set for that particular transcript. When two different primer sets did not work, the gene was excluded for qPCR analysis. A list of all primers which are used can be found in Appendix 6.1

Table 3: PCR protocol used to test primers on *S. tuberosum* CxE genomic DNA

Temperature (°C)	Time	Cycles
94	3'	
94 55 72	30" 30" 1'	40x
10	infinite	

Because of the genomic differences between RH and progeny of a CxE population, the primer sets were also tested by PCR on genomic DNA of S. tuberosum CxE. Previously isolated DNA for both round and long tubers was used to test primer sets. The reaction mixtures consisted of 1ul 10x Dreamtaq PCR buffer, 0.4 ul dNTP's (5mM), 0.4 ul Primer Forward (10mM), 0.4 ul Primer Reverse (10mM), 0.1 ul Dreamtaq-polymerase, 6,7 ul Milli-Q and 1 ul DNA (+/- 50 ng/ul). The PCR protocol which was used is shown in Table 3.

Resulting PCR-products were stained with loading dye and separated by electrophoresis on a 2% agarose using 0.5 x TBE buffer for 30 minutes at 100V. A 1 kb+ ladder was used as a reference. Gels were photographed using an OCTOPUS Fll.

2.6 qPCR on developmental series

To investigate the expression of the genes located in the *Ro*-locus during the development of a tuber, qPCR experiments have been performed on a developmental series of *Solanum tuberosum* RH. The results of these qPCR experiments have been compared with RNA-seq data generated by PGSC (2011).

RNA quality test

RNA samples were provided by Marion Oortwijn. This RNA was isolated from *Solanum tuberosum* RH tubers and stolons. The tubers/stolons were sampled at three different stages: stolons, tubers of 1 cm and tubers bigger than 1 cm. Concentration (ng/ul) and quality (ratios 280/260 and 260/230) of these RNA samples were assessed by measuring 1 ul of sample using the NanoDrop™ 1000 Spectrophotometer(Thermo Scientific). Samples which had enough volume were dyed with gelrad and were loaded on a 2% agarose gel using 0.5 x TBE buffer and separated by gel electrophoresis for 30 minutes at 100V. Trays were cleaned and filled with fresh 0.5 x TBE buffer in order to reduce the breakdown of RNA. After electrophoresis the gels were photographed using an OCTOPUS Fll and images were analysed. RNA samples were classified as intact when two clear bands were visible.

DNase digestion

DNase digestion was performed using the kit "DNase l" (Invitrogen). The final concentrations of the reaction wells were 50 ng/ul RNA, 1X/ul DNase l Reaction Buffer and 0.1 U/ul DNase l, Amp Grade in a total reaction volume of 10 ul. These reaction mixtures were incubated for 15 minutes at room temperature. After incubation, DNase l was inactivated with 1 ul of 25 mM EDTA. Reaction mixtures were heated for 10 minutes at 65°C. After DNase digestion samples were ready for cDNA synthesis.

cDNA synthesis

The iScript™ cDNA Synthesis Kit (Bio-Rad) was used to synthesize cDNA. 4 ul of 5x iScript reaction mix, 1 ul of iScript reverse transcriptase and 4 ul of Milli-Q were directly added to the DNase digestion mixtures. The new reaction mixtures were incubated in a GeneAmp® PCR-system 2700 (Applied Biosytems) following the PCR protocol shown in table 4. Prior to qPCR cDNA was diluted 10x

Temperature (°C)	Time
25	5'
42	30'
85	5'
4	5'
85	5'
10	Infinite

Table 4: PCR protocol used for cDNA synthesis

qPCR

The reaction mixtures used for the qPCR consisted of 5 ul IQ™ SYBR® Green Supermix, 0.2 ul Primer Forward (10 uM), 0.2 ul Primer Reverse (10 uM), 3.6 ul Milli-Q and 1 ul cDNA(+/- 50 ng/ul).The PCR protocol which has been used is shown in Table 5. Also melting curves of the

PCR-products have been performed in order to check the specificity of the primers. The PCR protocol used for the melting curves is also shown in Table 5. A BioRad CFX 96 has been used to perform the qPCR.

Results were analysed in Microsoft Excel. As a reference gene NAC has been used, because Bourke (2014) states NAC is very stable in its expression. Using Ct-values and the The 2–ΔΔCT (Livak) Method, relative expression levels were determined. Statistical significance was not tested, due to the lack of biological replicates. Melting curves have been analysed in CFX manager

To compare the expression profiles determined by qPCR with RNA-seq data, FKPM values for *Solanum tuberosum* RH stolons, young tubers and tubers mature tubers were extracted from the Supplementary Table 19B of the paper written by PGSC (2011). These values were visualised in Microsoft Excel.

Table 5: PCR protocol used for qPCR. In green the PCR protocol is shown which is used for the melting curves.

Temperature (°C)	Time	Cycles
95	3’	
95 60	15” 1”	39x
95	10”	
65	5”	ΔT+0.5°C/Cycle 60x

2.7 qPCR on round and long tubers

To get more information about the differences in expression of the genes located in the *Ro*-locus, qPCR experiments have been performed. Prior to the qPCR experiments RNA had to be isolated and cDNA had to be synthesized.

RNA isolations

Two different protocols were used to isolate RNA, RNA samples which had the highest quality on gel were used for cDNA synthesis and qPCR. In both RNA isolation 4 round and 4 long tubers were sampled. The tubers which were sampled, were dormant tubers which originated from a cross between C x E. For both RNA extractions three tubers per genotype were pooled and four biological replicates per tuber shape class were used.

RNA isolation (Z6-extraction), RNA quality test (spectrophotometer), DNase digestion and RNA cleanup

For the first RNA isolation, a Z6-extraction buffer (8 m guanidinium-HCl, 20 mM MES, 20 mM EDTA, 50 mM β-mercaptoethanol, pH 7.5; Logemann et al. 1987) was used in combination with the “RNeasy Minikit”(Qiagen). Plant tissue was frozen using liquid nitrogen and grinded

using pestle and mortar. The samples were then thawed using the Z6-extraction buffer and placed on ice. The samples were centrifuged at 15 000 g for 20 min at 4°C. The cleared supernatant was transferred to new Eppendorf tubes. From this cleared supernatant nucleic acids were precipitated by mixing with 0.1 vol 3 m sodium acetate pH 5 (DEPC-treated) and 0.8 vol isopropanol. After the precipitation, samples were incubated on ice for 5 minutes and centrifuged at 15 000 g for 20 min at 4°C. The supernatant was poured off and the pellet was washed with 80% EtOH. When the pellets were dry again, these pellets were treated with the RNeasy Mini kit as if they were normal cells. The protocol which goes along with this RNeasy Mini kit (Qiagen) can be found in Appendix 6.2

Quality and concentration of the RNA samples (1 ul) were measured using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific), however quality parameters (ratios 280/260 and 260/230) were too low for further use of this RNA. RNA had to be treated with DNase and cleaned up for further use. DNase digestion was again performed using "DNase l" (Invitrogen) kit. This protocol is already described in chapter 2.6. After DNase digestion, RNA was cleaned up using the "RNeasy Minikit" (Qiagen) in combination with the protocol "RNA cleanup" which is provided in Appendix 6.3 Again Quality and concentration of the RNA samples (1 ul) were measured using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific).

RNA isolation (Oñate-Sánchez *et al.* 2008), RNA quality test (spectrophotometer) and DNase digestion

The second RNA isolation protocol was developed by Oñate-Sánchez *et al.* (2008). Plant tissue was frozen using liquid nitrogen and grinded using pestle and mortal. Grounded tissue was transferred to cooled Eppendorf tubes and 550 ul extraction buffer (0.4M LiCl, 0.2 M Tris pH:8, 25 mM EDTA, 1% SDS) and 550 ul chloroform were added. The samples were vortexed for 10 seconds and kept on ice. After vortexing, samples were centrifuged for 3 minutes at full speed. Supernatants were transferred to new Eppendorf tubes and 500 ul of water-saturated acidic phenol was added. Samples were vortexed thoroughly and 200 ul choroform was added. Again the samples were centrifuged for 3 minutes at maximum speed. Supernatant was transferred to new Eppendorf tubes again and 1/3 volume of 8m LiCl was added, The samples were mixed by turning and precipitated for 1 hour at -20°C. After precipitation, the supernatant was poured off and 500 ul RNase free water was added. To precipitate carbohydrates, 7 ul 3M NaAc pH:5.2 and 250 ul ethanol were added. After mixing the samples by turning, the samples were centrifuged at maximum speed for 10 minutes at 4°C. Again the supernatant was transferred to new Eppendorf tubes. 43 ul 3 M NaAc pH:5.2 and 750 ul ethanol were added to this supernatant and the samples were mixed. Nucleic acids were precipitated by incubating the samples at -20°C for 1 hour. After precipitation, samples were centrifuged at maximum speed for 20 min at 4°C. Supernatant was poured off and the pellet was washed with 70% EtOH. Samples were vacuum dried for 15 minutes and pellets were re-suspended in 20 ul RNase free water. Quality and

concentration of the RNA samples (1 ul) were measured using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific). Quality parameters (ratios 280/260 and 260/230) and concentrations were within the workable range. DNase digestion was again performed as is described in chapter 2.6, using the "DNase I" (Invitrogen) kit.

RNA quality test by gel electrophoresis

All dyed RNA samples from both RNA isolations were loaded on a 2% agarose gel using TBE x 0.5 buffer and separated by gel electrophoresis for 30 minutes at 100V. Trays were cleaned and filled with fresh 0.5 x TBE buffer in order to reduce the breakdown of RNA. After electrophoresis the gels were photographed using an OCTOPUS FII and images were analysed. The three samples with the highest quality in every tuber shape class were used for cDNA synthesis.

cDNA synthesis

iScript™ cDNA Synthesis Kit (Bio-Rad) was used to synthesize cDNA. 500 ng RNA was used to synthesize cDNA. The final concentration of the reaction wells was: 25 ng/ul RNA, 4 ul of 5x iScript reaction mix, 1 ul of iScript reverse transcriptase in a total volume of 20 ul. The reaction mixtures were incubated in a GeneAmp® PCR-system 2700 (Applied Biosystems) following the PCR protocol shown in Table 4. Prior to qPCR cDNA was diluted 10x

qPCR

The reaction mixtures used for the qPCR consisted of 5 ul IQ™ SYBR® Green Supermix, 0.2 ul Primer Forward (10 uM), 0.2 ul Primer Reverse (10 uM), 2.6 ul Milli-Q and 2 ul cDNA(+/- 50 ng).The PCR protocol which has been used is shown in table 5 Also melting curves of the PCR-products have been performed in order to check the specificity of the primers. The PCR protocol used for the melting curves is also shown in Table 5. A BioRad CFX 96 has been used to perform the qPCR.

Results were analysed in Microsoft Excel. As a reference gene NAC has been used, because Bourke (2014) states NAC is very stable in its expression. Using Ct-values and the The $2^{-\Delta\Delta CT}$ (Livak) Method, relative expression levels were determined. Statistical significance was tested using a Student's t-test. Melting curves have been analysed in CFX manager

2.8 Peroxidase Activity Measurements

To investigate the differences in apoplastic peroxidase activity between round and long tubers, peroxidase activity measurements have been performed. These peroxidase activity measurements have been performed on apoplastic protein extractions of dormant tubers of progeny of a CxE population. Proteins were boiled down to test whether observed activity was caused by enzymes. Enzyme activity of an enzyme specific to the symplast of cells was measured to monitor potential symplastic contamination of the apoplastic protein

extractions. In order to localize apoplastic peroxidase activity tubers were cut in 4 different parts and peroxidase activity was measured separately for the for different parts.

Apoplastic protein extraction

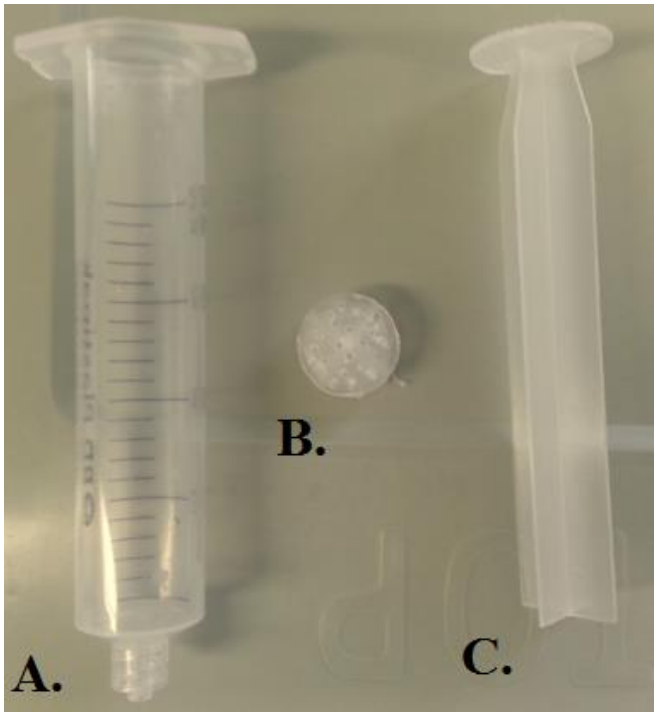


Figure 2: Syringes(A.) and the filters(B.). Part C. was used to submerge plant tissue, in case tissue started to float.

First syringes were placed in centrifuge tubes. Dormant tubers resulting from a CxE cross were cut transversally and were placed in the syringes. Three tubers per genotype were pooled and three biological replicates per tuber shape class were used. The syringes had a filter placed at the bottom in order to let fluid pass. Both a syringe and the filter are shown in figure 3. The plant tissue in the syringe was washed by filling the syringes with water and let the water pass through the filter. This washing step was repeated 2 times with the same water. After washing, the syringes were filled with 0.01 M Mes buffer (pH 5.5) containing 0.2 M KCL. This buffer was vacuum infiltrated for 15 minutes in the tissue. After vacuum infiltration, excess buffer was flown away trough the filter and last bits of excess buffer were removed syringe with paper. Syringes were placed in clean centrifuge tubes and centrifuged for 20 minutes at 1000g at 4°C. The protein concentrations of the resulting fluids were measured using NanoDrop™ 1000 Spectrophotometer (Thermo Scientific) and protein concentrations were normalized to 0.15 mg/ml.

Enzyme activity test (boiling)

In order to prove peroxidase activity and the fact that it was caused by enzymes, 20 ul of both a long and a round protein extraction was incubated at 98°C. The boiled extracts and untreated extracts were dripped on a piece of nitrocellulose membrane (BioRad 45um thick). This piece of nitrocellulose membrane was soaked in 1:1 (v/v) mixture of luminol/enhancer and stable peroxide buffer coming from the "SuperSignal West Femto Chemiluminescent Substrate" kit (Thermo Scientific) This soaked membrane was directly photographed by a G:BOX (Syngene). Lens was opened for 2 minutes in order to capture enough light emitted by the reaction. Pictures were analysed using the software Genesnap.

Symplastic contamination of apoplastic protein extractions

In order to test whether the apoplastic protein extractions were contaminated with symplastic material, the activity of the symplastic enzyme Glucose-6-phosphate

dehydrogenase (G6PD) was tested. Absorbance at a wavelength of 340 nm was measured using a BioRad Model 680 XR microplate reader at 30°C. Measurements were taken every 6 seconds during 6 minutes. Reaction mixtures consisted of 10 ul protein extract (0.15 mg/ml) and 190 ul substrate buffer (1mM Glucose-6-phosphate, 0.2mM $NADP^+$ and 1 mM MgCl2 in 0.1 M Tricine-NaOH ;pH 8.0) (Takahama, 1993) Also a positive control was incorporated by using 10 ul of G6PD (1 unit/ml). Raw data was analysed using Microsoft Excel.

Peroxidase activity measurements

Peroxidase activity was measured using 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (AzBTS-(NH4)2) which was bought at Sigma-Aldrich (A9941) as a substrate. Substrate buffer was prepared by dissolving 1 tablet of AzBTS-(NH4)2 in 100 ml 0.05 M phosphate-citrate buffer, pH 5.0. 25 ul of 30% H_2O_2 was added to the solution prior to the reactions and measurements took place. Absorbance at a wavelength of 405 nm was measured using a BioRad Model 680 XR microplate reader at 30°C. Measurements were taken every 6 seconds during 6 minutes. Reaction mixtures consisted of 10 ul protein extract (0.15 mg/ml) and 190 ul substrate buffer. Raw data was analysed using Microsoft Excel. Statistical significance was tested using a Student's t-test.

The protein extractions and the peroxidase activity measurements were also done on a long and a round dormant CxE tuber. Of these tubers peroxidase extractions and peroxidase activity was done on the pith, the sides (pooled), distal part and proximal (stolon side) part separately.

2.9 *In vitro* tuber induction

To investigate whether *in vitro* tubers could act as a way to study tuber shape, first information had to be gathered about the possibility to induce tuber formation *in vitro* for CE-clones. Also the development of the tubers had to be simultaneous to be able to compare different genotypes sampled at a certain time point. The different genotypes had to show the same differences in tuber shape as are observed in the greenhouse.

Table 6: CE-clones and their phenotypes for shape and earliness. Plants were available from the in *vitro core* collection(Plant Breeding).

CE clone	Shape	Earliness	Genotype (Earliness)
27	5	8.25	Vv
165	5	8.83	Vv
250	4.8	8.33	Vv
141	1.2	9	Vv
196	1	9	Vv
233	1	7.5	Vv

To induce tuber formation *in vitro*, the plants shown in Table 6 were grown on KI minimal medium containing 8% sucrose, cytokinin BA (10 mg/L) and agar. 5 explants per genotype were made by cutting above and below an axillary bud. In this way the explant contained a axillary bud and a leaf. Plates containing the medium and the explants were incubated in the dark at 24°C.

Respectively 13 days, 21 days and 27 days after cutting of the explants pictures were taken. Plants used in this experiment were comparable in their earliness of tuber formation.

3. Results

3.1 Searching for unannotated peroxidases

In Figure 4 the results are shown of the peroxidase transcript BLAST to the whole 280 kb long QTL. The transcripts of the different peroxidases located in the 280 kb long QTL map back to several places in this QTL. Mainly other annotated peroxidases were target to have BLAST hits for transcripts of other annotated peroxidases located in the same region. Also other regions show BLAST hits from several peroxidase transcripts. All regions containing BLAST hits outside annotated peroxidase genes were studied with both Fgenesh and the NCBI conserved domain finder. The locations which are captured by the red boxes are regions which have genes predicted by Fgenesh and these predicted genes also contain conserved peroxidase domains. Gene predictions, intron/exon structures and exact locations can be found in Appendix 6.5 The regions captured by the red boxes are considered as newly found peroxidase genes. In this study they will be named as peroxidase A, B, C, D and E

Figure 3: BLAST hits of the different peroxidase transcripts located in the Ro-locus mapped back to the Ro-locus. Newly found unannotated peroxidases are captured by the red boxes and named as is written below the red boxes. The green box captures BLAST hits mapping back to 2 different places within 1 annotated gene.

Another remarkable founding is the fact of peroxidase transcripts having BLAST hits to two different locations within the annotated gene PGSC0003DMG400020795 (captured by the green box). By studying this region again with both Fgenesh and the NCBI conserved domain finder, two functional genes containing conserved peroxidase domains are predicted. Gene predictions, intron/exon structures and exact locations can be found in Appendix 6.4 Newly found peroxidase genes and the two genes predicted from the annotated gene

PGSC0003DMG400020795(795A and 795B) are considered as existing genes and studied like all other annotated genes.

3.2 *In silico* localization of the peroxidases

The subcellular location of above described protein sequences, obtained with FGENESH or by using PGSC predicted proteins, was predicted using SignalP (Table 7). In order to get more information were peroxidases located in the *Ro*-locus were located and were peroxidase-activity should be measured.

Almost all transcripts are predicted to be secreted with a low RC value. A low RC value means that the prediction is reliable. Some transcripts are predicted to be located to the chloroplast or to "any other location", however these predictions go along with high RC values. A high RC value means that the prediction is not reliable. In fact all peroxidase transcripts which have reliable predictions go along with a protein localization to the secretory pathway. Only both transcript predictions of PGSC0003DMG400006681 are predicted to be located to "any other location with a relatively low RC value.

Table 7: Subcellular Locations of the different peroxidase transcripts. The names of the transcripts refer to the last 3 digits of their annotated gene names shown in table 1, followed by the prediction program which is used(FG: Fgenesh; msu: PGSC Representative Gene Models) In the second column the subcellular location of the transcripts is shown (S: Secretory; C: Chloroplast; _: Any other location) The third column contains the reliability of the prediction (1: diff > 0.8; 2: 0.800 > diff > 0.600; 3: 0.600 > diff > 0.400; 4: 0.400 > diff > 0.200; 5: 5 : 0.200 > diff).

Name	Loc	RC
679msu	S	1
679FG	S	1
680msu	S	1
680FG	S	1
681msu	_	2
681FG	_	2
800msu	_	5
800FG	S	1
799msu	_	4
799FG	S	1
798msu	_	5
798FG	_	3
795msu	C	5
795AFG	S	1
795BFG	S	1
801msu	C	4
801FG	S	1
A	S	1
B	S	1
C	S	1
D	S	1
E	S	1

3.3 Peroxidase Domain Study

The functionality of the peroxidases located in the *Ro*-locus was studied by aligning all peroxidase proteins and search for the conserved peroxidase domains as described by Dellanoy *et al*. (2003) and Passardi *et al*. (2004a). Alignments of all peroxidase proteins located in the Ro-locus is shown in Appendix 6.6. In Table 8 the results of this study are shown. Only the transcripts from the genes B, D, E, PGSC0003DMG400020801 and the Fgenesh prediction of gene PGSC0003DMG400020800 seem to have all four conserved peroxidase domains or peroxidase domains with some minor changes.

For most genes both the Fgenesh predictions and the predictions PGSC prediction contain the same domains, except for the genes PGSC0003DMG400020800 and PGSC0003DMG400020795. The Fgenesh prediction of PGSC0003DMG400020800 contains the domain FHDC where the other gene prediction of this gene does not contain this domain. For PGSC0003DMG400020795 Fgenesh splits the gene in two separate genes containing the domains FHDC, GGHT and GQIR, whereas the prediction made by "PGSC Representative Gene Models" only predicts one transcript containing only the FHDC domain.

So domains are present in the genomic sequence, however these domains are not included by some gene prediction programmes.

Table 8: Conserved peroxidase domains present in the different peroxidase transcripts. The names of the transcripts refer to the last 3 digits of their annotated gene names shown in Table 1, followed by the prediction program which is used(FG: Fgenesh; msu: PGSC prediction) In column 2 to 4 the different conserved peroxidase domains as described by Dellanoy et al. (2003) and Passardi et al. (2004a) are shown. The fourth column contains some remarks on the reference sequence.

Name	FHDC	VSCAD	GAHT	GEIR	Remarks sequence
679FG	FHDC	-	GGHT	GQIR	
679msu	FHDC	-	GGHT	GQIR	
680FG	FHDC	-	-	-	gap
680MSU	FHDC	-	-	-	gap
681FG	-	-	GGHT	GQIR	Mist TSS
681MSU	-	-	GGHT	GQIR	Mist TSS
795AFG	FHDC	-	GGHT	GQIR	
795BFG	FHDC	-	GGHT	GQIR	
795MSU	FHDC	-	-	-	2 split
798FG	-	VSCAD	GAHT	-	
798MSU	-	VSCAD	GAHT	-	
799FG	-	VSCAD	GAHT	GQIR	
799MSU	-	VSCAD	GAHT	GQIR	
800FG	FHDC	VSCAD	GAHT	GQIR	
800MSU	-	VSCAD	GAHT	GQIR	
801FG	FHDC	VSCAD	GGHT	GQIR	
801MSU	FHDC	VSCAD	GGHT	GQIR	
AFG	FHDC	-	GGHT	GQIR	edge of scaffold
BFG	FHDC	VSCAD	GSHT	GQIR	
CFG	FHDC	-	-	-	
DFG	FHDC	VSCAD	GSHT	GQIR	
EFG	FHDC	VSCAD	GSHT	GQIR	

3.4 Peroxidase localization by tissue-prints

The localization of peroxidase activity in dormant tubers from CxE progeny has been studied by tissue-printing. In Figure 5 the results of the tissue-prints are shown. From these results it is hard to make clear where in the tissue specific peroxidases are active. Almost the whole tissue-print emits high intensity's of light. It is even hard to distinguish the tuber shape of the long and the round potato.

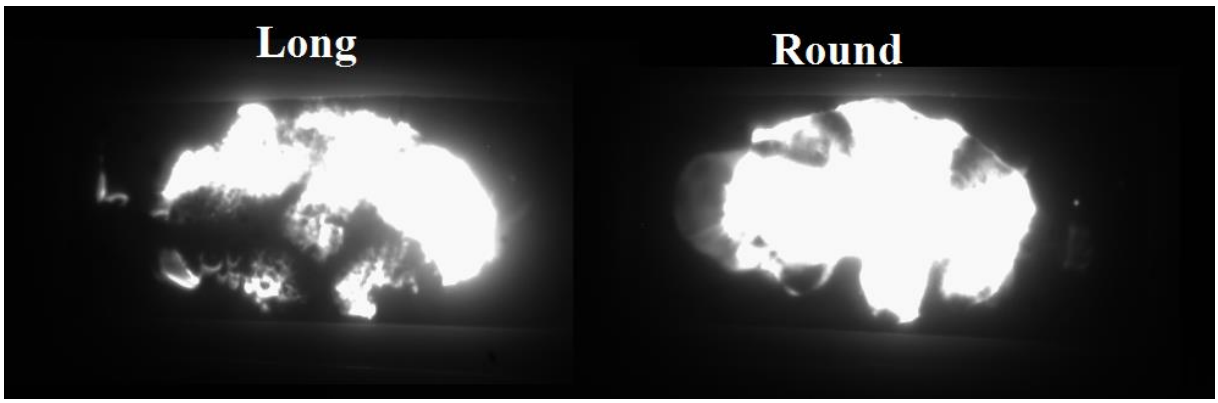


Figure 4: Results of the tuber tissue-prints. On the left side of the figure light emission caused by peroxidases catalysing the reaction between H_2O_2 and luminol of a long genotype and on the right side of a round genotype

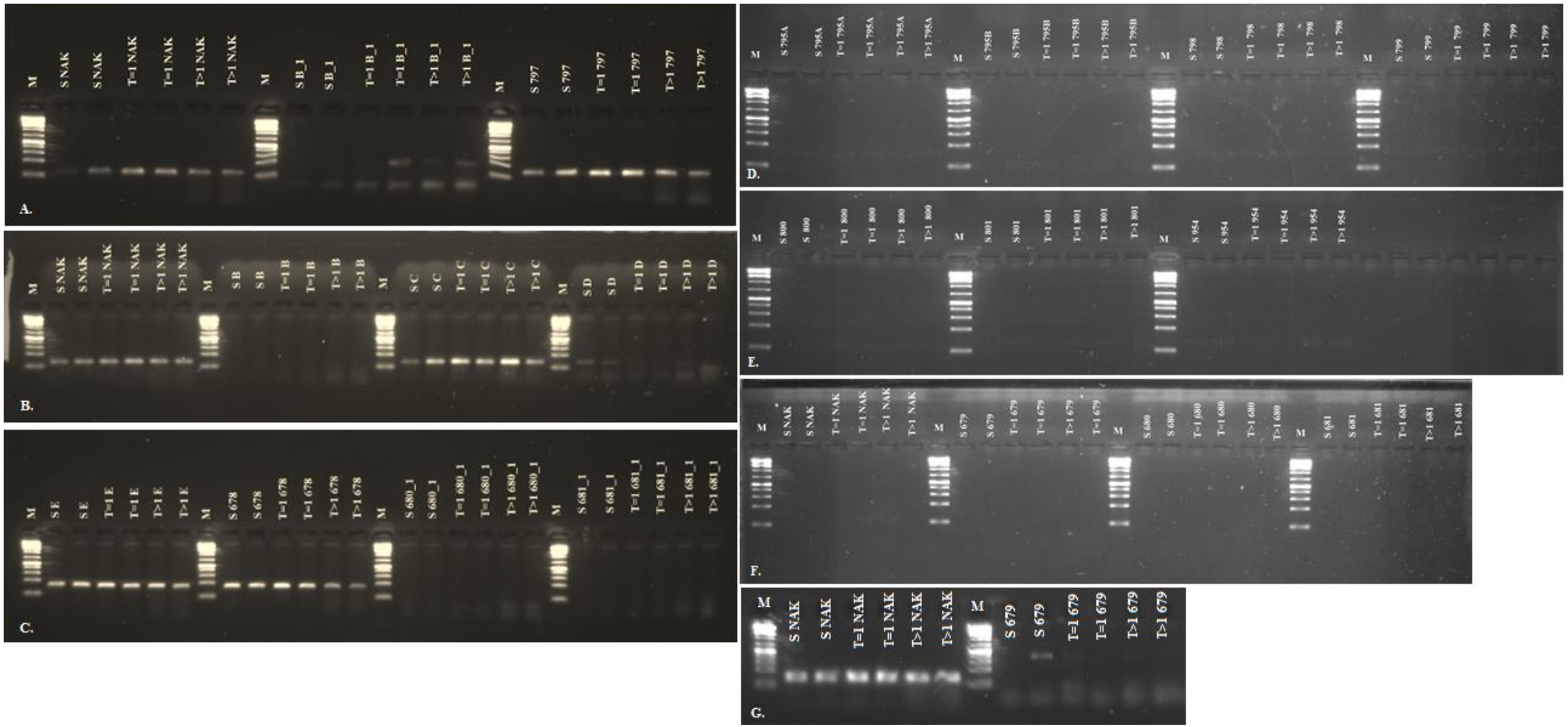


Figure 6: Results of the PCR on *Solanum tuberosum* RH cDNA. The names above the wells' refer to the cDNA samples (S: Stolon; T=1: Tubers of 1 cm; T>1: tubers bigger than 1 cm) followed by the last 3 digits of the annotated gene names shown in table 1. Gels A., B. and C. are the product of one PCR. The same applies to the gels D., E. and F. NAC has been used as a positive control.

3.5 Primer development

Primers used for qPCR have been tested by normal PCR. PCRs were performed on cDNA of a developmental series of *Solanum tuberosum* RH and on progeny of a CxE population. The results of the PCRs on RH are shown in Figure 6 For both the genes PGSC0003DMG400006680 and PGSC0003DMG400006681 two different primer sets were designed, however never a band was observed for these genes. For all other genes specific bands were observed of the expected sizes. PCR-product sizes which go along with the different primer combinations can be found in Appendix 6.1. Really low PCR-product signals are observed for the genes on gel D., E. and F of figure 6 However, PCR-products for these genes can be observed for all genes except for the genes PGSC0003DMG400006680, PGSC0003DMG400006681 and PGSC0003DMG400006679. Specificity of the primers designed for PGSC0003DMG400006679 is confirmed on gel G. of figure 6.

The only gene which is missing, is the newly found peroxidase A. This gene showed 99.99% similarity with the annotated gene PGSC0003DMG400006681, so no specific primers for this gene could be designed. In Figure 7 a part of the alignment of the transcript of gene A and PGSC0003DMG400006681 is shown. After working with the *in silico* data and after primer design some major remarks concerning the sequence were found which are shown in table 8 A gap in the reference sequence is observed close to the gene PGSC0003DMG400006680, PGSC0003DMG400006681 is missing its transcription start-site, PGSC0003DMG400020795is split up in two different genes by Fgenesh and the newly found peroxidase A is located on the edge of a scaffold DMS100.

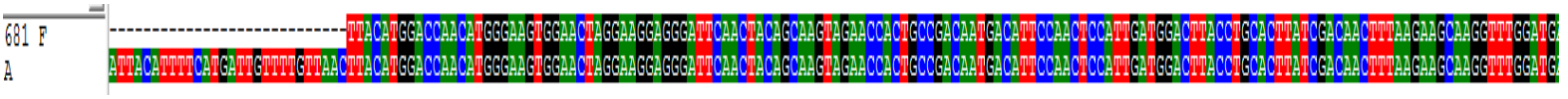


Figure 7: Small part of the alignment of the transcripts of peroxidase A and PGSC0003DMG400006681. Transcripts had 99.99% similarity. The transcript of PGSC0003DMG400006681 is missing the first part of the protein compared to the transcript of A and there is one SNP in the end of the transcript.

In Figure 8 the results of the PCR on *Solanum tuberosum* CxE genomic DNA is shown. In this PCR primer sets are less specific compared to the PCR's on *Solanum tuberosum* RH cDNA. For the genes PGSC0003DMG400006679, PGSC0003DMG400020798, PGSC0003DMG400020799, PGSC0003DMG400020800 and PGSC0003DMG400020801 specific PCR products are found of the same size. All these genes also show some unspecific bands or smears. For the gene PGSC0003DMG400020795A no PCR products are observed. The genes C, D, E, PGSC0003DMG400006678, PGSC0003DMG400020799 and PGSC0003DMG400040954 show PCR-products only in the round background. The PCR-product of PGSC0003DMG400020795B shows a small difference in length between the long and the round background. Only the PCR-products of C and PGSC0003DMG400020797 are specific in the round background.

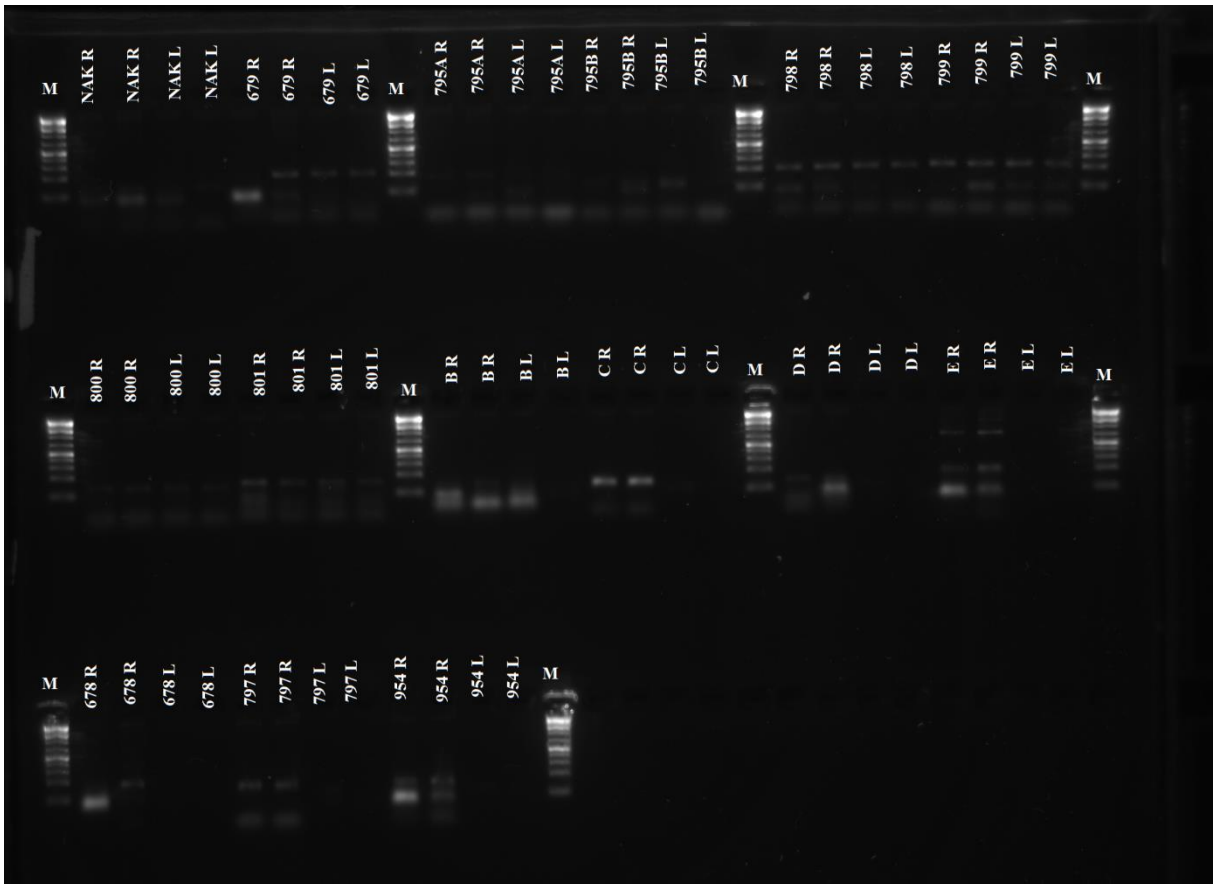


Figure 8: Results of the PCR on genomic DNA isolated from *Solanum tuberosum* CxE. Names refer to the first 3 digits of the annotated genes shown in table 1 followed by the genotype (Ro-locus) of the tuber. R refers to a tuber having 2 different round alleles and L refers to a tuber having two long alleles.

3.6 qPCR on developmental series

qPCR results

To get more information about the expression of the genes located in the Ro-locus, qPCRs have been performed on a developmental series of *Solanum tuberosum* RH. In Figure 9 the results of the qPCR on the developmental series are shown. No biological replicates were available, so statistical significance is not tested. The error bars shown in the figure reflect the variance between technical replicates. In this developmental range two different expression patterns are observed for the genes. The first pattern is a pattern were an expression decrease is observed for tubers of 1 cm. This pattern is observed for the genes PGSC0003DMG400020799, PGSC0003DMG400020801, PGSC0003DMG400020797, PGSC0003DMG400040954, C and E. The second pattern which is observed is a gradual increase of expression during tuber development. This pattern is observed for the genes PGSC0003DMG400020795A, PGSC0003DMG400020795B, PGSC0003DMG400020800 and PGSC0003DMG400006678. For the genes B and C it is hard to conclude anything about the expression patterns during development because variance between technical replicates was too high to have reliable results. In Appendix 6.7 the results of the melting-curves are shown. All genes shown in Figure 9 show just one melting-temperature, which only varies in a range of 0.5°C between different samples tested with the same primer set. This implies

PCR-products are specific. Aberrant melting-curves have been observed for B, which could be the cause for the large differences between technical replicates.

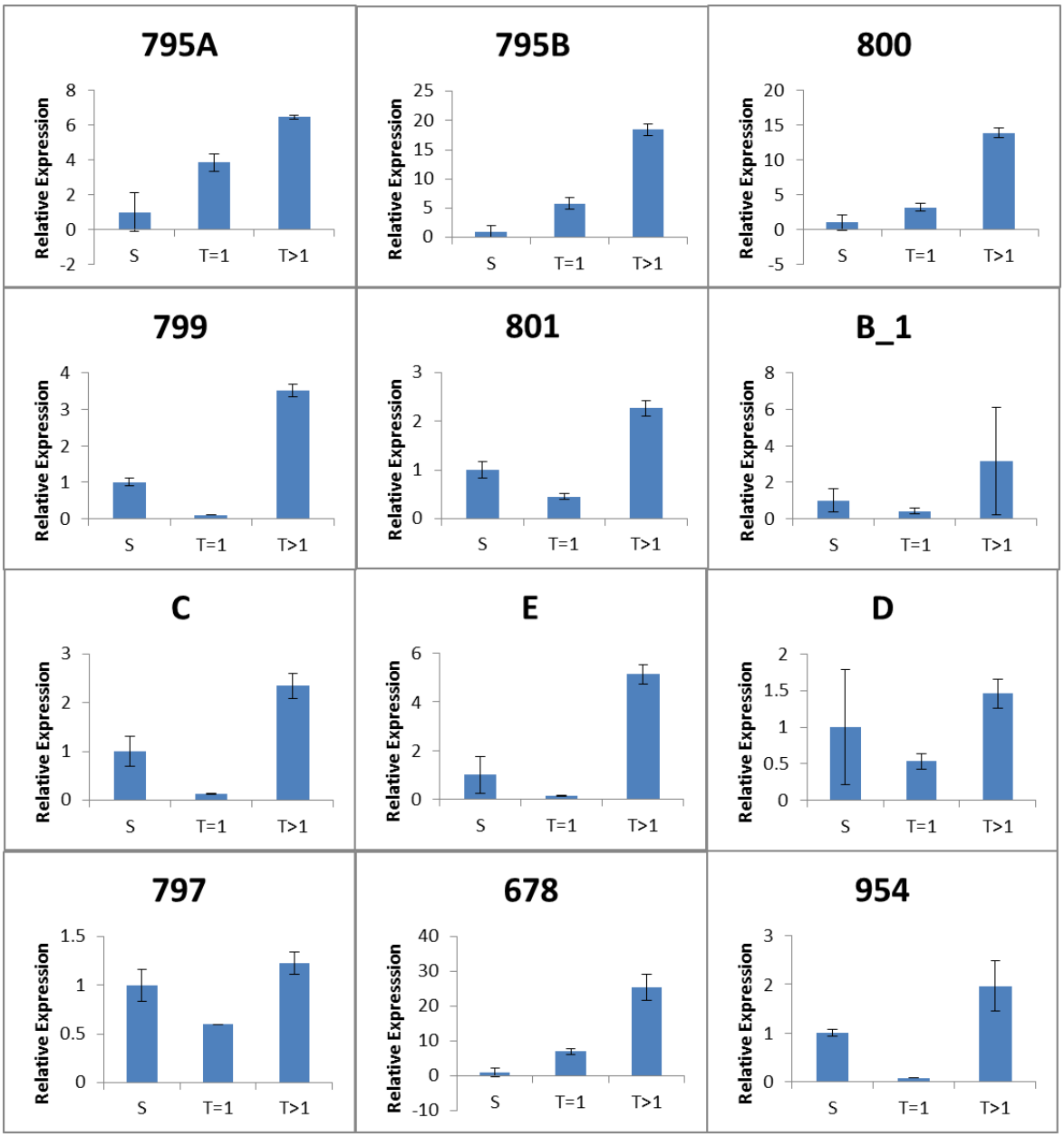


Figure 9: Expression study on a developmental range of *Solanum tuberosum* RH. Titles of the graphs correspond to the last 3 digits of the annotated gene names shown in table 1. Relative gene expression was measured in stolons (S), tubers of 1 cm(T=1) and tubers larger than 1 cm(T>1). Error bars represent the standard error of the mean of the 3 technical replicates which are used.

RNA-seq data

To compare the expression profiles determined by qPCR with RNA-seq data, FKPM values for *Solanum tuberosum* RH stolons, young tubers and tubers mature tubers were studied (PGSC, 2011) .In Figure 10 the FPKM-values of the all annotated genes located in the Ro-locus which have RNA-seq support is shown. Again the tissues stolons, tubers of 1 cm and tubers bigger than 1 cm are used for the analysis. In these data 4 expression patterns can be

observed. The first pattern is an gradual increase of expression during development. This expression pattern is observed for the genes PGSC0003DMG400006678, PGSC0003DMG400020798, PGSC0003DMG400006681, PGSC0003DMG400006679 and PGSC0003DMG400006680. The second pattern which is observed is a gradual decrease of expression during development. This pattern is observed for the genes PGSC0003DMG400020797, PGSC0003DMG400020795 and PGSC0003DMG400020801. The third expression pattern which is observed is a pattern where an expression dip in tubers of 1 cm is observed. This pattern is observed for the genes PGSC0003DMG400020799 and PGSC0003DMG400020800. The last expression pattern which is observed is observed for gene PGSC0003DMG400040954. No expression is observed for this gene in all the tested tissues.

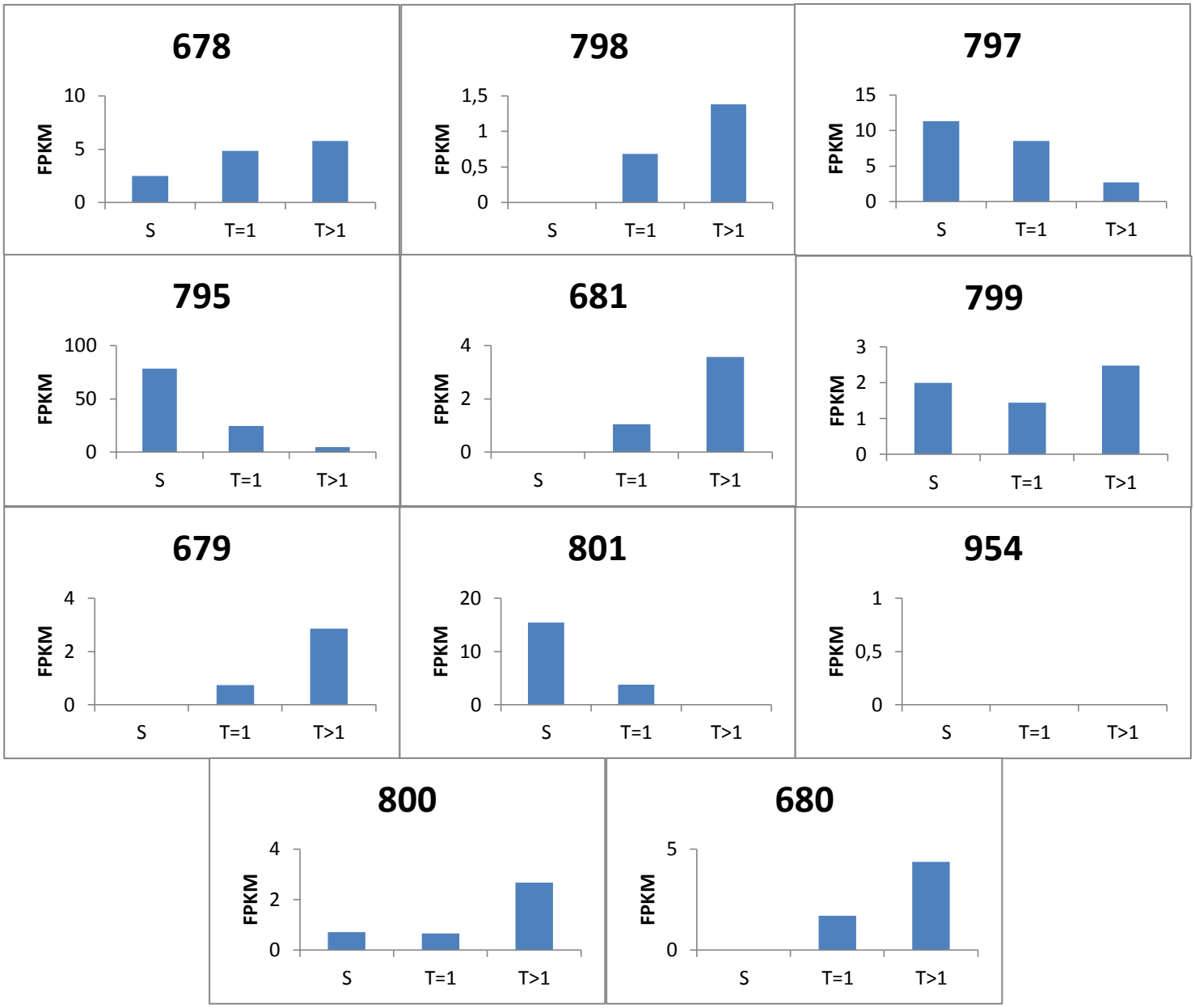


Figure 10: FKPM values of the different genes having RNA-seq support located in the Ro-locus. The titles of the graphs refer to the last three digits of the annotated gene names shown in table 1. FPKM-values are shown for stolons(S), young tubers (T=1)and mature tubers(T>1)(PGSC, 2011)

3.7 qPCR on round and long tubers

RNA-isolations

Two different RNA-isolations were used to isolate RNA for cDNA synthesis. In Figure 11 the results of both RNA isolations are shown. The three samples with the highest quality per tuber shape class are used for cDNA synthesis. From the RNA-isolation shown in Figure 10A sample R2, R3 and R4 are used for cDNA synthesis. From the RNA-isolation shown in Figure 10B sample L1, L2 and L3 are used for cDNA synthesis

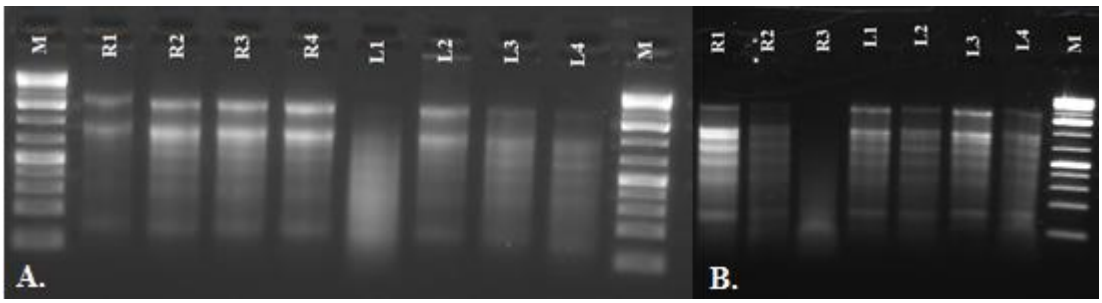


Figure 11: Results of the two different RNA-isolations. are shown. In A. the results of the RNA-isolations using the Z6 extraction buffer are shown. In B. the results of the RNA-isolations using the method developed by Oñate-Sánchez *et al.* (2008)

qPCR results

Results of the qPCR on long and round dormant tubers resulting from a CxE cross is shown in Figure 12 Only four genes are shown in this figure because hardly any expression was observed for the other genes located in the Ro-locus. However, when expression was observed, it was always observed in the round background. In fact, the same rule applies to the genes shown in Figure 12. Always a higher expression is observed in the round background compared to the long background. Only for the gene PGSC0003DMG400006678

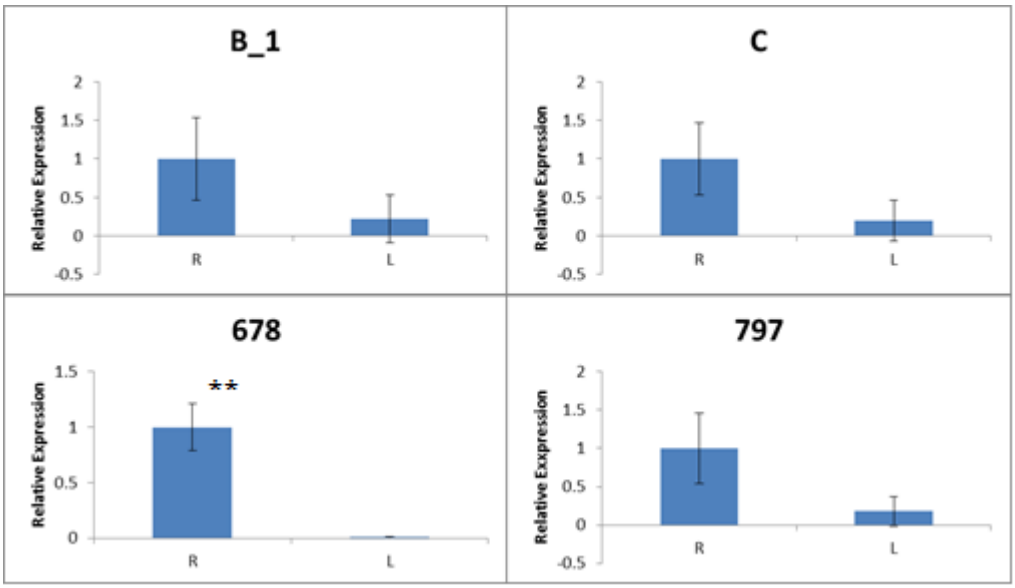


Figure 5: Expression study of the genes located in the Ro-locus in both genotypes bearing long (L) and round tubers (R). Titles of the graphs correspond to the last 3 digits of the annotated gene names shown in table 1. Error bars represent the standard error of the mean between biological replicates. Differences were tested using a Student's t-test. (*) denotes significant (p<0.05) and (**) denotes highly significant (p<0.01).

a significant difference is observed (P<0.01). However, all other genes show the same trend. In Appendix 6.8 the results of the melting-curves are shown. The genes PGSC0003DMG400006678, Peroxidase C and PGSC0003DMG400020797 show just one melting-curve, does not vary between different technical replicates tested with the same. This implies PCR-products are specific. For Peroxidase B unspecific melting-curves are observed. This could be caused by unspecificity of the primers or by the formation of primer dimer. When some expression was observed for the genes not shown in Figure 12, the melting temperatures of the PCR-products as result of a particular primer set were the same.

3.8 Peroxidase Activity Measurements

Enzyme activity test (boiling)

To test whether peroxidases could be extracted via the apoplastic protein extraction described in chapter 2.8, reactions with luminol were performed. Protein extracts were boiled down to check whether the light emission was caused due to the presence of proteins. In Figure 13A the results of this experiment are shown. Clear red spots were observed for both the extractions of the a long and a round potato. This indicates peroxidase activity. When protein extracts were boiled down these red spots disappeared. As is shown in Figure 13B also the light emission caused by the reaction is not observed anymore after boiling the protein extracts down. Proteins (peroxidases) were considered as the cause for the chemical reaction.

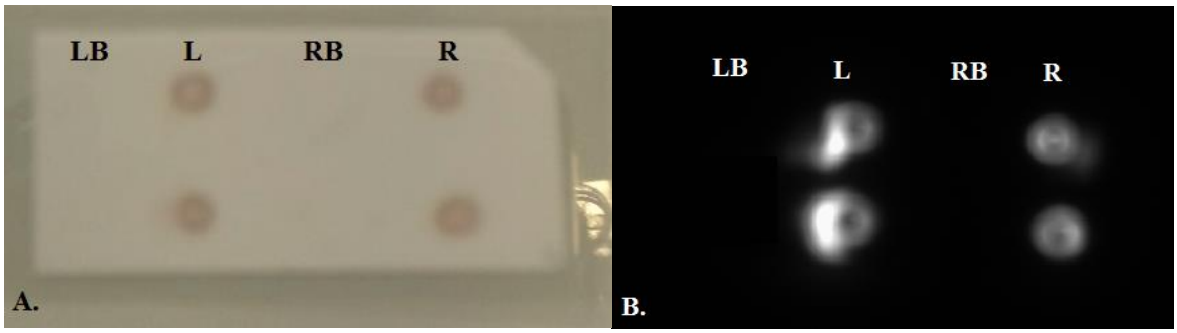


Figure 13: Results of the boiling down of apoplastic protein extractions.(L=long tuber protein extract, R=round tuber protein extract, LB=long tuber protein extract boiled and RB=round tuber protein extract boiled) In A. a picture by a normal camera is shown. In B. a picture by the G:Box is shown.

Symplastic contamination of apoplastic protein extractions

The results of the experiments to test symplastic contamination of the apoplastic protein extractions are shown in Figure 14. An increase of absorbance is observed at 340 nm for the positive control which indicates activity of the enzyme G6PD. For all other samples any increase in absorbance at 340 nm was not observed. So no activity of G6PD has been observed in the apoplastic protein extractions. Apoplastic protein extractions are considered clean enough from contamination with symplastic material to make conclusions about the apoplast itself.

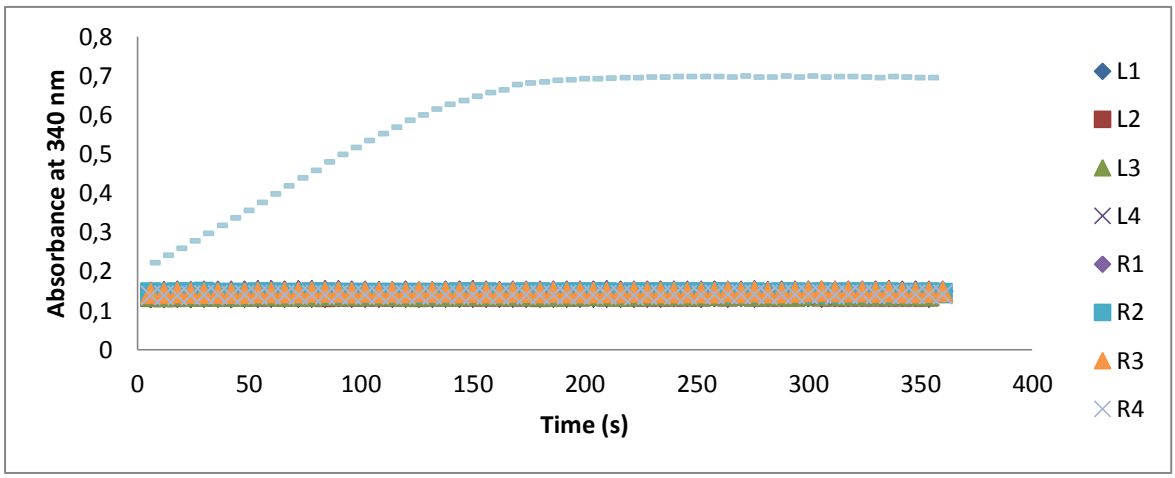


Figure 14: Results of the G6PD activity measurements. L1 to R4 represent the different genotypes used for the peroxidase extractions. C represents a positive control.

Peroxidase activity measurements

To study the differences in apoplastic peroxidase activity in dormant tubers from CxE progeny, peroxidase activity measurements were performed on apoplastic protein extractions. In Figure 15 the results of the peroxidase activity measurements are shown. The average of the four round genotypes is higher compared to the average of the four long genotypes. Quite a large variation in apoplastic peroxidase activity between the round genotypes and between the long genotypes is observed. As a result of the large variation between genotypes in a tuber class relative large standard errors are observed. No significant differences were found in this experiment ($p=0.17$).

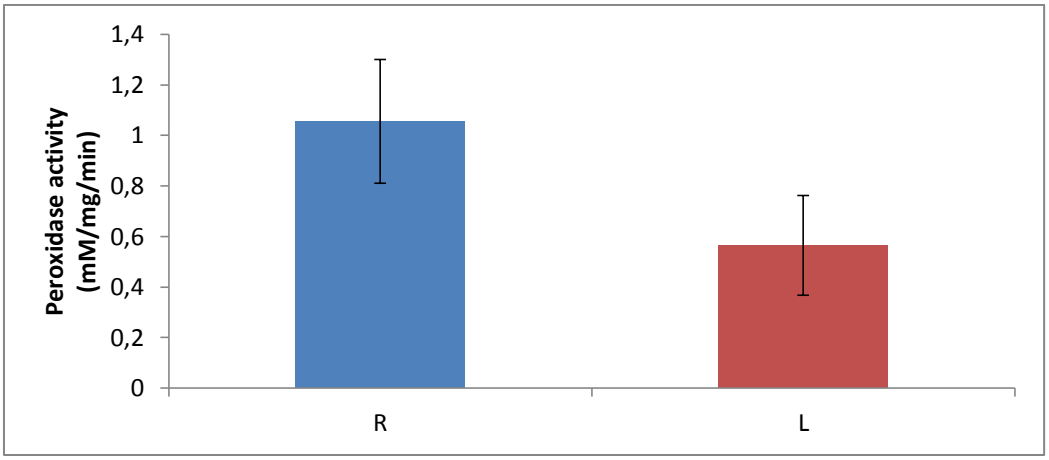


Figure 15: Apoplastic peroxidase activity of both genotypes bearing long (L) and round tubers (R). Error bars represent the standard error of the mean between biological replicates. Differences were tested using a Student's t-test. No significant differences have been found (p=0.17).

To get more information about the localization of apoplastic peroxidase activity in dormant tubers of CxE progeny, peroxidase activity of apoplastic protein extractions from four different parts of the tuber were measured. In Figure 16 the results of the apoplastic peroxidase activity of different parts of a dormant potato tubers are shown. Again the round tuber shows the trend to have a higher apoplastic peroxidase activity compared to the long tuber. Proximal and distal part of the tubers show for both genotypes the highest peroxidase activity.

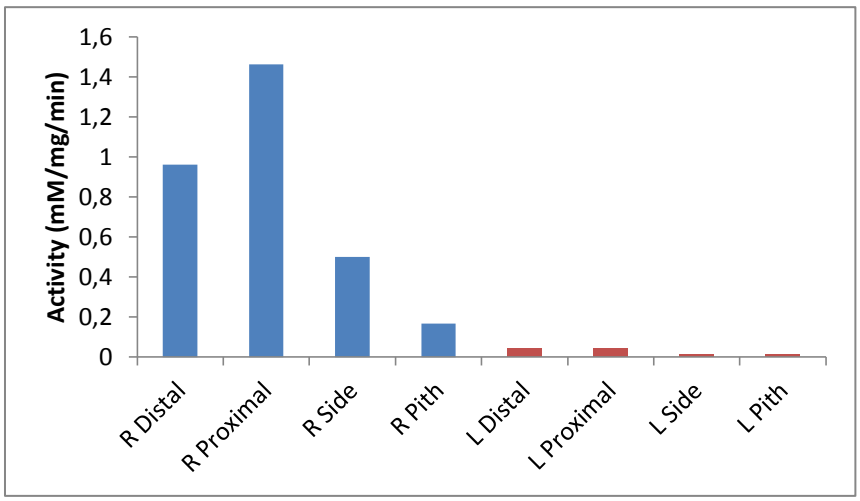


Figure 16: Apoplastic peroxidase activity of different parts of a dormant potato tuber. R states a genotype with a round phenotype for tuber shape and L states a genotype with a long phenotype for tuber shape. No biological replicates were available.

3.9 *In vitro* tuber induction

To investigate whether *in vitro* tubers could act as a way to study tuber shape, several CE-clones were treated under tuber inducing conditions like high sugars, availability of Ck and abundancy of light. The development of the potato explants cut from the plants shown in Table 6 is shown in Figure 17. After 13 days of treatment with the tuber inducing conditions all genotypes were already producing stolons. The genotype CE27 already produced small tubers on axillary buds of the stolons. Also genotype CE233 started to show development of tubers on the stolon tips, however these tubers were rather small.

After 21 days of treatment even more genotypes started to show tuber development. The genotype C196 had formed little tubers on the stolon tips. The tubers of the genotypes CE27 and CE233 were grown bigger. Genotype CE141 shows a really branchy phenotype. Axillary buds of the stolons formed even more stolons.

After 27 days of treatment also CE250 was bearing small tubers at the axillary buds of the stolons. The genotypes CE141 and CE165 did not have developed tubers at all. Despite the plants had the same genotype for earliness in tuber formation, tubers did not develop at the same time *in vitro.* It was also not (yet) possible to phenotype the tuber shape of the *in vitro* developed tubers of different genotypes.

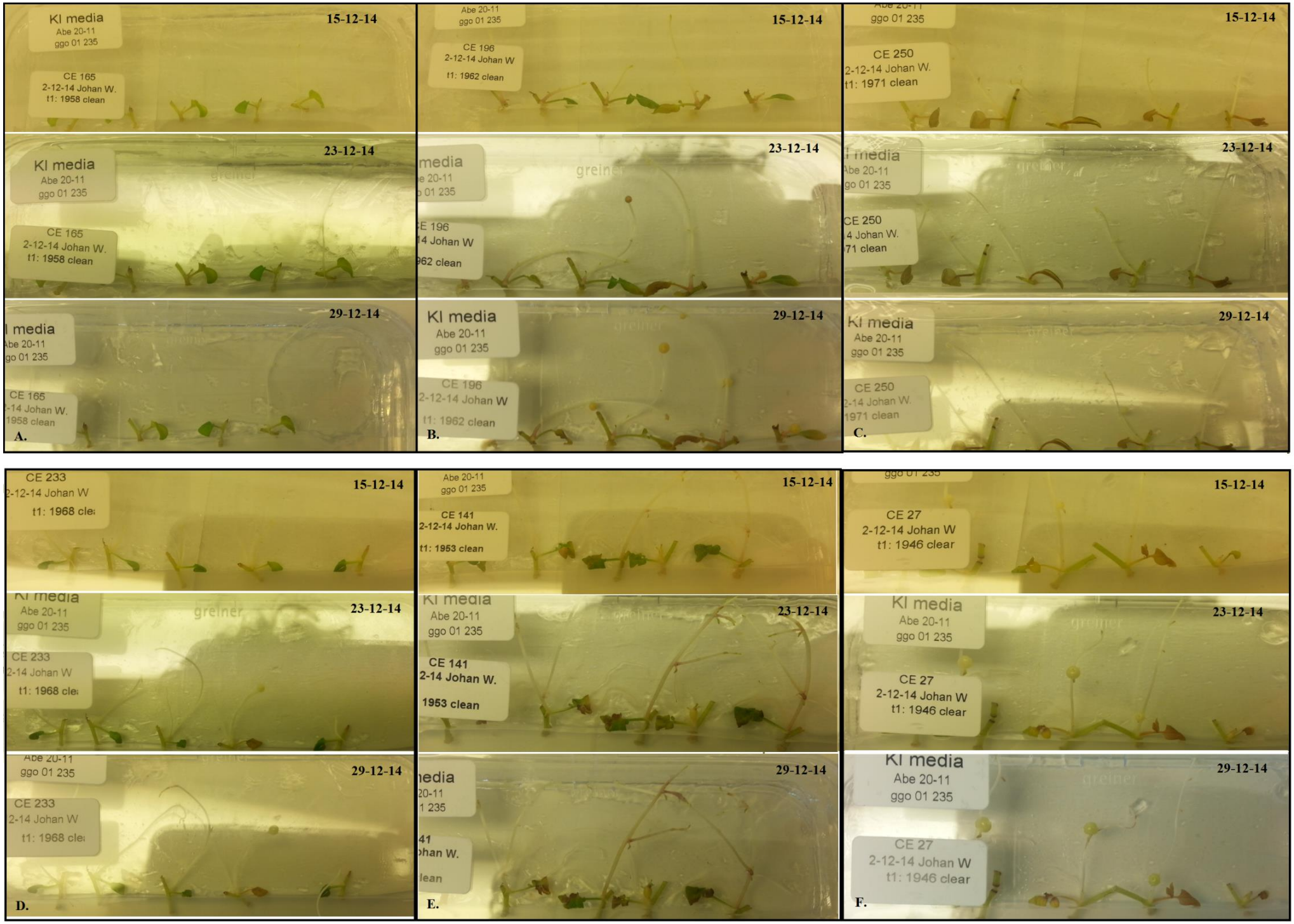


Figure 17: *In vitro* tuber development of different genotypes. A. CE165; B. CE196; C. CE250; D. CE233; E. CE141 and F. CE27. Characteristics of the genotypes can be found in Table 6.

4. Discussion

4.1 Functional annotation of the candidate region

From earlier research it is known that peroxidases tend to cluster in the genome (Welinder *et al.* 2002). Such repetitive regions in the genome are often very hard to assemble after genome sequencing. This problem is caused by the fact that sequenced fragments are often relatively short (200 bp to 700 bp). These reads are too short to capture the repetitive elements in its entirety. In this way many reads consist of only small parts of the repetitive elements, which can be mapped back to several places on the genome. Assembly of repeat regions is often of a lower quality compared to other parts of the genome. After assembly, sequence of repeat regions is often characterized by gaps.

The problem described above is causing some troubles in the assembly of the reference genome and is also observed in the different *in silico* experiments which have been performed in this study. Many gaps in the sequence have been observed in the Ro-locus. The annotated gene PGSC0003DMG400006681 was missing its transcription start site. On the other hand the unannotated peroxidase A had a 99.99% sequence similarity with PGSC0003DMG400006681. Surprisingly, the unannotated peroxidase gene A does have a transcription start site. It should be noted that unannotated peroxidase A was located on the edge of scaffold DMS100. The presence of both peroxidase A and PGSC0003DMG400006681 and the location of peroxidase A could indicate a wrong assembly of the genome. Transcription of both the transcript of PGSC0003DMG400006681 and the transcript of peroxidase A is not observed in the developmental range of *Solanum tuberosum* RH.

For the functionality of the peroxidases four different conserved domains are described in literature (Dellanoy *et al*., 2003)(Passardi et al. 2004a). The only genes which contained all four domains were B, D, E, PGSC0003DMG400020801 and the Fgenesh prediction of gene PGSC0003DMG400020800. The rest of the peroxidases. All other peroxidases missed 1 or more of the conserved domains. For these genes the functionality of the protein can be altered or even lost, which can result in pseudogenes. Surprisingly, for most genes lacking peroxidase domains expression is observed in the developmental series of RH. Of course, missing of such important peroxidase domains can also be caused by errors in the assembly of the reference sequence. It also depends which gene prediction software is used because for example the Fgenesh prediction of the gene PGSC0003DMG400020800 contains all four domains, where the prediction by “PGSC Representative Gene Models” is missing the domain characterized by the sequence “FHDC”. Also Fgenesh predicts two separate genes for PGSC0003DMG400020795 where “PGSC Representative Gene Models” only predicts one gene.

The reference sequence of the *Ro*-locus is still of a relatively low quality compared to other parts of the genome. In fact, it is still not clear how many functional peroxidases there are

really located in the Ro-locus. In this study all peroxidases which showed a PCR-product on gel after a PCR on the developmental series of *Solanum tuberosum* RH cDNA were also tested in the two different qPCRs. The genes which did not show a PCR-product in the primer test on the developmental series still could be functional peroxidases. These genes could for example be expressed in other tissues or during other stages of tuber development.

4.2 *In silico* prediction of the subcellular location of the peroxidase proteins

In order to extract the peroxidases Ro-locus from the correct subcellular locations, we first had to know were these proteins were located at subcellular level. From the predictions shown in Table 1 we can conclude that all peroxidase proteins will be located to the secretory pathway. Only the gene predictions which are located to the secretory pathway go along with a high reliability of the prediction. The secretory pathway means that proteins will most likely be located to the apoplast of the cells. From literature it is known that peroxidases can act on the stiffness of cell walls(Passardi *et al*., 2004). Cell walls are located in the apoplast of the cells. Peroxidases located in the Ro-locus are considered as Class lll Peroxidases, because they are predicted to be secreted most likely to the apoplast of the cells. Therefore, enzyme extractions done for peroxidase activity measurements will specifically target the apoplast.

4.3 Peroxidase localization in tubers by tissue-prints

Unfortunately, it is not visible where peroxidases activity is localized in a tuber. Both the stamps of a round and the stamps of the long tubers emit to much light to see any pattern. This quantity of light could possibly be caused by symplastic peroxidases which leaked out of the ruptured cells on the cut surface. Also the amount of substrate added or the opening time of the lens could be the cause of the high light intensions on the pictures.

It was hard to make a nice tissue-print of the whole cut-surface because tubers were only cut in two parts. Some parts of the cut-surface did not touch the nitrocellulose membrane because the tuber-half was not flexible enough. This problem can possibly be solved by cutting thin flexible sections of a potato tuber and use these for the tissue-prints. When somewhere in the future a confirmation of the involvement of peroxidases in determination of tuber shape is reached, it is really important for the understanding of these mechanisms were peroxidases activity is localized in the tubers.

In order to get more knowledge about the localization of peroxidase activity in potato tubers, it is important to improve the method of tissue-printing. A protocol as is described by Liu et al. (2014) could be used. In this protocol H2O2 is stained of which the concentrations are regulated by peroxidases. This protocol can also be amended with another substrate like luminol or o-dianisidine/H2O2 to visualize peroxidase activity directly (Dunand *et al.* 2007). DAB has to be replaced by one of these substrates.

4.4 Primer development

In order to perform qPCRs, primers were designed and tested by PCR on a developmental series of *Solanum tuberosum* RH. The genes annotated PGSC0003DMG400006680 and PGSC0003DMG400006681 did not show a PCR-product after a PCR on the developmental series. However, for the newly found peroxidases B, C, D and E transcription PCR-products were observed after the PCR on the developmental series. Which is an indication of the presence of these unannotated genes in the Ro-locus. PCR-products have not been sequenced, so it is still possible that the products are a mix of different peroxidases located elsewhere on the genome. The chance that the primers targeted several peroxidases is rather small, because the primers were designed in the 3'-UTR of the transcripts. UTR is known to be less prone to selection because changes in these parts of the transcript do not result in amino acid changes. So the highest variation between transcripts of the highly repetitive cluster of peroxidases can be found in the UTR.

Another PCR has been done on genomic *Solanum tuberosum* CxE DNA to prove the primers would also work for this race. Only the primer sets of the genes PGSC0003DMG400020797 and C could be validated in the round background. None of the other primer sets could be validated by this PCR, because almost all PCR's show unspecific bands or did not have any PCR-product. This could be caused by the lower annealing temperature (55°C instead of 60°C) which has been used for this PCR. By lowering the annealing temperature the specificity of the primers decreased and in this way primers could be able to target other peroxidases on the genome. Also genomic DNA was used which already was stored at 4°C for approximately two years (2013), so some degradation of the DNA could have taken place during that storage time. Unfortunately, no other genomic DNA nor cDNA was available of *Solanum tuberosum* CxE. Primer sets confirmed by the PCR's on Solanum tubersum RH cDNA were used for the expression experiments.

Melting-curves are specific for the different genes which showed expression in the developmental range of RH. Also almost all genes which showed expression in the dormant tubers of CxE progeny showed specific melting-curves, except for peroxidase B. This could be caused by unspecific amplification in cDNA of dormant CxE tubers. However, this aberrant melting curves could also be caused by primer-dimer. When no expression of the gene is present in the tissue, primer-dimer could be amplified. Of course also a mistake in the procedure of qPCR could cause troubles with the amplification of the particular transcript.

4.5 qPCR on developmental series

Two different groups of transcription patterns are observed during the development of a *Solanum tuberosum* RH tuber. The first transcription pattern is characterized by an expression dip for the tubers of 1 cm. This first transcription pattern has its expression peak in the tubers bigger than 1 cm. The second transcription is characterized by an gradual

increase of expression during development. In both classes of transcription peroxidases can be found. This differential expression of peroxidases during development of a tuber could indicate that the expression of just one gene or a couple of genes is responsible for the shape of a tuber. Due to the differential expression of peroxidases a collective increase or decrease in expression of peroxidases seems to be sorted out. However, no biological replicates are used in this experiment and no divergence in round and long genotypes has been made in this experiment.

By comparing both the expression profiles which are generated by qPCR and the expression profiles which are generated by the FKPM-values clear differences can be observed. The genes PGSC0003DMG400020795 and PGSC0003DMG400020801 show a gradual decrease of expression for the FKPM expression data whereas both genes show totally different expression pattern in the qPCR experiments. PGSC0003DMG400020795 is showing a gradual increase during development and PGSC0003DMG400020801 is showing a pattern with an expression dip in tubers of 1 cm. Another interesting point is the fact that no expression is observed for PGSC0003DMG400040954 in the FPKM expression data, whereas clearly expression is observed in the qPCR data for this gene. The expression patterns of PGSC0003DMG400020797 do also not match when comparing both expression data sets. For the genes PGSC0003DMG400020799, PGSC0003DMG400020800 and PGSC0003DMG400006678 the expression profiles of both the FPKM and the qPCR data sets somehow show the same expression patterns. For the genes PGSC0003DMG400020798, PGSC0003DMG400006681, PGSC0003DMG400006679, PGSC0003DMG400006680, B, C, D and E no comparison could be made, because in one of both data sets information for that particular gene is missing.

Many differences in trends are observed for the expression patterns of the different genes in the Ro-locus in the two different expression data-sets. These differences could originate from several different possibility's. The qPCR data set is based on a RNA-extraction of one single genotype. The only thing what is known about this genotype is the fact that it was isolated for *Solanum tuberosum* RH. So no biological replicates are included in this qPCR experiment. Differences in expression trends could also be caused by differences in genotype within *Solanum tuberosum* RH or caused by differences in sampling time. Also the differences could be caused by small differences in developmental stage of the tubers between the two different data-sets. The data-set generated by PGSC (2011) was also produced for another purpose than describing the expression differences of a gene within a developmental range. The data-set generated by PGSC (2011) was generated to annotate as many genes as possible and that is why so many different tissues and treatments are included.

4.6 qPCR on round and long tubers

Two different RNA-extractions have been used to extract RNA from the different long and round genotypes. In the RNA-extraction using the Z6-extraction buffer(Logemann et al. 1987), tissue of the long genotypes was grinded first. Only the quality of the RNA-extractions of the round genotypes was high enough to synthesize cDNA. In the RNA-extraction as described by Oñate-Sánchez *et al.* (2008), tissue of the round genotypes was grinded first. From this extraction only the quality of the long genotypes was high enough for cDNA synthesis. After all tissues were grinded, the rest of the protocol was proceeded, therefore samples which were grinded first were able to thaw before proceeding with the rest of the protocol. This is most probably the cause of the degradation of the samples which were grinded first in both RNA-extractions.

Grinding of the tissue took quite a long time in both RNA-extractions, because frozen tuber-tissue is very hard to grind with pestle and mortar. In the future, it is better to grind tuber tissue by wrapping tuber tissue in aluminium foil, freezing it in liquid nitrogen and smash it with a hammer. In this way grinding of the tissue would take less time. It is also better to directly add the first buffer of the protocol to the grinded tissue and to place it directly on ice. For future extractions of tuber RNA, it is also better to use the protocol described by Oñate-Sánchez *et al.* (2008). Quality parameters (ratios 280/260 and 260/230) were directly good enough for cDNA synthesis. The protocol using the Z6-buffer resulted in RNA-samples with poor values for the quality ratios 280/260 and 260/230. This RNA first had to be cleaned up before further use.

For only a few genes expression results could be generated by qPCR on the dormant *Solanum tuberosum* CxE tubers. Most probably this is caused by a differential expression of the genes located in the *Ro*-locus during development. Some genes are only expressed early in development. Expression of these genes can't be measured anymore in later stages of development.

For every gene, the trend seems to be a higher expression in the round genotypes compared to the long genotypes. For gene PGSC0003DMG400006678 this difference is even highly significant. When expression was observed in the other genes located in the Ro-locus this was most of the times observed in the round genotypes. This could indicate a general enhancement of gene expression over the whole region. The most logical explanation for a general enhancement of gene expression over the whole region would be an artefact in the procedure of qPCR in this case. Hardly any expression data is available about the peroxidase genes located in the *Ro*-locus. The expression data which is available about the peroxidases is unreliable because of the high variance between biological replicates. However, such a general enhancement of gene expression over the whole region can also have another cause.

For example, such an general enhancement of gene expression could be caused by epigenetics. A region can be methylated which effects the expression of the genes in this region. When epigenetics are the cause of such an expression difference, revertants should have been observed. This means that a long genotype becomes a round genotype or vice versa. However, these revertants have never been observed in CxE populations.

Another possible explanation could be the presence of an enhancer which effects the expression of the whole region. By "looping" of the DNA such enhancers could affect further located promoters on the DNA (De Laat *et al.* 2008)(Bulger and Groudine, 2011). Mendes *et al.* (2013) has studied this phenomena for the gene VERDANDI (VDD) in *Arabidopsis thaliana*. The interaction between enhancer and promotor of the gene VDD is on a short range . However, Mendes *et al.* (2013) is stating that most of the loop formations that have been intensively studied are related to long-range DNA looping. Whether such enhancer-protein interactions could also exist in a 280 kb long region of *Solanum tuberosum* is not known. Before more research is done to these possibilities of general enhancements of gene expression, first more information about the expression of the genes located in the Ro-locus has to be gathered.

4.7 Peroxidase Activity Measurements

In the dormant tubers resulting from a C x E cross a higher apoplastic peroxidase activity in the round tubers compared to long tubers seems to be the trend. Unfortunately, no significant difference has been found in this study. By repeating the apoplastic peroxidase activity experiments with a larger number of replicates, it is maybe possible to find a significant difference. Of course, many more peroxidase which are located on other places on the genome are located to the apoplast at a subcellular level. Peroxidase activity measured in the apoplast is therefore not only caused by the peroxidase genes located in the *Ro*-locus.

Normally, expression of certain genes and the presence of certain proteins is correlated. In the dormant *Solanum tuberosum* CxE tubers this correlation is not found between the expression of peroxidases in the qPCR experiments and the peroxidase activity experiments. The lack of this correlation could be caused by the abundancy of peroxidase proteins in the apoplast transcribed on other places on the genome. In this way expression of the peroxidases located in the Ro-locus can be low and apoplastic peroxidase activity can be high. Another explanation can be the fact that peroxidases are relatively stable proteins (Modal *et al.* 1993)(Modi et al. 1990)(Chattopadhyay and Mazumdar, 2000) . In this way expression of certain genes can be low, but the proteins are still present at their subcellular location.

It is important to perform the peroxidase activity measurements at the same time to get reliable results. Both H_2O_2 and the substrate AzBTS are relatively unstable. Both

concentrations of H_2O_2 are affecting the kinetics of the peroxidase enzymes. When for example round and long tubers are measured on a different time-point, concentrations of the substrates can be altered. It is also important to perform the peroxidase extraction of the samples at the same time. Over time, proteins will degrade. In this way it is possible that differences due to breakdown of proteins will interfere with your actual activity measurements.

Clear differences in peroxidase activity are observed between different parts of the tuber. However, only one tuber of both tuber shape class is tested. No statistical test have been done on this data due to the unavailability of biological replicates. In order to get a better understanding about the differences in peroxidase activity in the different parts of a tuber, biological replicates have to be measured for both tuber shape classes. Also the way tubers will be cut has to be taken into account because of the shape differences, for example the distal part of a long tuber is longer compared to the distal part of a round tuber. A better way to divide the tubers in the four different parts is to cut the tuber in two halves longitudinally and cut above and below the pith which can be seen in the halves. In this way a clear deviation of the tubers is obtained.

4.8 *In vitro* tuber induction

For most genotypes we managed to induce tuberization *in vitro.* Only 2 genotypes had not yet formed tubers. Unfortunately, the different genotypes did not set tubers simultaneously, despite their identical genotype for earliness of tuber formation. For the development of for example a developmental range it is really important to sample plants at an equal developmental stage. When these developmental stages are reached at the same time in several different genotypes, sampling becomes much easier. It isn't necessary anymore to measure the tubers and relate the size of a tuber to the of the developmental stage of a tuber. It seems like it depends strongly on the genotype of a plant when it starts to tuberize and whether it starts to tuberize.

After the explants were treated for 27 days under tuber inducing conditions, it was still not visible whether tubers were round or long. This could be concerning for the use of *in vitro* tubers for the different measurements, because experiments where round and long tubers are separated are not possible anymore. Maybe after some more days differences between round and long genotypes could been made. *In* vitro plants lack secondary tuber growth and whole plant dynamics. Tubers shape could be determined during secondary growth or determined by signals from other parts of the plant (Xu *et al.* 1998)(Kloosterman *et al.,* 2005). In that case differences in phenotypes would not be observed at all *in vitro*. The same genotypes which have been used for the *in vitro* experiments were also used *in vivo. In vivo* phenotypes did not correlate with the phenotypes as described in Table 6. Maybe plants have been switched in the *in vitro* core collection. Unfortunately, no hard conclusions can be drawn concerning the possibility to study tuber shape *in vitro*.

4.9 Conclusions and Recommendations

This research provides some interesting clues about a possible involvement of peroxidases influencing tuber shape. A higher apoplastic peroxidase activity in the round genotypes compared to the long genotypes is the trend which is observed in this study. Unfortunately, no significant differences have been observed in apoplastic peroxidase activity between round and long dormant tubers of CxE progeny.

Also a trend of higher expression of all the genes located in the *Ro*-locus is observed, however this could also be caused by some kind of artefact in the procedure. Dormant tubers were used for these expression analysis, where it was better to use active developing tubers in terms of gene expression. Genes of the other categories aspartate aminotransferase, non-specific-lipid-transfer protein and genes of unknown function could not be excluded as candidate genes by the expression studies nor the peroxidase activity measurements.

By studying the reference sequence of the *Ro*-locus many gaps in the sequence were observed. Also 5 unannotated peroxidase genes were found in the *Ro*-locus. One of these unannotated peroxidase genes had 99.99% similarity with the annotated gene PGSC0003DMG400006681 and was located on the edge of scaffold DMS100. The gene PGSC0003DMG400020795 was split up in two separate genes by a gene prediction using Fgenesh. Many peroxidase genes located in the *Ro*-locus are missing conserved peroxidase domains in their sequence. All these observations are indicating problems with the assembly of the reference sequence of the *Ro*-locus. It is still not known how many functional peroxidases are located in the *Ro*-locus.

In vitro tuber formation has been induced in this study, however no differences in tuber shape have been observed. The same genotypes which were grown *in vitro* were also grown *in vivo*. *In vivo* differences in tuber shape differences were not observed as was expected. It is not known whether *in vitro* tubers are suitable to study differences in tuber shape.

To characterize the gene or the group of genes it is necessary to do further research. A good way to study which gene(s) is influencing tuber shape is to produce over-expression lines of the four different groups of genes located in the *Ro*-locus. This can be done by transforming both a round and a long genotype with four different constructs respectively a peroxidase, the aspartate aminotransferase, the gene of unknown function and the non-specific-lipid-transfer protein. A tuber-specific promoter can be built in the construct to specifically over-express the genes in the tubers. By studying the phenotypes of the over-expression lines and comparing these with the untransformed genotypes, it might be possible to characterize the gene influencing tuber shape. Also a developmental series of these over-expression lines can be set up, which can be studied with all experiments described in this study. In this way also over-expression and (in case of peroxidase) higher activity can be related to the phenotypes.

5. References

- Aksenova N.P., Konstantivova T.N., Golyanovskaya S.A., Sergeeva L.I. and Romanov G.A. (2012). Hormonal Regulation of Tuber Formation in Potato Plants, Russian Journal of Plant Physiology 59, 451–466.
- Aksenova, N.P., Konstantinova, T.N., Golyanovskaya, S.A., Kossmann, J., Willmitzer, L., and Romanov, G.A. (2000). Transformed Potato Plants as a Model for Studying the Hormonal and Carbohydrate Regulation of Tuberization, Russ. J. Plant Physiol 47, 370–379.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215, 403-410
- Asada K. (1992). Ascorbate peroxidase – a hydrogen peroxide-scavenging enzyme in plants. Physiologia Plantarum 85, 235 – 241.
- Bulger M. and Groudine M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell. 144(3), 327–339.
- Chattopadhyay K. and Mazumdar S. (2000). Structural and Conformational Stability of Horseradish Peroxidase: Effect of Temperature and pH. Biochemistry 39 (1), pp 263–270.
- De Jongh and Row, P. R. (1972). Genetic markers in inbred clones of cultivated diploid potatoes. Potato Res. 15, 200-108.
- De Laat W, Klous P, Kooren J, Noordermeer D, Palstra RJ, Simonis M, Splinter E and Grosveld F. (2008). Threedimensional organization of gene expression in erythroid cells. Curr, Top, Dev, Biol. 82, 117–139.
- Delannoy E., Jalloul A., Assigbetsé K., Marmey P., Geiger J.P., Lherminier J., Daniel J.F., Martinez C. and Nicole M. (2003). Activity of Class lll Peroxidases in the Defense of Cotton to Bacterial Blight. MPMI 16, 1030 –1038.
- D'hoop, B. B., Keizer, P. L. C., Paulo, M. J., Visser, R. G. F., van Eeuwijk, F. A., & van Eck, H. J. (2014). Identification of agronomically important QTL in tetraploid potato cultivars using a marker-trait association analysis. Theoretical and Applied Genetics.
- Dunand C., Crèvecoeur M. and Penel C.(2007). Distribution of superoxide and hydrogen peroxide in Arabidopsis root and their influence on root development: possible interaction with peroxidases. New Phytologist 174. 332–341.
- Emanuelsson O., Brunak S., von Heijne G., Nielsen H. (2007). Locating proteins in the cell using TargetP, SignalP, and related tools, Nature Protocols 2, 953-971.
- Ewing, E., and Struik, P. (1992). Tuber formation in potato: induction, initiation, and growth. Horticultural Reviews 14, 197.
- FAOSTAT (2015). FAOSTAT. Retrieved January 26, 2015, from http://faostat3.fao.org/browse/Q/QC/E
- Hirsch C.D., Hamilton J.P., Childs K.L., Cepela J., Crisovan E., Vaillancourt B., Hirsch C.N., Habermann M., Neal B. and Buell C.R. (2014) Spud DB: a Resource for Mining Sequences, Genotypes and Phenotypes to Accelerate Potato Breeding, The Plant Genome 7, 1-12.
- Jackson, S.D. (1999). Multiple signaling pathways control tuber induction in potato. Plant Physiology 119, 1-8.
- Kloosterman, B., Vorst, O., Hall, R.D., Visser, R.G., and Bachem, C.W. (2005). Tuber on a chip: differential gene expression during potato tuber development. Plant Biotechnology Journal 3, 505-519.

- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) ClustalW and ClustalX version 2. Bioinformatics 23(21): 2947-2948
- Lin, K., Zhang, N., Severing, E. I., Nijveen, H., Cheng, F., Visser, R. G., … Bonnema, G. (2014). Beyond genomic variation - comparison and functional annotation of three Brassica rapa genomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics, 15(1), 250.
- Liu Y., Offler C.E. and Ruan Y.(2014). A simple, rapid and reliable protocol to localize hydrogen peroxide in large plant organs by DAB-mediated tissue printing. Front. Plant Sci. 5:745.
- Logemann, J., Schell, J. and Willmitzer, L. (1987). Improved method for the isolation of RNA from plant tissue. Anal. Biochem. 163, 16-20.
- Marchler-Bauer A., Derbyshire M.K., Gonzales N.R., Lu S., Chitsaz F., Geer L.Y., Geer R.C., He J., Gwadz M., Hurwitz D.I., Lanczycki C.J., Lu F., Marchler G.H., Song J.S., Thanki N., Wang Z., Yamashita R.A., Zhang D., Zheng C., Bryant S.H.. (2015). "CDD: NCBI's conserved domain database.", Nucleic Acids Res.43(D)222-6
- Masson, M. F. (1985). Mapping, combining abilities, heritabilities and heterosis with 4x X 2x crosses in potato. University of Wisconsin--Madison.
- Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., … Colombo, L. (2013). MADS Domain Transcription Factors Mediate Short-Range DNA Looping That Is Essential for Target Gene Expression in Arabidopsis. The Plant Cell, 25(7), 2560–2572.
- Modi, S., Saxena, A. K., Behere, D. V., and Mitra, S. (1990) Binding of thiocyanate and cyanide to manganese(III)-reconstituted horseradish peroxidase: a 15N nuclear magnetic resonance study. Biochim. Biophys. Acta 1038, 164.
- Mondal, M. S., Mazumdar, S., and Mitra, S. (1993). Binding of cyanide and thiocyanate to manganese reconstituted myoglobin and formation of peroxide compound: optical spectral, multinuclear NMR, and kinetic studies. Inorg. Chem. 32, 5362.
- NCBI (2014). BLAST. Retrieved February 2, 2015, from http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
- Obata-Sasamoto, H. and Suzuki, H. (1979) Activities of Enzymes Relating to Starch Synthesis and Endogenous Levels of Growth Regulators in Potato Stolon Tips during Tuberization, Physiol. Plant 45, 320–324.
- Oñate-Sánchez L. and Vicente-Carbajosa J. (2000) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. BMR Research Notes, 1:93.
- Passardi, F., Penel, C., & Dunand, C. (2004). Performing the paradoxical: how plant peroxidases modify the cell wall. TRENDS in Plant Science, 9 (11).
- Passardi F., Longet D., Penel C., and Dunand C. (2004a). The class III peroxidase multigenic family in rice and its evolution in land plants. Phytochemistry 65, 1879–1893.
- Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. Nature, 475(7355), 189–95.
- Prashar, A., Hornyik C., Young, V., McLean, K., Sharma S. K., M. Finlay B. Dale, Bryan, G. J. (2014). Construction of a dense SNP map of a highly heterozygous diploid potato population and QTL analysis of tuber shape and eye depth. Theor Appl Genet.

- Sharma S.K., Bolser D., de Boer J., Sønderkær M., Amoros W., Carboni M.F., D'Ambrosio J.M., de la Cruz G., Di Genova A., Douches D.S., Eguiluz M., Guo X., Guzman F., Hackett C.A., Hamilton J.P., Li G., Li Y., Lozano R., Maass A., Marshall D., Martinez D., McLean K., Mejía N., Milne L., Munive S., Nagy I., Ponce O., Ramirez M., Simon R., Thomson S.J., Torres Y., Waugh R., Zhang Z., Huang S., Visser R.G.F., Bachem C.W.B., Sagredo B., Feingold S.E., Orjeda G., Veilleux R.E., Bonierbale M., Jacobs J.M.E., Milbourne D.,Martin D.M.A. and Bryan G.J. (2013). Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. G3 3, 2031-2047.
- Śliwka, J., Wasilewicz-Flis, I., Jakuczun, H., & Gebhardt, C. (2008). Tagging quantitative trait loci for dormancy, tuber shape, regularity of tuber shape, eye depth and flesh colour in diploid potato originated from six Solanum species. Plant Breeding, 127(1), 49–55.
- Solovyev V, Kosarev P, Seledsov I, Vorobyev D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7, Suppl 1, 10.1-10.12
- Takahama U. (1993) Redox state of ascorbic acid in the apoplast of stems of Kalanchoë daigremontiana. Physiologia Plantarum 89, 791-798.
- Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30, 2725-2729
- Taylor, L. M. (1978). Variation patterns of parthenogenetic plants derived form "unreduced" embryesacosf Solanum tubemsum subspecies andigma (Jwe.t B uk.) Hawkes. Theor. Appl. Genet, 5 (2), 241-249.
- The European Cultivated Potato Database. (2015). The European Cultivated Potato Database. Retrieved From http://www.europotato.org/display_character.php?char_no=32&character=Tuber shape
- Thompson, D. S., Davies, W. J., & Ho, L. C. (1998). Regulation of tomato fruit growth by epidermal cell wall enzymes. Plant, Cell and Environment, 21, 589–599
- Uitdewilligen, J. (2012). Discovery and Genotyping of Existing and Induced DNA Sequence Variation in Potato. PhD thesis.
- United States Department of Agriculture (2015). USDA. Retrieved on February 5, 2015 from http://www.ars.usda.gov/is/graphics/photos/dec00/k9152-1.htm
- Van Eck, H. J. (2007). Genetics of Morphological and Tuber Traits. In M. A. T. and H. A. R. Dick Vreugdenhil, John Bradshaw, Christiane Gebhardt, Francine Govers, Donald K.L. Mackerron (Ed.), Potato Biology and Biotechnology (pp. 91–115).
- Van Eck, H. J., Jacobs, J. M. E., Stam, P., Ton, J., Stiekema, W. J., & Jacobsen, E. (1994). Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. Genetics, 137(1), 303–309.
- Vreugdenhil, D. and Struik, P.C. (1989). An Integrated View of the Hormonal Regulation of Tuber Formation in Potato (*Solanum tuberosum*), *Physiol. Plant* 75, 525–531.
- Welinder K.G. (1992). Superfamily of plant, fungal and bacterial peroxidases. Current Opinion in Structural Biology 2, 388-393
- Welinder, K. G., Justesen, A. F., Kjaersgård, I. V. H., Jensen, R. B., Rasmussen, S. K., Jespersen, H. M., & Duroux, L. (2002). Structural diversity and transcription of class III peroxidases from Arabidopsis thaliana. European Journal of Biochemistry, 269(24), 6063–6081.

- Willemsen, J. (2014). The identification of candidate genes responsible for variation in potato tuber shape.
- Xu X., Vreugdenhil D. and Van Lammeren A.A.M. (1998). Cell division and cell enlargement during potato tuber formation. J. Exp. Bot. 49 (320), 573-582.

6. Appendices

6.1 Primer list

Gene	Primer Name	Sequence	Chromosome	Start	End	Length of fragment
PGSC0003DMG400006679	JC_679_Rev	tgcctcttaaccacagtcca	10	49022380	49022399	174
PGSC0003DMG400006679	JC_679_For	tcatgtggcgagtgaacaaag	10	49022534	49022554	
PGSC0003DMG400006680	JC_680_Rev	ctgatcagtagacgacgaaaacaa	10	49040627	49040650	79
PGSC0003DMG400006680	JC_680_For	atcttacaacgaaatagtgcgac	10	49040684	49040706	
PGSC0003DMG400006681	JC_681_Rev	AAACATTATCTTGATTCACTGAC	10	49064429	49064451	163
PGSC0003DMG400006681	JC_681_For	GGTGAACTAAtgaaatcaagtgtg	10	49064569	49064592	
PGSC0003DMG400020795	JC_795A_Rev	CCTTGCTTCTTGAAGTTGTTG	10	49247717	49247736	145
PGSC0003DMG400020795	JC_795A_For	TCTGTAGTTGCATTAGGTGGACC	10	49247840	49247931	
PGSC0003DMG400020795	JC_795B_Rev	GGTCCACCTAACTCCGACTTG	10	49253762	49253931	86
PGSC0003DMG400020795	JC_795B_For	TGACAGTGAAAAGACTGCGATAC	10	49253975	49253997	
PGSC0003DMG400020798	JC_798_Rev	ccacacaactctcagttcacaat	10	49172288	49172310	179
PGSC0003DMG400020798	JC_798_For	attgacaggaaatgaaggctaga	10	49172445	49172467	
PGSC0003DMG400020799	JC_799_Rev	ctatttatcggaacaactcggaag	10	49151372	49151395	192
PGSC0003DMG400020799	JC_799_For	tcaagaatttggttatttgcaa	10	49151543	49151564	
PGSC0003DMG400020800	JC_800_Rev	ccagcagcgcgtacTTAGTT	10	49084179	49084198	100
PGSC0003DMG400020800	JC_800_For	GCTAAGTCTATGATTAAGATGGGC	10	49084256	49084279	
PGSC0003DMG400020801	JC_801_Rev	caatcaaattcttcaagatcaaca	10	49259883	49259906	121
PGSC0003DMG400020801	JC_801_For	TTTGCCGAGTCTATGATTAAGAT	10	49259785	49259807	
PGSC0003DMG400040954	JC_954_Rev	acgccactatgaccgaaaaa	10	49171777	49171796	118
PGSC0003DMG400040954	JC_954_For	tcgacgcataaaattggtgt	10	49171678	49171697	
B	JC_B_Rev	caccattctgatgaaatacaacac	10	49218387	49218410	124
B	JC_B_For	gaagtgctaatgatgttgatgaaat	10	49218487	49218511	
C	JC_C_Rev	gtaacgacaactcaagcaattacc	10	49209575	49209598	125
C	JC_C_For	GCTACGATTACTTTTTCATGATTG	10	49209657	49209680	
D	JC_D_Rev	CAATTATTTCATCAGCACTTCTTCT	10	49114314	49114338	103
D	JC_D_For	TGATTAAGATGGGTGATATCAAGC	10	49114395	49114418	
E	JC_E_Rev	attgacgtatagtggacgcata	10	49178993	49179014	147
E	JC_E_For	ACCGAACCACACCTCCAAA	10	49179122	49179140	
PGSC0003DMG400006678	JC_678_Rev	TACATCCTACCGAACAGCCTC	10	48982429	48982449	147
PGSC0003DMG400006678	JC_678_For	GGCAAGTGCAATCCATGAAG	10	48982302	48982321	
PGSC0003DMG400006680	JC_680_1_Rev	tcaatccataTCATGTTGCCTCTA	10	49040776	49040799	116
PGSC0003DMG400006680	JC_680_1_For	CCAACGAGCAAAATAATGATACTC	10	49040869	49043146	
PGSC0003DMG400006681	JC_681_1_Rev	GCACAATTAACTTAACCAAGGCTA	10	49064495	49064518	97(i.c.w. JC_801_For)
PGSC0003DMG400020797	JC_797_F	TGCTGAGGAGCTGAGAGGTT	10	49199342	49199361	110
PGSC0003DMG400020797	JC_797_R	ATAGAGTAATTGTGCATACACATGA	10	49199251	49199276	
B	JC_B_1_F	GCTGAGTCTATGATTAAGATGGGTG	10	49199163	49199188	111
B	JC_B_1_R	atttcatcaacatcattagcacttc	10	49218487	49218511	

6.2 RNeasy Mini Kit Quick-Start Protocol

Quick-Start Protocol

RNeasy® Mini Kit, Part 1

The RNeasy Mini Kit (cat. nos. 74104 and 74106) can be stored at room temperature (15–25°C) for at least 9 months.

For more information, additional and more detailed protocols, and safety information, please refer to the *RNeasy Mini Handbook*, which can be found at www.qiagen.com/handbooks.

For technical assistance, please call toll-free 00800-22-44-6000, or find regional phone numbers at www.qiagen.com/contact.

Notes before starting

- If purifying RNA from cell lines rich in RNases, or tissue, add either 10 µl β-mercaptoethanol (β-ME), or 20 µl 2 M dithiothreitol (DTT)*, to 1 ml Buffer RLT. Buffer RLT with β-ME or DTT can be stored at room temperature for up to 1 month.
- Add 4 volumes of ethanol (96–100%) to Buffer RPE for a working solution.
- Remove RNA*later*®-stabilized tissue from the reagent using forceps.
- For RNeasy Protect Mini Kit (cat. nos. 74124 and 74126), please start with the *Quick-StartProtocol RNAlater RNA Stabilization Reagent, RNAlater TissueProtect Tubes, and RNeasy Protect Kits*.

* This option not included for cells in handbook; handbook to be updated.

1. **Cells:** Harvest a maximum of 1 x 10^7 cells, as a cell pellet or by direct lysis in the vessel. Add the appropriate volume of Buffer RLT (see Table 1).
   Tissues: Do not use more than 30 mg tissue. Disrupt the tissue and homogenize the lysate in the appropriate volume of Buffer RLT (see Table 1). Centrifuge the lysate for 3 min at maximum speed. Carefully remove the supernatant by pipetting, and use it in step 2.
2. Add 1 volume of 70% ethanol to the lysate, and mix well by pipetting. Do not centrifuge. Proceed immediately to step 3.
3. Transfer up to 700 µl of the sample, including any precipitate, to an RNeasy Mini spin column placed in a 2 ml collection tube (supplied). Close the lid, and centrifuge for 15 s at ≥8000 x *g*. Discard the flow-through.

Optional: For DNase digestion, follow steps 1–4 of "On-column DNase digestion" in *Quick-StartProtocol RNeasy Mini Kit, Part 2*.

4. Add 700 µl Buffer RW1 to the RNeasy spin column. Close the lid, and centrifuge for 15 s at ≥8000 x *g*. Discard the flow-through.

5. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid, and centrifuge for 15 s at ≥8000 x *g*. Discard the flow-through.

6. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid, and centrifuge for 2 min at ≥8000 x *g*.

   Optional: Place the RNeasy spin column in a new 2 ml collection tube (supplied). Centrifuge at full speed for 1 min to dry the membrane.

7. Place the RNeasy spin column in a new 1.5 ml collection tube (supplied). Add 30–50 µl RNase-free water directly to the spin column membrane. Close the lid, and centrifuge for 1 min at ≥8000 x *g* to elute the RNA.

8. If the expected RNA yield is >30 µg, repeat step 7 using another 30–50 µl of RNase-free water, or using the eluate from step 7 (if high RNA concentration is required). Reuse the collection tube from step 7.

Table 1. Volumes of Buffer RLT for sample disruption and homogenization

Sample	Amount	Dish	Buffer RLT	Disruption and homogenization
Animal cells	<5 x 10^6	<6 cm	350 µl	Add Buffer RLT, vortex (≤1 x 10^8 cells); or use QIAshredder, TissueRuptor®, or needle and syringe
	≤1 x 10^7	6–10 cm	600 µl	
Animal tissues	<20 mg	–	350 µl*	TissueLyser LT; TissueLyser II; TissueRuptor, or mortar and pestle followed by QIAshredder or needle and syringe
	≤30 mg	–	600 µl	

* Use 600 µl Buffer RLT for tissues stabilized in RNA*later*, or for difficult-to-lyse tissues.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual.

"RNA*later*®" is a trademark of AMBION, Inc., Austin, Texas and is covered by various U.S. and foreign patents.

6.3 RNA cleanup protocol

RNA cleanup

Notes before starting

- Add 4 volumes of ethanol (96–100%) to Buffer RPE for a working solution.

1. Adjust the sample to a volume of 100 µl with RNase-free water. Alternatively, follow steps in "DNase digestion of RNA before RNA cleanup" in Appendix E of *RNeasy Mini Handbook*. Add 350 µl Buffer RLT, and mix well.
2. Add 250 µl ethanol (96–100%) to the diluted RNA, and mix well by pipetting. Do not centrifuge. Proceed immediately to step 3.
3. Transfer the sample (700 µl) to an RNeasy Mini spin column placed in a 2 ml collection tube (supplied). Close the lid. Centrifuge for 15 s at ≥8000 x *g*. Discard the flow-through.

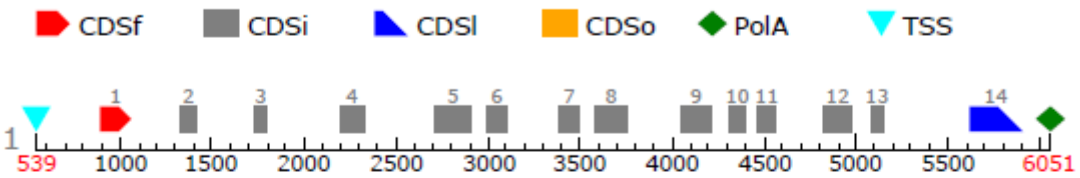
   Optional: If performing optional on-column DNase digestion, follow steps 1–4 of "On-column DNase digestion" (above) after this step.
4. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid. Centrifuge for 15 s at ≥8000 x *g* to wash the membrane. Discard the flow-through.
5. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid. Centrifuge for 2 min at ≥8000 x *g* to wash the membrane.

   Optional: Place the RNeasy spin column in a new 2 ml collection tube (supplied). Close the lid, and centrifuge at full speed for 1 min.
6. Place the RNeasy spin column in a new 1.5 ml collection tube (supplied). Add 30–50 µl RNase-free water directly to the spin column membrane. Close the lid, and centrifuge for 1 min at ≥8000 x *g* to elute the RNA.
7. If the expected RNA yield is >30 µg, repeat step 6 using another 30–50 µl of RNase-free water. Alternatively, use the eluate from step 6 (if high RNA concentration is required). Reuse the collection tube from step 6.

6.4 Fgenesh prediction of annotated genes

PGSC0003DMG400006678

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:48977300..48983500
Length of sequence: 6201
Number of predicted genes 1: in +chain 1, in -chain 0.
Number of predicted exons 14: in +chain 14, in -chain 0.
Positions of predicted genes and exons: Variant 1 from 1, Score:134.438867



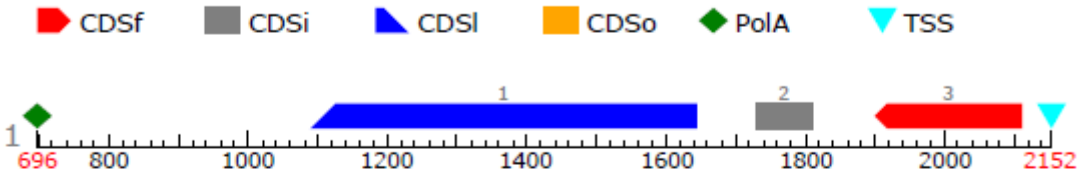
1	+		TSS	539		-6.78			
1	+	1	CDSf	886	1056	16.57	886	1056	171
1	+	2	CDSi	1324	1409	16.30	1324	1407	84
1	+	3	CDSi	1727	1793	-5.08	1728	1793	66
1	+	4	CDSi	2188	2324	6.67	2188	2322	135
1	+	5	CDSi	2700	2906	31.19	2701	2904	204
1	+	6	CDSi	2992	3097	5.97	2993	3097	105
1	+	7	CDSi	3384	3488	11.02	3384	3488	105
1	+	8	CDSi	3578	3750	13.47	3578	3748	171
1	+	9	CDSi	4047	4206	21.39	4048	4206	159
1	+	10	CDSi	4302	4397	6.68	4302	4397	96
1	+	11	CDSi	4463	4557	13.18	4463	4555	93
1	+	12	CDSi	4819	4975	6.49	4820	4975	156
1	+	13	CDSi	5081	5139	-4.41	5081	5137	57
1	+	14	CDSl	5610	5901	16.30	5611	5901	291
1	+		PolA	6051		-4.29			

Predicted protein(s):
>FGENESH:[mRNA] 1 14 exon (s) 886 - 5901 1911 bp, chain +
ATGAACATGTGGAGTAGACATGTTAGATGGCTGTCTACCTCAGTAGCAACAAATAGGAAA
AACAATATGGGATGGTGGGACCATGTGCAGCCAGCTCCTAAAGACCCAATCACCAGTGTC
ACTCAGGCTTTTCTTTCTGACCCAACTCCTCTCAAGCTCAATCTTGGTGTGGGAGCTTAC
AGAGATGATAAAGGAAAACCAGTAACTCTTGAGTGTGTTAGAAGAGCTGCTGAAAAGATT
TCTGGCTGTGAAATCTTAATTTCCTTATCTAATTACTCGAATTTCTCATTTCTTACTTTT
CTGAAGTCTGGCTGGCCATTACATGTAATTAAGTACTTTAGATATACTGTTGGTACACCA
GTCATTCCAGACACCAAGTCGAGGTATCTAGATGTGCACTCTCAAAGTCGTTGGAGTGCT
TACTTGCCAACTGTTTATGTATTATGCCTTTATTTTACTCTGGAATCAACTAAAGCAACA
ACAAAATCAAAGTTTGTGCAGGACTGTGTTAAATTGGCTTATGGAAATGACTCTAGTGTT
GTTAGAGAAGACAGGTTTGCTGGTGTTCCAGCTCTTTCTGGGACTGGTGCGTGCCGGCTG
TTTGCGGAGTTCCAAAGGCGCTTCTATCCTGACTCACAAATGTTTCTACCAATGCCAACT
TGGTCCAACCACCACGACATTTGGAGGGATTCTCAAGTTTGTACAAGAACCTACCACTAT
TATGATCCTGGTACAAAAGGATTGCGATTTCAAGCCCTCGTAAATGATATCAAGAATGCT
CCAGATCGTTCCTTTTTCTTACTTCATCCTTGTGCTCACAACCCAACTGGTGTCGACCCC
AACTATGAGCAGTGGAAAGAAATCTCACACATATTTAAGATCAAGAATCATTTTCCTTTT
TTTGATATGGCTTATCAAGGAATTGCTAGCGGAGATGTTGAGAGAGATGCCACTGCTATT
CGAATATTTCTAGAAGATGGACATCTATTGGGCTGTGCCCAATCTTTTTCCAAAAACATG
GGCTTGTATGAACACCGAGTAGGTTGTGTCAGTATTGTCTCCTGGGATGATAAGCAGGCC

ACAGCAATCAAGAGCCAGCTGCAGCGAATTGTCAGGGCAATGTACAGTAGCTCACCGGTT
CATGGCCCATTATTAGTGTCAACAATCTTGAATGATGCTGATTTAAAGGCACTTTGGGAA
GAGGAAGTTAAGGTCATGGTGGATCGCTTAATCAGTATGAGAATCACATTACGTCAAACT
CTTGAGGAGTTGAACTCATCTTCGAGCTGGGAACATATAACGAAGCAGGTTGGAATGTTT
TACTTCTCTGGTTTATCCCCTGAGGAAGTTAACCACTTGCAAAGGGATTTTCATATATAC
ATGACTAACGATGGGCGCATCAGAAAGCTTGCTCCATCGTATGTCGATTTAGTTGTTTTT
CACAAGAAACTGAAGAAAGAAAACAGAAAGAACCAAAGGATTCTCAAATTTCCTCTTTTT
TTCATGTTTTTTGCAATTGACATGGCTGATGCAACATTGTCATATGGCAGCATGGCAGGG
CAATTCTTATTGAGATCTTGTGCAGATAACCATTGTCTAATACTGTTAGAGGCTGTTCGA
TGCAAATTTAACAAATATCTTGGAACAAAAGAAGTTTGGGACGCGCTTGAGGCTTCACTA
TTTTTCCCTCCAACACTAGAATCCGTTCATCTGCAGCTTCAAAACCTAAAGCAGGATGAT
CCAATTGTCACACAATATCTAGACAAGGCAAAATTGATTTTGGGTGAGTTAACCTCTATT
GGACGTCCTTTATGCCTTGCAAACCAGAACATATGCATATTCAAGGGGTCGGGTTCAGAA
TTTAAAGATATTATCACCGCCTTATTGGCGCGATCTCAACCTGTTACATAG
>FGENESH: 1 14 exon (s) 886 - 5901 636 aa, chain +
MNMWSRHVRWLSTSVATNRKNNMGWWDHVQPAPKDPITSVTQAFLSDPTPLKLNLGVGAY
RDDKGKPVTLECVRRAAEKISGCEILISLSNYSNFSFLTFLKSGWPLHVIKYFRYTVGTP
VIPDTKSRYLDVHSQSRWSAYLPTVYVLCLYFTLESTKATTKSKFVQDCVKLAYGNDSSV
VREDRFAGVPALSGTGACRLFAEFQRRFYPDSQMFLPMPTWSNHHDIWRDSQVCTRTYHY
YDPGTKGLRFQALVNDIKNAPDRSFFLLHPCAHNPTGVDPNYEQWKEISHIFKIKNHFPF
FDMAYQGIASGDVERDATAIRIFLEDGHLLGCAQSFSKNMGLYEHRVGCVSIVSWDDKQA
TAIKSQLQRIVRAMYSSSPVHGPLLVSTILNDADLKALWEEEVKVMVDRLISMRITLRQT
LEELNSSSSWEHITKQVGMFYFSGLSPEEVNHLQRDFHIYMTNDGRIRKLAPSYVDLVVF
HKKLKKENRKNQRILKFPLFFMFFAIDMADATLSYGSMAGQFLLRSCADNHCLILLEAVR
CKFNKYLGTKEVWDALEASLFFPPTLESVHLQLQNLKQDDPIVTQYLDKAKLILGELTSI
GRPLCLANQNICIFKGSGSEFKDIITALLARSQPVT

PGSC0003DMG400006679

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49021600..49025000
Length of sequence: 3401
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 3: in +chain 0, in -chain 3.
Positions of predicted genes and exons: Variant 1 from 1, Score:85.589081

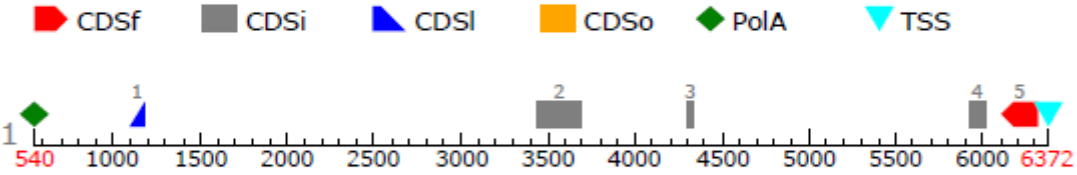


1	-		PolA	696		1.61			
1	-	1	CDSl	1088	- 1643	59.44	1088	- 1642	555
1	-	2	CDSi	1729	- 1808	9.53	1731	- 1808	78
1	-	3	CDSf	1898	- 2110	19.00	1898	- 2110	213
1	-		TSS	2152		-1.88			

Predicted protein(s):
>FGENESH:[mRNA] 1 3 exon (s) 1088 - 2110 849 bp, chain -
ATGGCTTCAAGTTATAGCTTGGTGTTTCTTCATGTGTTAGTCATATTTTCTCTAGCACGC
ATCGCTTTTTCGGATTTGTCAGATGATTTCTACGATGATATTTGTCCCCAAGCTTTACCA
ACCATTAATCGGGTTGTTGAGGATGCAATCAGGCAAGAGAGGCGAATGGGCGCCTCTTTG
CTACGATTACATTTTCATGATTGTTTCGTTAACGGTTGTGACGCTTCAATGGGCGTCCAG
TTGTATCTTGTGCAGACATATTGGCTGTTGCAGCTCGTGACTCCGTTGTTGCTACAGTTA
CATGGACCAACATGGGAAGTGGAACTAGGAAGGAGGGATTCAACTACAGCAAGTAGAACC
ACTGCCGACAATGACATTCCAACTCCATTGATGGACTTACCTGCACTTATCGACAACTTT
AAGAAGCAAGGTTTGGATGAAGAAGACCTCGTTGCTCTCTCTGGTGGCCACACACTAGGG
TTTGCTCAGTGTTCCACCTTTAGGAATCGCATCTACGATGAGACTAACAACATTGACTCA
ACCTTTGCAAGTCAACGTCAAGCAAATTGTCCGCGTAGTGGTGGTGATTCCAATCTTGCT
TCCCTTGATCCTACACCTGCTCTTTTTGACTCCAAATATTTCAGTAACTTGGTGTCTAAG
AAAGGGCTATTGCATTCTGATCAAGCACTGTTTAGCGGAGGAGAGACTGATGAGCTTGTT
AAAACCTATAGTACTAACCTAAGAACTTTCTCGAAAGATTTTTCTGAGTCTATGATTAAG
ATGGGTAATATCAAGCCATTGACTGGAAATGAGGGCCAAATTCGTGTCGATTGCAGGAAA
GTGAACTAA
>FGENESH: 1 3 exon (s) 1088 - 2110 282 aa, chain -
MASSYSLVFLHVLVIFSLARIAFSDLSDDFYDDICPQALPTINRVVEDAIRQERRMGASL
LRLHFHDCFVNGCDASMGVQLYLVQTYWLLQLVTPLLLQLHGPTWEVELGRRDSTTASRT
TADNDIPTPLMDLPALIDNFKKQGLDEEDLVALSGGHTLGFAQCSTFRNRIYDETNNIDS
TFASQRQANCPRSGGDSNLASLDPTPALFDSKYFSNLVSKKGLLHSDQALFSGGETDELV
KTYSTNLRTFSKDFSESMIKMGNIKPLTGNEGQIRVDCRKVN

PGSC0003DMG400006680

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49039700..49047300
Length of sequence: 7601
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 5: in +chain 0, in -chain 5.
Positions of predicted genes and exons: Variant 1 from 1, Score:31.237720

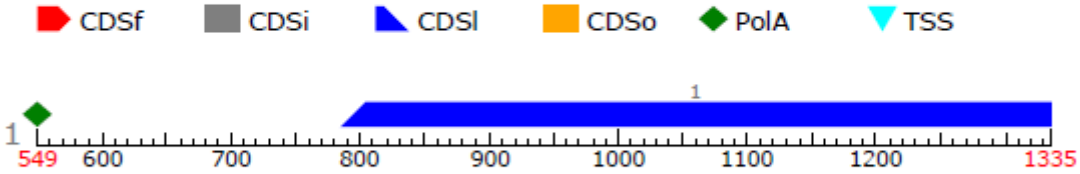


1	-		PolA	540		1.61			
1	-	1	CDSl	1087	1176	4.93	1087	1176	90
1	-	2	CDSi	3431	3687	7.84	3431	3685	255
1	-	3	CDSi	4294	4333	-5.51	4295	4333	39
1	-	4	CDSi	5917	6009	15.69	5917	6009	93
1	-	5	CDSf	6099	6311	22.80	6099	6311	213
1	-		TSS	6372		-1.08			

Predicted protein(s):
>FGENESH:[mRNA] 1 5 exon (s) 1087 - 6311 693 bp, chain -
ATGGCTTCAAGTTATAGCTTGGTGTTTCTTCATGTGTTAGTCATGTTTTGTCTAGCAAAC
ATCGCTTTTTTCGGATTTATCAGATGATTTCTACGATGATATTTGTCCCCAAGCTTTACCA
ACCATTAGAAGGGTTGTTGAGGATGCAGTCAGCCAAGAGAGGCGAATGGGTGCCTCTTTG
CTACGATTACATTTTCATGATTGTTTCGTTAACGGTTGTGACGCTTCAATTCTTCTTGAT
CAAACGGCTACTATTGACAGTGAAAAGACTGCTCGTCCTAATAACAATTCTGCTAGAGGA
TTTGAAGGGAAAAGGGTATATGTGAGCCATTTATATATAAGTAGGGTGAGATTCAAACTG
AACACAAATGGATGCAGCAAAGGAAATCCTGAGAGTGTTGGTGGAAGAGGTTTCCCTAGA
GATGAAAATGGACCGTTAATCATGTCATACTCAGACTACTATGGAGAATGCTCTAATAAT
AATGTTGCAGAATTGAAAGCCATACTCTATGAAATTCAATGGTGCATCACAAATGGATAT
AAAAATGTGGATATTGAATCTGACTCTATGATCATTATTAAAATGACCAACGAGCAAAAT
AATGATACTCATGCTCCTGATATTTTGAGAATTTTAGTTACAAGAAAAGCTATTGGTGAT
CCGGAGGAGATTGAAACATTAGAGGCAACATGA
>FGENESH: 1 5 exon (s) 1087 - 6311 230 aa, chain -
MASSYSLVFLHVLVMFCLANIAFSDLSDDFYDDICPQALPTIRRVVEDAVSQERRMGASL
LRLHFHDCFVNGCDASILLDQTATIDSEKTARPNNNSARGFEGKRVYVSHLYISRVRFKL
NTNGCSKGNPESVGGRGFPRDENGPLIMSYSDYYGECSNNNVAELKAILYEIQWCITNGY
KNVDIESDSMIIIKMTNEQNNDTHAPDILRILVTRKAIGDPEEIETLEAT

PGSC0003DMG400006681

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49063800..49065900
Length of sequence: 2101
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 1: in +chain 0, in -chain 1.
Positions of predicted genes and exons: Variant 1 from 1, Score:52.841321



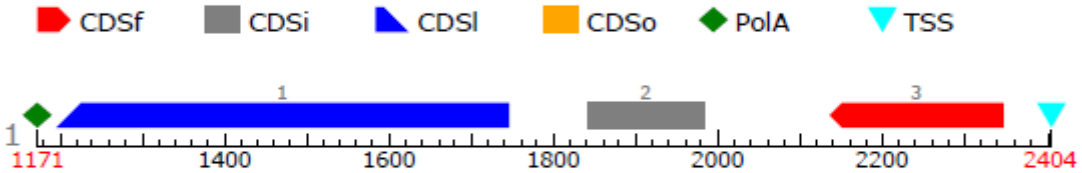

```
1 -      PolA       549              -3.59
1 -    1 CDSl       784 -      1335   66.17       784 -      1335    552
```

Predicted protein(s):

```
>FGENESH:[mRNA] 1 1 exon (s) 784 - 1335 552 bp, chain -
TTACATGGACCAACATGGGAAGTGGAACTAGGAAGGAGGGATTCAACTACAGCAAGTAGA
ACCACTGCCGACAATGACATTCCAACTCCATTGATGGACTTACCTGCACTTATCGACAAC
TTTAAGAAGCAAGGTTTGGATGAAGAAGACCTCGTTGCTCTCTCTGGTGGCCACACACTA
GGGTTTGCTCAGTGTTCCACCTTTAGGAATCGCATTTACGATGAGACTAACAACATCGAC
TCAACCTTTGCAAGTCAACGTCAAGCAAATTGTCCGCGTAGTGGTGGTGATTCCAATCTT
GCTTCCCTTGATCCTACATCTGCTCTTTTTGACTCCAAATATTTCAGTAACTTGGTGTCT
AAGAAAGGGCTATTGCATTCCGATCAAGCACTGTTTAGCGGAGGAGAGACTGATGAGCTT
GTTAAAACCTATAGTACGAACCTAAGAACTTTCTCGAAAGATTTTGCTGAGTCTATGATT
AAGATGGGTAATATCAAGCCATTGACTGGAAATGAAGGCCAAATTCGTGTCGATTGCAGG
AAGGTGAACTAA
>FGENESH: 1 1 exon (s) 784 - 1335 183 aa, chain -
LHGPTWEVELGRRDSTTASRTTADNDIPTPLMDLPALIDNFKKQGLDEEDLVALSGGHTL
GFAQCSTFRNRIYDETNNIDSTFASQRQANCPRSGGDSNLASLDPTSALFDSKYFSNLVS
KKGLLHSDQALFSGGETDELVKTYSTNLRTFSKDFAESMIKMGNIKPLTGNEGQIRVDCR
KVN
```

PGSC0003DMG400020800

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49083000..49086200
Length of sequence: 3201
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 3: in +chain 0, in -chain 3.
Positions of predicted genes and exons: Variant 1 from 1, Score:87.659674

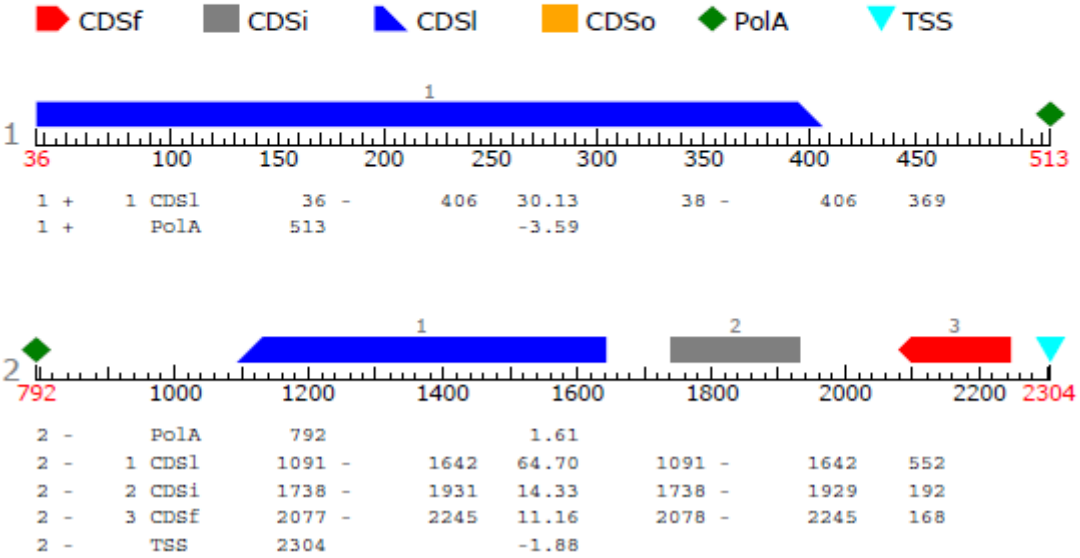


1	-		PolA	1171		1.61			
1	-	1	CDSl	1194	- 1745	61.07	1194	- 1745	552
1	-	2	CDSi	1841	- 1981	4.69	1841	- 1981	141
1	-	3	CDSf	2134	- 2346	23.60	2134	- 2346	213
1	-		TSS	2404		-2.38			

Predicted protein(s):
>FGENESH:[mRNA] 1 3 exon (s) 1194 - 2346 906 bp, chain -
ATGGCTTCAAGTTATAGCTTCATGCTTCTTCATATGTTAGTCATGTTTAGTCTAGCAAGT
ATCGCTTTTTCGGATTTGTCAGATGATTTCTACGATGATATTTGTCCCCAAGCTTTGCCA
ACCATTAGACGGGTTGTTGAGGATGCAGTCAGGCAAGAGAGGCGAATGGGTGCCTCTTTG
CTACGATTACATTTTCATGATTGTTTCGTTAACACTGCTCGTCCTAATAACAATTCTGCT
AGAGGATGTGAGGTGATCGATAGAATTAAATCGGAGGTTGATAGAGTATGTGGAGGTCCA
GTTGTATCTTGTGCGGACATCTTGGCTGTTGCAGCTCGTGACTCCGTTGTTGCTTTACAT
GGACCAACATGGGAAGTGGAACTAGGAAGAAGGGATTCGACTACAGCAAGTAGAACCACT
GCCAATAATGACATTCCAACTCCATTGATGGACTTACCTGCACTTATCGACAACTTTAAG
AAGCAAGGTTTGGATGAGGAAGACCTCGTTGCTCTCTCCGGTGCCCACACTCTAGGGTTT
GCTCAGTGTTCCACCTTTAGGAATCGCATCTACAATGAGATTACCAACATCGACTCCACC
TTTGCAAGTCAACGTCAAGCAAATTGTCCGCGTAGTGGAGGTGATTCCAATCTTGCTTCC
CTTGATCCTACACCTGCTCTTTTCGACTCCAAATATTTCAGTAACTTGGTTTCCAAGAAA
GGGCTTTTGCATTCTGATCAAGCACTGTTTAGTGGAGGAGAGACTGATGAGCTTGTTAAA
ACCTATAGTACGAACCTAAGGACTTTTTCGAAAGATTTTGCTAAGTCTATGATTAAGATG
GGCAATATCAAACTATTGACCGGAAATCAAGGCCAGATTCGTGTCGATTGCAGGAAAGTG
AACTAA
>FGENESH: 1 3 exon (s) 1194 - 2346 301 aa, chain -
MASSYSFMLLHMLVMFSLASIAFSDLSDDFYDDICPQALPTIRRVVEDAVRQERRMGASL
LRLHFHDCFVNTARPNNNSARGCEVIDRIKSEVDRVCGGPVVSCADILAVAARDSVVALH
GPTWEVELGRRDSTTASRTTANNDIPTPLMDLPALIDNFKKQGLDEEDLVALSGAHTLGF
AQCSTFRNRIYNEITNIDSTFASQRQANCPRSGGDSNLASLDPTPALFDSKYFSNLVSKK
GLLHSDQALFSGGETDELVKTYSTNLRTFSKDFAKSMIKMGNIKLLTGNQGQIRVDCRKV
N

PGSC0003DMG400020799

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49150500..49153300
Length of sequence: 2801
Number of predicted genes 2: in +chain 1, in -chain 1.
Number of predicted exons 4: in +chain 1, in -chain 3.
Positions of predicted genes and exons: Variant 1 from 1, Score:106.149902



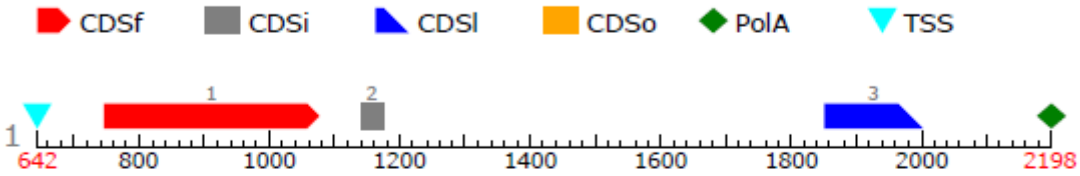
1	+	1	CDSl	36	-	406	30.13	38 -	406	369
1	+		PolA	513			-3.59			
2	-		PolA	792			1.61			
2	-	1	CDSl	1091	-	1642	64.70	1091 -	1642	552
2	-	2	CDSi	1738	-	1931	14.33	1738 -	1929	192
2	-	3	CDSf	2077	-	2245	11.16	2078 -	2245	168
2	-		TSS	2304			-1.88			

Predicted protein(s):
>FGENESH:[mRNA] 1 1 exon (s) 36 - 406 369 bp, chain +
GAGCTGACTTGGCTCAATCGACTTCTTCTTGATCTTGATTATCCACTGGAGCTTCCCATT
CCTGTTCATTCTGATAGTCAATCAGCAATTCACACTGCCCGAAACCCCATCTTTCATGAG
AGAACCAAGCAAGTGGAGCTCGACTGCCATTTTGTCAGGCAACAATTCCTTGCTGGTCTC
ATTTCACTCTCTTATGTACCGGCCTCCTCCCAGCTTGTCGATCTTTTTACTAAGGCACTA
TTTGGACCTTCTCATCACACCATTCTTCCCAAGTTGGGTCTCTCTGCACTGCCCTTCAAC
TTGGGGGAAGGGGTGTTGAAAATGGTGTTTCGTCGATCTCACAAAAGAACAAAGAATTGG
AGAAGATGA
>FGENESH: 1 1 exon (s) 36 - 406 122 aa, chain +
ELTWLNRLLLDLDYPLELPIPVHSDSQSAIHTARNPIFHERTKQVELDCHFVRQQFLAGL
ISLSYVPASSQLVDLFTKALFGPSHHTILPKLGLSALPFNLGEGVLKMVFRRSHKRTKNW
RR
>FGENESH:[mRNA] 2 3 exon (s) 1091 - 2245 915 bp, chain -
ATGGCTTCAAGTTATAGCTTCCTGTTTCTTCATATGTTAGTCATGTTTAGTCTAGCAAGT
ATCGCTTTTTCGGATTTGTCAGATGATTTCTACGATGATATTTGTCCCCAAGCTTTGCCA
ACCATTAGACGGGTTGTTGAGGATACAGTCAGGCAAGAGAGGCGAATGGGTTGTGACGCT
TCCATTCTTCTTGATCAAACGGCTACTATTGACAGTGAAAAGACTGCTCGTCCTAATAAC
AATTCTGCTAGAGGATTTGAGGTGATCGATATAATTAAATCGGAGGTTGATAGAGTATGT
GGACGTCCAGTTGTATCTTGTGCGGACATCTTGGCTGTTGCAGCTCGTGACTCCGTTGTT
GCTTTACATGGACCAACATGGGAAGTGGAACTAGGAAGAAGGGATTCGACTACAGCAAGT
AGAACCACTGCCAATAATGACATTCCAACTCCATTGATGGACTTACCTGCACTTATCGAC

AACTTTAAGAAGCAAGGTTTGGATGAGGAAGACCTCGTTGCTCTCTCCGGTGCCCACACT
CTAGGGTTTGCTCAGTGTTCCACCTTTAGGAATCGCATCTACAATGAGATTAACAACATC
GACTCCACCTTTGCAAGTCAACGTCAAGCAAATTGTCCGCGTAGTGGAGGTGATTCCAAT
CTTGCTTCCCTTGATCCTACATCTGCTCTTTTCGACTCCAAATATTTCAGTAACTTGGTT
TCCAAGAAAGGGCTTTTGCATTCTGATCAAGCACTCTTTAGTGGAGGAGAGACTGATGAG
CTTGTTAAAACCTATAGTACGAACCTAAGGACTTTTTCGAAAGATTTTGCTAAGTCTATG
ATTAAGATGGGTAATATCAAACTATTGACCGGAAATCAAGGCCAAATTCGTGTCGACTGC
AGGAAAGTGAACTAA
>FGENESH: 2 3 exon (s) 1091 - 2245 304 aa, chain -
MASSYSFLFLHMLVMFSLASIAFSDLSDDFYDDICPQALPTIRRVVEDTVRQERRMGCDA
SILLDQTATIDSEKTARPNNNSARGFEVIDIIKSEVDRVCGRPVVSCADILAVAARDSVV
ALHGPTWEVELGRRDSTTASRTTANNDIPTPLMDLPALIDNFKKQGLDEEDLVALSGAHT
LGFAQCSTFRNRIYNEINNIDSTFASQRQANCPRSGGDSNLASLDPTSALFDSKYFSNLV
SKKGLLHSDQALFSGGETDELVKTYSTNLRTFSKDFAKSMIKMGNIKLLTGNQGQIRVDC
RKVN

PGSC0003DMG400040954

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49169800..49172200
Length of sequence: 2401
Number of predicted genes 1: in +chain 1, in -chain 0.
Number of predicted exons 3: in +chain 3, in -chain 0.
Positions of predicted genes and exons: Variant 1 from 1, Score:27.565140

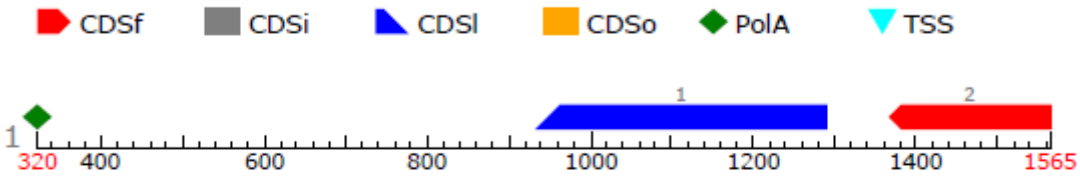


1	+		TSS	642			-1.78		
1	+	1	CDSf	744	-	1075	28.61	744 - 1073	330
1	+	2	CDSi	1139	-	1172	-0.64	1140 - 1172	33
1	+	3	CDSl	1849	-	2001	4.89	1849 - 2001	153
1	+		PolA	2198			-0.59		

Predicted protein(s):
>FGENESH:[mRNA] 1 3 exon (s) 744 - 2001 519 bp, chain +
ATGGCTGGGAAAATTGCATGTTTTGTGGTTTTGTGCATGGTGATGGTTGCACCCTATGCA
TCGGAGGCGTTTAGCTGCGGTCAAGTTCAGTCCGGGGTGGTTAAGTGTCTCCCTTATTTG
CAAAATCGCGGCCCTGTGGGACAGTGTTGTGATGTTGTTAAAGGTCTCCTGAATTCTGCT
AAGACAACACAGGATCGTAGAACTACATGCTCTTGCCTTAAATCAGCTGCTTCTATTATA
AAGGGCATTGATATGAGCAAAGCTGCTGGTCTCCCTGCTCTTTGTGGTGTCAAAAGTCCC
TTCAAGATTAGCCTCTCCACTGACTGCACTAATTTAACAAAGACTTTTGAAACATTTTGT
CTAGAGGGTCCAGTGAGAGATGAAAAATCGATTTATCGACGCATAAAATTGGTGTTAGAC
GTGCTATCATTGTTGCGAAGAATAAGATGCATATGGATCGAGTTTGATCCTGCTTTATGT
CTTGTTGTCTTCTTTTTTTCGGTCATAGTGGCGTTGTAA
>FGENESH: 1 3 exon (s) 744 - 2001 172 aa, chain +
MAGKIACFVVLCMVMVAPYASEAFSCGQVQSGVVKCLPYLQNRGPVGQCCDVVKGLLNSA
KTTQDRRTTCSCLKSAASIIKGIDMSKAAGLPALCGVKSPFKISLSTDCTNLTKTFETFC
LEGPVRDEKSIYRRIKLVLDVLSLLRRIRCIWIEFDPALCLVVFFFSVIVAL

PGSC0003DMG400020798

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49171700..49173500
Length of sequence: 1801
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 2: in +chain 0, in -chain 2.
Positions of predicted genes and exons: Variant 1 from 1, Score:50.032358



1	-		PolA	320			-0.59		
1	-	1	CDSl	931	-	1290	36.84	931 - 1290	360
1	-	2	CDSf	1365	-	1565	23.84	1365 - 1565	201

Predicted protein(s):

>FGENESH:[mRNA] 1 2 exon (s) 931 - 1565 561 bp, chain -

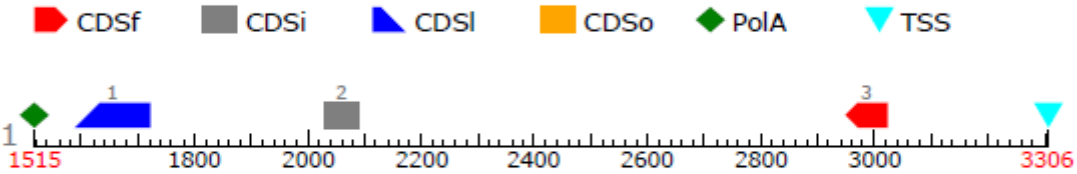
ATGTCGGGTTGTGACGCTTCCATTCTTCTTGATCAAACGGCTACTATTGACTGTGAAAAG
ACTGCTCGTCCTAATAACAATTCTGCTAGAGGATGTGAGGTGATCGATAGAATTAAATCG
GAGGTTGATAGAGTATGTGGAGGTCCAGTTGTATCTTGTGCGGACATCTTGGCTGTTGCA
GCTCGTGACTCCGTTGTTGCTTATTTTGTTACATTGAGACAGTTACATGGACCAACATGG
GAAGTGGAACTAGGAAGAAGGGATTCGACTACAGCAAGTAGAACCACTGCCAATAATGAC
ATTCCAACTCCATTGATGGACTTACCTGCACTTATCGACAACTTTAAGAAGCAAGGTTTG
GATGAAGAAGACCTCGTTGCTCTCTCCGGTGCCCACACTCTAGGGTTTGCTCAGTGTTCC
ACCTTTAGGAATCGCATTCCACCTTTAGGAATCGCATCTACAATGAGACTAACAACATTG
ACTCAACCTTTGCAAGTCAGTCCGCATAGTGGTGGTGATTCCAATCTTGCTTCCCTTGAT
CCTACACCTTTTTTACTCTAA

>FGENESH: 1 2 exon (s) 931 - 1565 186 aa, chain -

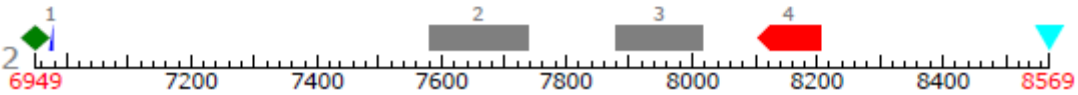
MSGCDASILLDQTATIDCEKTARPNNNSARGCEVIDRIKSEVDRVCGGPVVSCADILAVA
ARDSVVAYFVTLRQLHGPTWEVELGRRDSTTASRTTANNDIPTPLMDLPALIDNFKKQGL
DEEDLVALSGAHTLGFAQCSTFRNRIPPLGIASTMRLTTLTQPLQVSPHSGGDSNLASLD
PTPFLL

PGSC0003DMG400020797

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49195600..49204800
Length of sequence: 9201
Number of predicted genes 2: in +chain 0, in -chain 2.
Number of predicted exons 7: in +chain 0, in -chain 7.
Positions of predicted genes and exons: Variant 1 from 1, Score:10.462855



1	-		PolA	1515		-0.59			
1	-	1	CDSl	1586	- 1720	5.31	1586	- 1720	135
1	-	2	CDSi	2029	- 2087	0.91	2029	- 2085	57
1	-	3	CDSf	2947	- 3022	5.14	2948	- 3022	75
1	-		TSS	3306		-4.28			



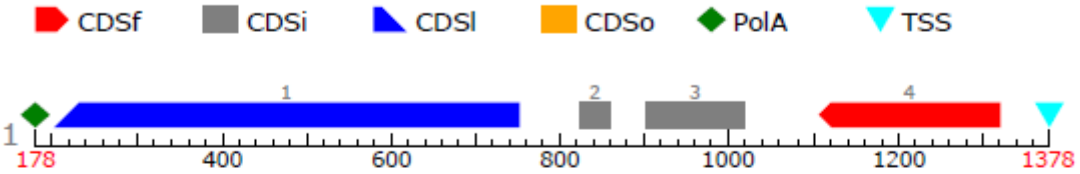
2	-		PolA	6949		1.61			
2	-	1	CDSl	6971	- 6976	-2.13	6971	- 6976	6
2	-	2	CDSi	7578	- 7736	11.24	7578	- 7736	159
2	-	3	CDSi	7878	- 8014	3.25	7878	- 8012	135
2	-	4	CDSf	8102	- 8204	3.92	8103	- 8204	102
2	-		TSS	8569		-6.08			

Predicted protein(s):
>FGENESH:[mRNA] 1 3 exon (s) 1586 - 3022 270 bp, chain -
ATGGCGGTAACTCTACCTAGCAGACTGGAGTTCCTAAGAGCTAGGTTCAAGGAGAAAAAC
AAAGTTGTGACTAATGTGTACTGTACTGCTGTTGAGTGGTTCGCTGACACCGTGATTTTA
GGTACCGACACACCGCTAAATTTGGATTTGCGTCGGGAAATCCCACATTATATTGGGGAT
AAAGCGTTCCTTAACAAAGGCGACTCCGTACCCAAAGGGACTCAAACCCGAGATCTCTTG
GTTAAGGATGTTTTTTTGTGGTACATGTAA
>FGENESH: 1 3 exon (s) 1586 - 3022 89 aa, chain -
MAVTLPSRLEFLRARFKEKNKVVTNVYCTAVEWFADTVILGTDTPLNLDLRREIPHYIGD
KAFLNKGDSVPKGTQTRDLLVKDVFLWYM
>FGENESH:[mRNA] 2 4 exon (s) 6971 - 8204 405 bp, chain -
ATGTTGATCATCAGTAGAGACACTAATGACATAAATGCTACTAAACATATGCTAGAAAGC
AAGTTTGATATGAAAGATCTTGGAGTTGCTGAAGTGATCTTAGATTTCAATATTGTCAAG
ACTCCAATAAATGTGAGCTTTGCATTTCAAATGAGTGAAGGCGAAAGTGACTCACAATTG
AATTGTGCTAGAGTATTGGGAAGTATGATGTATATCAAGAAGTGTATGCGATCAAATATA
GATATAGTGATGAAAATTGGATCACCGGGATCAAATGAAGTAAAATCTACGAGTGGATAT
GTATTTACTCTTGGTGGAGGAGCAGTCTCTTGGAAATCTACCAAAAAGACATGTATCGCT
AGTTCTACAATGATCTCTGAGCTAGGTGCTCTAGATAAGTTATAA
>FGENESH: 2 4 exon (s) 6971 - 8204 134 aa, chain -

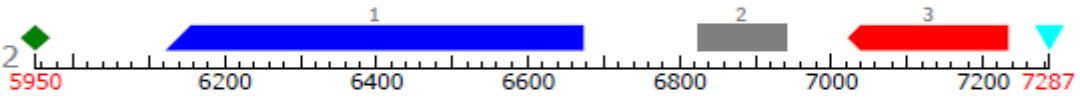
MLIISRDTNDINATKHMLESKFDMKDLGVAEVILDFNIVKTPINVSFAFQMSEGESDSQL
NCARVLGSMMYIKKCMRSNIDIVMKIGSPGSNEVKSTSGYVFTLGGGAVSWKSTKKTCIA
SSTMISELGALDKL

PGSC0003DMG400020795A (178-1378) and PGSC0003DMG400020795B (5950-7287)

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49247100..49254500
Length of sequence: 7401
Number of predicted genes 2: in +chain 0, in -chain 2.
Number of predicted exons 7: in +chain 0, in -chain 7.
Positions of predicted genes and exons: Variant 1 from 1, Score:215.549951



1	-		PolA	178		1.61		
1	-	1	CDSl	200	- 751	63.67	200 - 751	552
1	-	2	CDSi	822	- 859	13.80	822 - 857	36
1	-	3	CDSi	900	- 1017	12.79	901 - 1017	117
1	-	4	CDSf	1105	- 1320	17.76	1105 - 1320	216
1	-		TSS	1378		-1.68		



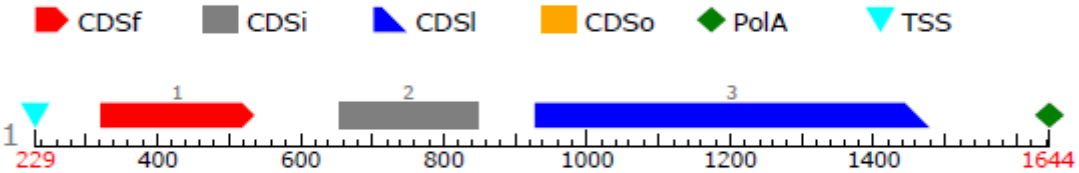
2	-		PolA	5950		1.61		
2	-	1	CDSl	6122	- 6673	67.17	6122 - 6673	552
2	-	2	CDSi	6823	- 6939	21.12	6823 - 6939	117
2	-	3	CDSf	7021	- 7233	21.14	7021 - 7233	213
2	-		TSS	7287		-1.28		

Predicted protein(s):
>FGENESH:[mRNA] 1 4 exon (s) 200 - 1320 924 bp, chain -
ATGGCTTTGAAATATAGCTTCTTTGTTCTTCATGTTTTAGTCATGTTTTCTCTAGCAAGC
ATGGCACTTTCAGATTCTCTTTCACCTTCTTTCTACAATCATGTTTGTCCCCAAGCTGTG
CCAGCCATAAAACGAGTTGTTGAGGATGCAGTCAGGAAAGAGAGGCGAATGGGTGCCTCT
TTGCTACGTTTACATTTTCATGATTGTTTTGTCAATGGGTGTGACGCTTCAATTCTTCTA
GACAAAACGGCTACTATTGACAGTGAAAAGACTGCAATACCTAATAACAATTCTATTAGA
GGATTTGATGTGATTGACAAAATCAAAGTCGGAGACATTGTGGCTGTTGCAGCTCGCGAC
TCTGTAGTTGCATTAGGTGGACCAACATGGGAAGTTCCACTAGGAAGAAGGGATTCTACT
ACAGCAAGTAGAACCACAGCAAACAATGATATCCCACCTCCATTTTTGGACTTACCTGCA
CTTATCAACAACTTCAAGAAGCAAGGATTGGATGAGAAAGATCTCGTCGCTCTCTCTGGT
GGCCACACTCTAGGATTTGCTCAGTGTTCCACCTTCAGGAATCGTATCTACAATGACACT
AACATCGACTCCACTTTTGCAAGCCAACGCAAAGCCAATTGTTCACGCAGTGGAGGTAAT
ACCAACCTTGCTCCACTCGATCCAACTCCAGCTCTTTTTGACTCAAAATATTTTAGTAAC
TTGGTGTCCAAAAAAGGACTTTTGCACTCTGATCAAGCACTATTTAATGGCGGTCAAACT
GATAATCTTGTTAAGAAATATAGTACCAGCCTCGGAAGTTTCTCTAAAGATTTTGCTGAG
TCTATGATTAAGATGGGAAATATCAAGCCATTGACAGGGAATCAAGGCCAAATTCGCGTA
AACTGCAGGAAGGTGAACGTCTAA
>FGENESH: 1 4 exon (s) 200 - 1320 307 aa, chain -

MALKYSFFVLHVLVMFSLASMALSDSLSPSFYNHVCPQAVPAIKRVVEDAVRKERRMGAS
LLRLHFHDCFVNGCDASILLDKTATIDSEKTAIPNNNSIRGFDVIDKIKVGDIVAVAARD
SVVALGGPTWEVPLGRRDSTTASRTTANNDIPPPFLDLPALINNFKKQGLDEKDLVALSG
GHTLGFAQCSTFRNRIYNDTNIDSTFASQRKANCSRSGGNTNLAPLDPTPALFDSKYFSN
LVSKKGLLHSDQALFNGGQTDNLVKKYSTSLGSFSKDFAESMIKMGNIKPLTGNQGQIRV
NCRKVNV
>FGENESH:[mRNA] 2 3 exon (s) 6122 - 7233 882 bp, chain -
ATGGCTTCGAATAGCTTCATTTTTCTTCATGTTCTAGTCATGTTTTCTCTAGCAAGCATG
GCACTTTCGGATTCTCTTTCACCATCTTTCTACAATCATGTTTGTCCCCAAGCTTTGCCT
GCCATAAAACGAGTCGTTGAAGATGCAGTCAGGAAAGAGAGGCGAATGGGTGCCTCTTTG
CTACGTTTACATTTTCATGATTGTTTCGTTAATGGGTGTGACGCTTCAATTCTTCTAGAC
AAAACAGCTACTATTGACAGTGAAAAGACTGCGATACCTAATAACAATTCTATTAGAGGA
TTTGATGTGATTGACAAAATCAAGTCGGAGTTAGGTGGACCAACATGGGAAGTTCCACTA
GGAAGAAGGGACTCTACAACAGCAAGTAGAACCACAGCAAACAATGATATCCCACCTCCA
TTTTTGGACTTACCAGCACTTATCAACAACTTCAAGAAGCAAGGATTGGATGAGAAAGAT
CTCGTTGCTCTCTCTGGAGGACACACACTAGGATTTGCTCAGTGTTTTACTTTCAGGAAT
CGCATCTACAATGACACTAACATTGACTCTACCTTTGCAAAGCAACTCCAAACCAATTGT
CCACGTAGTGGAGGTGATTCCAAGCTTGCTCCACTTGATCCAACTCCAGCTCTTTTCGAC
TCAAAATATTTTAGTAACTTGATGTCCAAGAAAGGACTTCTACACTCTGATCAAGCACTA
TTTAATGGCGGTCAAACTGATAATCTTGTTAAGAAATATAGTATCAGCCTCGGGAGTTTC
TCCAAAGATTTTGTTGAATCTATGATTAAGATGGGAAATATCAAGCCATTGACAGGGAAA
CAAGGCCAAATTCGCGTAAACTGCAGGAAGGTGAACGTCTAA
>FGENESH: 2 3 exon (s) 6122 - 7233 293 aa, chain -
MASNSFIFLHVLVMFSLASMALSDSLSPSFYNHVCPQALPAIKRVVEDAVRKERRMGASL
LRLHFHDCFVNGCDASILLDKTATIDSEKTAIPNNNSIRGFDVIDKIKSELGGPTWEVPL
GRRDSTTASRTTANNDIPPPFLDLPALINNFKKQGLDEKDLVALSGGHTLGFAQCFTFRN
RIYNDTNIDSTFAKQLQTNCPRSGGDSKLAPLDPTPALFDSKYFSNLMSKKGLLHSDQAL
FNGGQTDNLVKKYSISLGSFSKDFVESMIKMGNIKPLTGKQGQIRVNCRKVNV

PGSC0003DMG400020801

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49258400..49260800
Length of sequence: 2401
Number of predicted genes 1: in +chain 1, in -chain 0.
Number of predicted exons 3: in +chain 3, in -chain 0.
Positions of predicted genes and exons: Variant 1 from 1, Score:90.191656



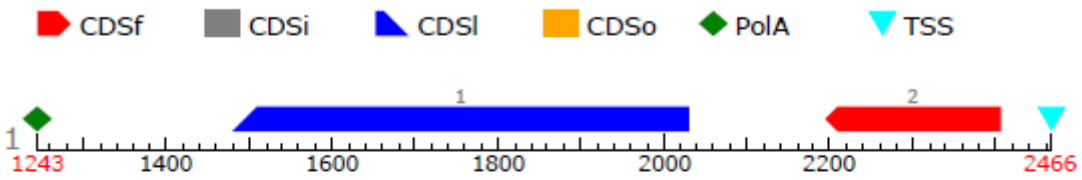
1	+		TSS	229		-1.68			
1	+	1	CDSf	320	- 535	7.85	320	- 535	216
1	+	2	CDSi	653	- 847	33.14	653	- 847	195
1	+	3	CDSl	926	- 1477	50.65	926	- 1477	552
1	+		PolA	1644		1.61			

Predicted protein(s):
>FGENESH:[mRNA] 1 3 exon (s) 320 - 1477 963 bp, chain +
ATGGCTTTGAAATATAGCTTCTTTGTTCTTCATGTTTTAGTCATGTTTTCTCTAGCAAGC
ATGGCACTTTCAGATTCTCTTTCACCTTCTTTCTACAATCATGTTTGTCCCCAAGCTTTG
CCAGCCATAAAACGAGTTGTTGAGGCTGCAGTCAGGAAAGAGAGGCGAATGGGTGCCTCT
TTGCTACGTTTACATTTTCATGATTGTTTCGTCAATGGGTGTGACGCTTCAATTCTTCTA
GACAAAACGGCTACTATTGACAGTGAAAAGACTGCAATACCTAATAACAATTCTATTAGA
GGATTTGATGTTATTGATAAAATCAAGTCGGAGGTTGATAAATGTTGTGGACGTTCTATT
GTGTCTTGTGCAGACATTGTGGCTGTTGCAGCTCGCGACTCTGTAGTTGCACTAGGTGGA
CCAACATGGGAAGTTCCACTAGGAAGAAGGGATGCTACTACAGCAAGCAGAACCACAGCA
AACAATGATATCCCACCTCCATTTTTGGACTTACCTGCGCTTATCAACAACTTCAAGAAG
CAAGGATTGGATGAGAAAGATCTCGTCGCTCTCTCTGGTGGCCACACTCTAGGATTTGCT
CAATGTTCCACCTTCAGGAATCGTATCTACAATGACACTAACATCGAATCCACTTTTGCA
AGCCAACGCAAAGCCAATTGTCCACGCAATGGAGGCAATACCAACCTTGCTCCACTCGAT
CCAACTCCAGCTCTTTTTGACTCAAAATATTTTAGTAACTTGTTGTCCAAGAAAGGACTT
TTGCACTCTGATCAAGCACTATTTAATGGCGGTCAAACTGATAATCTTGTTAAGAAATAT
AGTACCAACCTCGGGAGTTTCTCTAAAGATTTTGCCGAGTCTATGATTAAGATGGGAAAT
ATCAAGCCATTGACAGGGAATCAAGGCCAAATTCGCGTAAACTGCAGGAAGGTGAACGTC
TAA
>FGENESH: 1 3 exon (s) 320 - 1477 320 aa, chain +
MALKYSFFVLHVLVMFSLASMALSDSLSPSFYNHVCPQALPAIKRVVEAAVRKERRMGAS
LLRLHFHDCFVNGCDASILLDKTATIDSEKTAIPNNNSIRGFDVIDKIKSEVDKCCGRSI
VSCADIVAVAARDSVVALGGPTWEVPLGRRDATTASRTTANNDIPPPFLDLPALINNFKK
QGLDEKDLVALSGGHTLGFAQCSTFRNRIYNDTNIESTFASQRKANCPRNGGNTNLAPLD
PTPALFDSKYFSNLLSKKGLLHSDQALFNGGQTDNLVKKYSTNLGSFSKDFAESMIKMGN
IKPLTGNQGQIRVNCRKVNV

6.5 Fgenesh predictions of unannotated peroxidase genes

A.

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49073400..49077300
Length of sequence: 3901
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 2: in +chain 0, in -chain 2.
Positions of predicted genes and exons: Variant 1 from 1, Score:60.447363

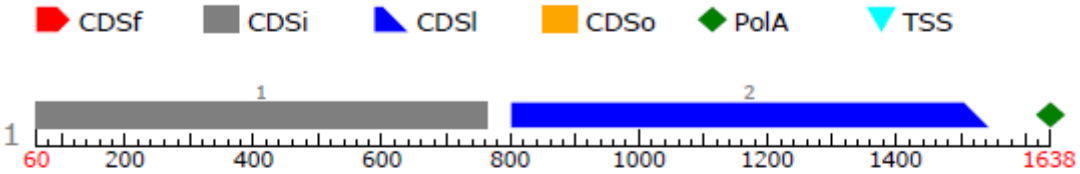


1	-		PolA	1243			-3.59		
1	-	1	CDSl	1478	-	2029	52.85	1478 - 2029	552
1	-	2	CDSf	2193	-	2405	13.80	2193 - 2405	213
1	-		TSS	2466			-1.08		

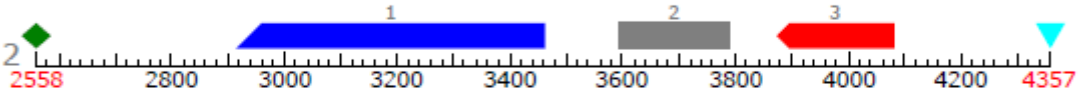
Predicted protein(s):
>FGENESH:[mRNA] 1 2 exon (s) 1478 - 2405 765 bp, chain -
ATGGCTTCAAGTTATAGCTTGGTGTTTCTTCATGTGTTAGTCATGTTTTGTCTAGCAAAC
ATCGCTTTTTCGGATTTATCAGATGATTTCTACGATGATATTTGTCCCCAAGCTTTACCA
ACCATTAGAAGGGTTGTTGAGGATGCAGTCAGCCAAGAGAGGCGAATGGGCGCCTCTTTG
CTACGATTACATTTTCATGATTGTTTTGTTAACTTACATGGACCAACATGGGAAGTGGAA
CTAGGAAGGAGGGATTCAACTACAGCAAGTAGAACCACTGCCGACAATGACATTCCAACT
CCATTGATGGACTTACCTGCACTTATCGACAACTTTAAGAAGCAAGGTTTGGATGAAGAA
GACCTCGTTGCTCTCTCTGGTGGCCACACACTAGGGTTTGCTCAGTGTTCCACCTTTAGG
AATCGCATTTACGATGAGACTAACAACATCGACTCAACCTTTGCAAGTCAACGTCAAGCA
AATTGTCCGCGTAGTGGTGGTGATTCCAATCTTGCTTCCCTTGATCCTACATCTGCTCTT
TTTGACTCCAAATATTTCACTAACTTGGTGTCTAAGAAAGGGCTATTGCATTCCGATCAA
GCACTGTTTAGCGGAGGAGAGACTGATGAGCTTGTTAAAACCTATAGTACGAACCTAAGA
ACTTTCTCGAAAGATTTTGCTGAGTCTATGATTAAGATGGGTAATATCAAGCCATTGACT
GGAAATGAAGGCCAAATTCGTGTCGATTGCAGGAAGGTGAACTAA
>FGENESH: 1 2 exon (s) 1478 - 2405 254 aa, chain -
MASSYSLVFLHVLVMFCLANIAFSDLSDDFYDDICPQALPTIRRVVEDAVSQERRMGASL
LRLHFHDCFVNLHGPTWEVELGRRDSTTASRTTADNDIPTPLMDLPALIDNFKKQGLDEE
DLVALSGGHTLGFAQCSTFRNRIYDETNNIDSTFASQRQANCPRSGGDSNLASLDPTSAL
FDSKYFTNLVSKKGLLHSDQALFSGGETDELVKTYSTNLRTFSKDFAESMIKMGNIKPLT
GNEGQIRVDCRKVN

B.

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49215600..49223200
Length of sequence: 7601
Number of predicted genes 2: in +chain 1, in -chain 1.
Number of predicted exons 5: in +chain 2, in -chain 3.
Positions of predicted genes and exons: Variant 1 from 1, Score:229.536450



1	+	1	CDSi	60	-	760	63.94	62 - 760	699
1	+	2	CDSl	799	-	1542	65.80	799 - 1542	744
1	+		PolA	1638			-3.99		



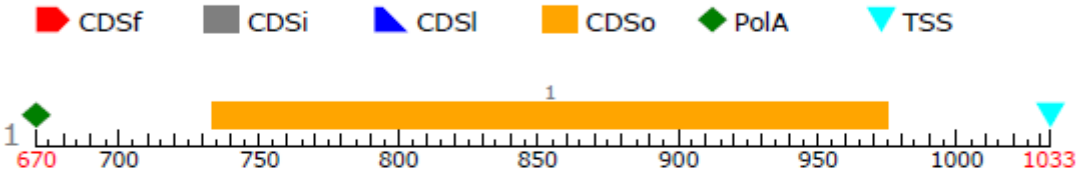
2	-		PolA	2558			1.61		
2	-	1	CDSl	2913	-	3461	74.57	2913 - 3461	549
2	-	2	CDSi	3593	-	3787	26.82	3593 - 3787	195
2	-	3	CDSf	3871	-	4080	16.17	3871 - 4080	210
2	-		TSS	4357			-4.28		

Predicted protein(s):
>FGENESH:[mRNA] 1 2 exon (s) 60 - 1542 1443 bp, chain +
AAGACTGCAATTGGTTGTAGGTGGATCTACAAAATAAAATACAAGTCTATTGGTGAAGTA
GAGAGGTATAAGGCAAGGTTAGTGGCAAAAGGATACAGTCAGAAAGAGGGGATAGATTAC
AAAGAAACTTTTTCACCTGTGGTAAAGATGGTGACTGTAAGAACTATCCTAGCTTTAGCT
GTGGCAAGGAAATGGCACATTCATCAAATGGATGTGTATAATGCATTTCTAAATGGTGAT
CTACAGGATGAAATATACATGACCTTGCCACAAGGATTTACCAGTCAGGGGGAGAAGAAA
GTATGCAGACTGGTAAAGTCTCTATATGGTCTTAAACAAGCACCTAGACAATGGAATCAA
AAGCTTAAAGAGGCATTACAAAAACTCAGTTTCAGTCAAAGTCAGCATGATCATTCTCTG
TTTACAAGAAAAACCTCTAAAGGAACTACAATAATTTTAGTATATGTTGATGATATCTTG
GTTACAGGAAGTAGTCTTGAGCTGATTAAGGAAACAAAAGAAGCCTTACAACAAGTATTC
AAGATGAAGGATTTGGGAGAATTAAGATTTTTTCTTAGAATAGAGTTTGCAAGGTGTGAG
GCTGGAATGGTTATGCATCAAAGGAAATATGCACTTCAACTGATTGCTGAAGTTGGAGTA
TCTGGTGCTAAGCCATCAGGCACTCCTATGGATGTGAATGAAAATCAAGGAGACAAGCTT
GTAGATCCAAGTGTCTACCAAAAATTGATTGGAAAACTTCTATATTTGAACATGACTAGA
CCTGACATTTCATTTAGTGTGCAAACATTAAGTCAGTTCTTACAGCAACCTAAAAAATCT
CACATGGATGCTGCACTAAGAGTAGTCAAGTACATCAAGAAACAACCAGGTCAGGGTGTG
TTGTTATCCAGCAGTTCAGGCACTGAAATCACAGCCTACTGTGATGCAGACTGGGCAGCA
TGTCCTACTACAAGGAGGTCTGTAACTGGATTTGTGATCAAGCTAGGAGAGTCTATGGTG
TCTTGGAAAGCTAAGAAGCAAACAACAATATCAAGGAGTTCAGCTGAAGCTGAGTACAGA
AGTTTGGCATCTATTGTAGCTGAGTTAGTATGGTTAGTAGGGCTGCTAAGAAGTTTGGAT

GCTGAAATAACTCTTCCAGTAAACATATACAGTGATAGTAAATCTGCAATTCAGTTAGCA
GCCAATCCAGTCTATCATGAACGTACCAAACACATTGAAATCGATTGTCATTTCATTAGA
GAGAAGTTGCAACAAGGGTTAATCAATATCAGTTATATACCAACTCAAAGTCAGCCAGCT
GATGTTCTAACGAAAGGACTAAACAGAGTACAACATGAGCTCTTATTGTCCAAGCTAGGA
GTTCTAAACATATTTGCTCCTCCAAGCTTGAGGGGGAGTGTTGATAGAAATAAGAGTCTG
TAA
>FGENESH: 1 2 exon (s) 60 - 1542 480 aa, chain +
KTAIGCRWIYKIKYKSIGEVERYKARLVAKGYSQKEGIDYKETFSPVVKMVTVRTILALA
VARKWHIHQMDVYNAFLNGDLQDEIYMTLPQGFTSQGEKKVCRLVKSLYGLKQAPRQWNQ
KLKEALQKLSFSQSQHDHSLFTRKTSKGTTIILVYVDDILVTGSSLELIKETKEALQQVF
KMKDLGELRFFLRIEFARCEAGMVMHQRKYALQLIAEVGVSGAKPSGTPMDVNENQGDKL
VDPSVYQKLIGKLLYLNMTRPDISFSVQTLSQFLQQPKKSHMDAALRVVKYIKKQPGQGV
LLSSSSGTEITAYCDADWAACPTTRRSVTGFVIKLGESMVSWKAKKQTTISRSSAEAEYR
SLASIVAELVWLVGLLRSLDAEITLPVNIYSDSKSAIQLAANPVYHERTKHIEIDCHFIR
EKLQQGLINISYIPTQSQPADVLTKGLNRVQHELLLSKLGVLNIFAPPSLRGSVDRNKSL
>FGENESH:[mRNA] 2 3 exon (s) 2913 - 4080 954 bp, chain -
ATGGCTTCTCGTAGCTTTCTCTTTATTTATGTGTTAGTCATGTTTTCTCTAGCAGGCATG
GCGTTTTCGGACTTGTCCGATGATTTCTACCACCATATTTGTCCCAAAGCTTTACCAACC
ATTAAACGGGTTGTTGAGGATGCAGTCAGAAAAGAGAGACGAATGGGTGCTTCTTTGCTA
CGTTTACATTTTCATGATTGTTTCGTTAATGGTTGTGATGCTTCAATTCTTCTTGATCAA
ACTTCTACTATTAATAGTGAAAAGACTTCTCGTGCTAATAACAATTCAGCCAGAGGATTT
GAGGTGATTGATAAAATTAAATCAGAGGTTGATAAAGTTTGTGGACGTCAAGTTGTGTCT
TGTGCTGACATCTTAGCTGTTGCAGCTCGTGACTCTGTAGTTGCTCTACATGGACCAAGT
TGGAAAGTGAAACTGGGAAGAAGAGACTCGACTACAGCAAGTAGAACCGCGGCCAACGAC
AATATTCCAACTCCGTTTATGGACTTACCTGCACTTATCAAAAACTTCAAGAAGCAAGGT
TTGGATGAGGAAGACCTCGTTGCTCTCTCCGGTTCCCATACACTAGGGTTTGCTCAATGT
TTCACCTTCAGGAATCGCATTTACAATGAGACTAACATTGATCCCACCTTTAGAAGACAA
CGCCAAGCAAATTGTCCACGTAGTGGAGGTGATTCCAATCTTGCTCCACTTGATCCAACA
CCAGCTCTTTTCGACTCAAAATATTTTAGTGACTTAAGGTCCAAGAAAGGGCTTTTACAT
TCTGATCAAGCACTATTTAGTGGAGGAAAAACCGATGATCTTGTGGAGAAATATAGTAAA
AACTTAGGAATGTTTTCGAAAGATTTTGCTGAGTCTATGATTAAGATGGGTGATATCAAG
CCATTGACCGGAAAGCGAGGCCAAATTCGTGTCAACTGCAGGAAGGTGAACTAA
>FGENESH: 2 3 exon (s) 2913 - 4080 317 aa, chain -
MASRSFLFIYVLVMFSLAGMAFSDLSDDFYHHICPKALPTIKRVVEDAVRKERRMGASLL
RLHFHDCFVNGCDASILLDQTSTINSEKTSRANNNSARGFEVIDKIKSEVDKVCGRQVVS
CADILAVAARDSVVALHGPSWKVKLGRRDSTTASRTAANDNIPTPFMDLPALIKNFKKQG
LDEEDLVALSGSHTLGFAQCFTFRNRIYNETNIDPTFRRQRQANCPRSGGDSNLAPLDPT
PALFDSKYFSDLRSKKGLLHSDQALFSGGKTDDLVEKYSKNLGMFSKDFAESMIKMGDIK
PLTGKRGQIRVNCRKVN

C.

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49208885..49210478
Length of sequence: 1594
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 1: in +chain 0, in -chain 1.
Positions of predicted genes and exons: Variant 1 from 1, Score:15.681310



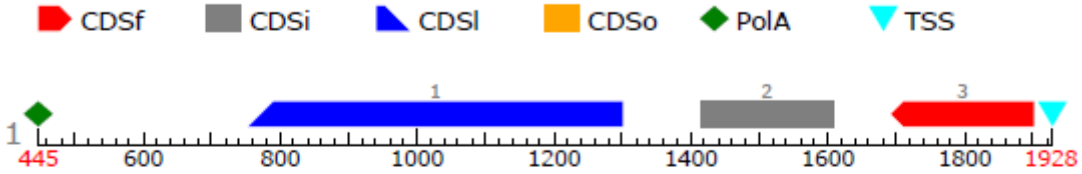

```
1 -      PolA       670                1.61
1 -    1 CDSo       733 -       975   16.34       733 -       975    243
1 -      TSS       1033               -1.88
```

Predicted protein(s):

```
>FGENESH:[mRNA] 1 1 exon (s) 733 - 975 243 bp, chain -
ATGGCTTCGAGTTATAGCTTCCTCTTTCTTCATATGTTAGTCATGTTTAGTCTAGCTAGT
ATCGCTTTTTTCGGATTTGTCAGATGATTTCTACGATGATATTTGTCCCCAAGCTTTGCCA
ACCATTAGACGGGTTGTTGAGGATACAGTCAGGCAAGAGAGGCGAATGGGTGCCTCTTTG
CTACGATTACTTTTTCATGATTGTTTCGTTAATGTCAGTTTAATTTCTTTCTCTCCTACC
TAG
>FGENESH: 1 1 exon (s) 733 - 975 80 aa, chain -
MASSYSFLFLHMLVMFSLASIAFSDLSDDFYDDICPQALPTIRRVVEDTVRQERRMGASL
LRLLFHDCFVNVSLISFSPT
```

D.

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49113591..49115820
Length of sequence: 2230
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 3: in +chain 0, in -chain 3.
Positions of predicted genes and exons: Variant 1 from 1, Score:112.338257



1 -		PolA	445		1.61			
1 -	1	CDSl	752 -	1300	75.37	752 -	1300	549
1 -	2	CDSi	1414 -	1608	26.82	1414 -	1608	195
1 -	3	CDSf	1692 -	1901	13.62	1692 -	1901	210
1 -		TSS	1928		-3.28			

Predicted protein(s):

>FGENESH:[mRNA] 1 3 exon (s) 752 - 1901 954 bp, chain -

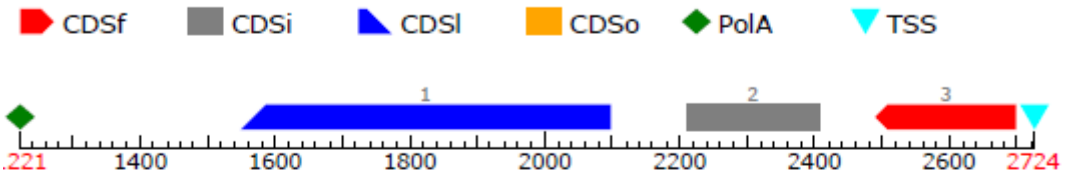
ATGGCTTCTCGTAGCTTTCTCTTTATTTATGTGTTAGTCATGTTTTCTCTAGCAGGCATG
GCGTTTTCGGACTTGTCCGATGATTTCTACCACCATATTTGTCCCAAAGCTTTACCAACC
ATTAAACGGGTTGTTGAGGATGCAGTCAGAAAAGAGAGACGAATGGGTGCTTCTTTGCTA
CGTTTACATTTTCATGATTGTTTCGTTAATGGTTGTGATGCTTCAATTCTTCTTGATCAA
ACTTCTACTATTAATAGTGAAAAGACTTCTCGTGCTAATAACAATTCAGCCAGAGGATTT
GAGGTGATTGATAAAATTAAATCAGAGGTTGATAAAGTTTGTGGACGTCAAGTTGTGTCT
TGTGCTGACATCTTAGCTGTTGCAGCTCGTGACTCTGTAGTTGCTCTACATGGACCAAGT
TGGAAAGTGAAACTGGGAAGAAGAGACTCGACTACAGCAAGTAGAACCGCGGCCAACGAC
AATATTCCAACTCCGTTTATGGACTTACCTGCACTTATCAAAAACTTCAAGAAGCAAGGT
TTGGATGAGGAAGACCTCGTTGCTCTCTCCGGTTCCCATACACTAGGGTTTGCTCAATGT
TTCACCTTCAGGAATCGCATTTACAATGAGACTAACATTGATCCCACCTTTAGAAGACAA
CGCCAAGCAAATTGTCCACGTAGTGGAGGTGATTCCAATCTTGCTCCACTTGATCCAACA
CCAGCTCTTTTCGACTCAAAATATTTTAGTGACTTAAGGTCCAAGAAAGGGCTTTTACAT
TCTGATCAAGCACTATTTAGTGGAGGAAAGACCGATGATCTTGTTGAGAAATATAGTAAA
AACTTAGGAATGTTTTCCAAAGATTTTGCTGAGTCTATGATTAAGATGGGTGATATCAAG
CCATTGACCGGAAAGCGAGGCCAGATTCGTGTCAACTGCAGGAAGGTGAACTAA

>FGENESH: 1 3 exon (s) 752 - 1901 317 aa, chain -

MASRSFLFIYVLVMFSLAGMAFSDLSDDFYHHICPKALPTIKRVVEDAVRKERRMGASLL
RLHFHDCFVNGCDASILLDQTSTINSEKTSRANNNSARGFEVIDKIKSEVDKVCGRQVVS
CADILAVAARDSVVALHGPSWKVKLGRRDSTTASRTAANDNIPTPFMDLPALIKNFKKQG
LDEEDLVALSGSHTLGFAQCFTFRNRIYNETNIDPTFRRQRQANCPRSGGDSNLAPLDPT
PALFDSKYFSDLRSKKGLLHSDQALFSGGKTDDLVEKYSKNLGMFSKDFAESMIKMGDIK
PLTGKRGQIRVNCRKVN

E.

FGENESH 2.6 Prediction of potential genes in Tomato genomic DNA
Seq name: chr10:49177729..49181521
Length of sequence: 3793
Number of predicted genes 1: in +chain 0, in -chain 1.
Number of predicted exons 3: in +chain 0, in -chain 3.
Positions of predicted genes and exons: Variant 1 from 1, Score:108.660205



1	-		PolA	1221		1.61		
1	-	1	CDSl	1548	- 2096	71.77	1548 - 2096	549
1	-	2	CDSi	2210	- 2404	26.82	2210 - 2404	195
1	-	3	CDSf	2488	- 2697	13.62	2488 - 2697	210
1	-		TSS	2724		-3.28		

Predicted protein(s):

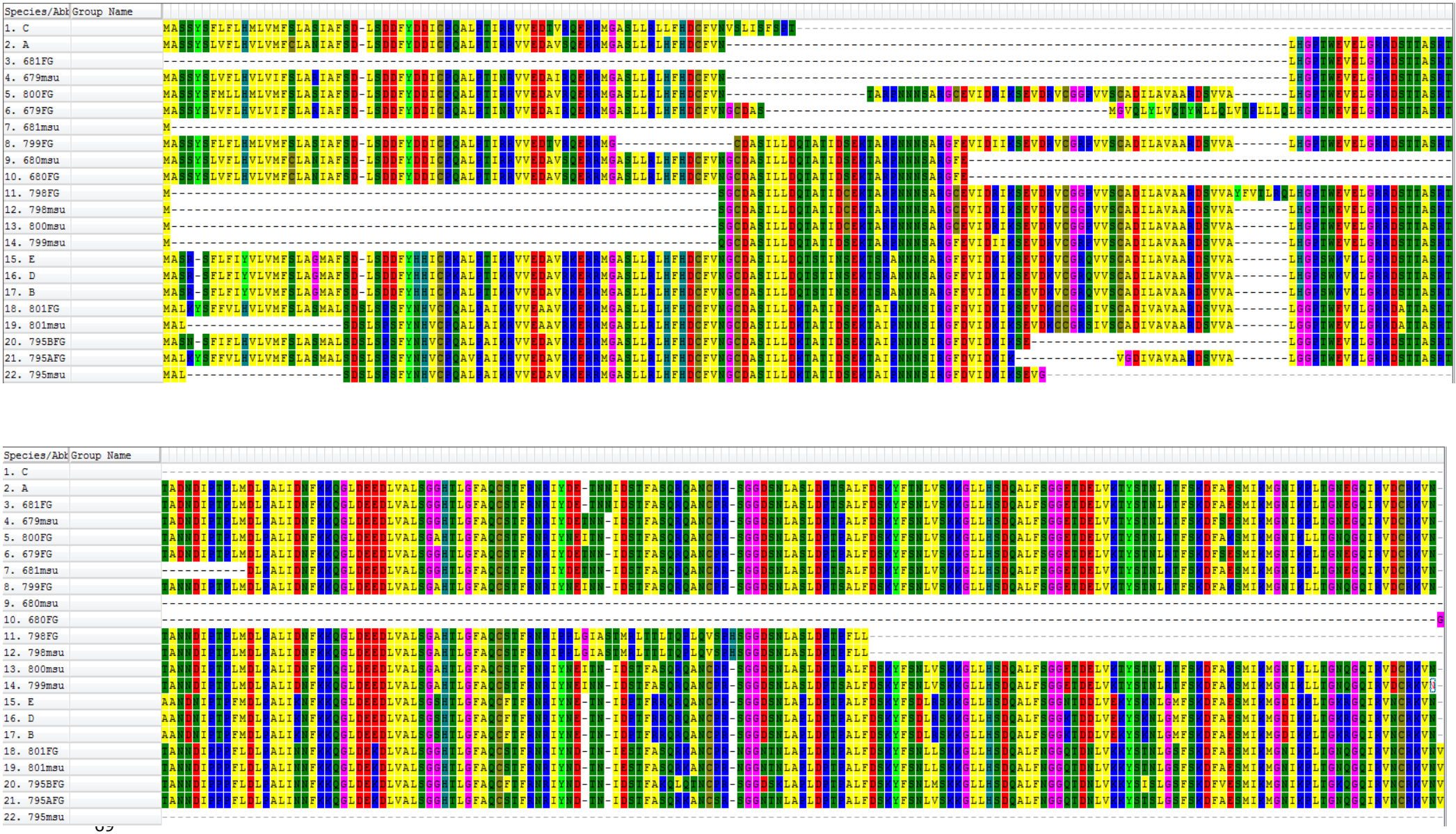
>FGENESH:[mRNA] 1 3 exon (s) 1548 - 2697 954 bp, chain -

ATGGCTTCTCGTAGCTTTCTCTTTATTTATGTGTTAGTCATGTTTTCTCTAGCAGGCATG
GCGTTTTCGGACTTGTCCGATGATTTCTACCACCATATTTGTCCCAAAGCTTTACCAACC
ATTAAACGGGTTGTTGAGGATGCAGTCAGAAAAGAGAGACGAATGGGTGCTTCTTTGCTA
CGTTTACATTTTCATGATTGTTTCGTTAATGGTTGTGATGCTTCAATTCTTCTTGATCAA
ACTTCTACTATTAATAGTGAAAAGACTTCTCGTGCTAATAACAATTCAGCCAGAGGATTT
GAGGTGATTGATAAAATTAAATCAGAGGTTGATAAAGTTTGTGGACGTCAAGTTGTGTCT
TGTGCTGACATCTTAGCTGTTGCAGCTCGTGACTCTGTAGTTGCTCTACATGGACCAAGT
TGGAAAGTGAAACTGGGAAGAAGAGACTCGACTACAGCAAGTAGAACCGCGGCCAACGAC
AATATTCCAACTCCGTTTATGGACTTACCTGCACTTATCAAAAACTTCAAGAAGCAAGGT
TTGGATGAGGAAGACCTCGTTGCTCTCTCCGGTTCCCATACACTAGGGTTTGCTCAATGT
TTCACCTTCAGGAATCGCATTTACAATGAGACTAACATTGATCCCACCTTTAGAAGACAA
CGCCAAGCAAATTGTCCACGTAGTGGAGGTGATTCCAATCTTGCTCCACTTGATCCAACA
CCAGCTCTTTTCGACTCAAAATATTTTAGTGACTTAAGGTCCAAGAAAGGGCTTTTACAT
TCTGATCAAGCACTATTTAGTGGAGGAAACACCGATGATCTTGTTGAGAAATATAGTAAA
AACTTAGGAATGTTTTCCAAAGATTTTGCTGAGTCTATGATTAAGATGGGTGATATCAAA
CCATTGACCGGAAAGCGAGGCCAAATTCGTGTCAACTGCAGGAAGGTGAACTAA

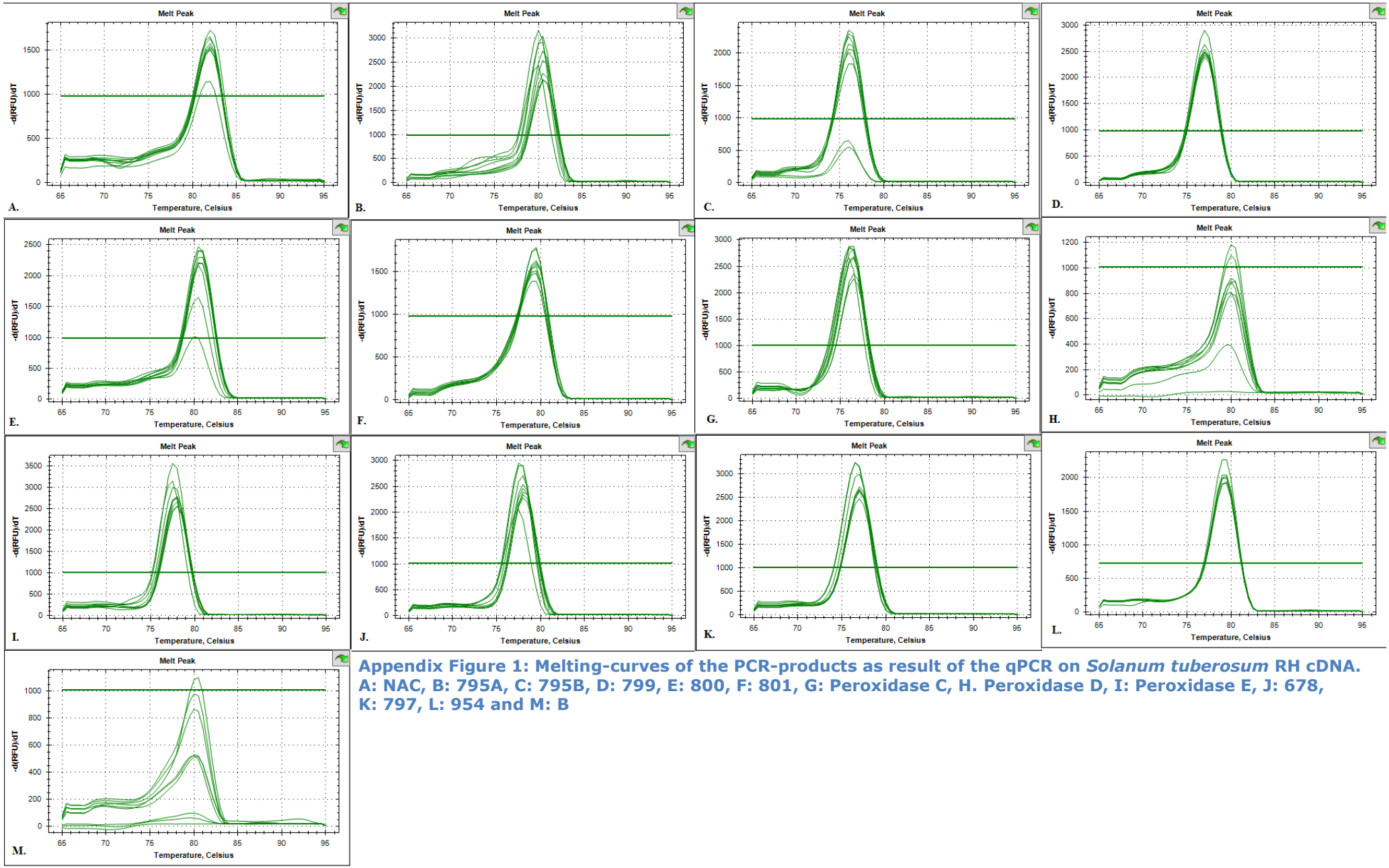
>FGENESH: 1 3 exon (s) 1548 - 2697 317 aa, chain -

MASRSFLFIYVLVMFSLAGMAFSDLSDDFYHHICPKALPTIKRVVEDAVRKERRMGASLL
RLHFHDCFVNGCDASILLDQTSTINSEKTSRANNNSARGFEVIDKIKSEVDKVCGRQVVS
CADILAVAARDSVVALHGPSWKVKLGRRDSTTASRTAANDNIPTPFMDLPALIKNFKKQG
LDEEDLVALSGSHTLGFAQCFTFRNRIYNETNIDPTFRRQRQANCPRSGGDSNLAPLDPT
PALFDSKYFSDLRSKKGLLHSDQALFSGGNTDDLVEKYSKNLGMFSKDFAESMIKMGDIK
PLTGKRGQIRVNCRKVN

6.6 Alignment of the peroxidase proteins

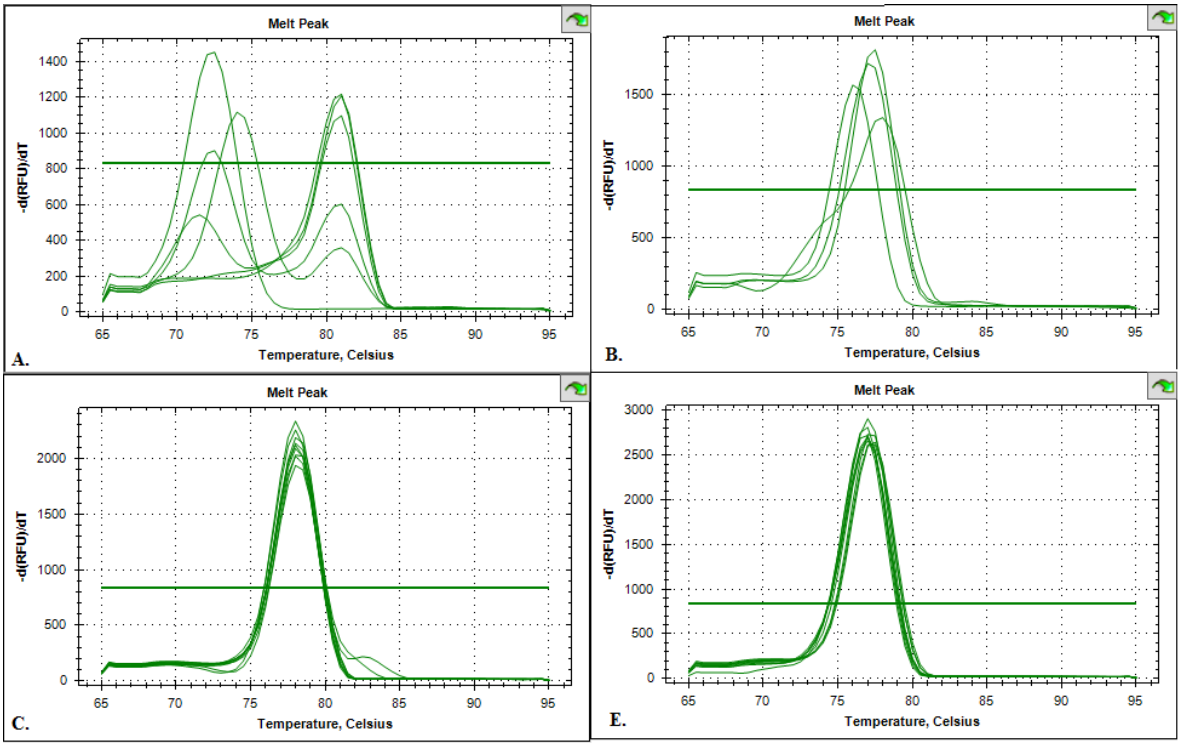


6.7 Melting-Curves of the qPCR-products (*Solanum tuberosum* RH)



Appendix Figure 1: Melting-curves of the PCR-products as result of the qPCR on *Solanum tuberosum* RH cDNA. A: NAC, B: 795A, C: 795B, D: 799, E: 800, F: 801, G: Peroxidase C, H. Peroxidase D, I: Peroxidase E, J: 678, K: 797, L: 954 and M: B

6.8 Melting-Curves of the qPCR-products (*Solanum tuberosum* CxE)



Appendix Figure 2: Melting-curves of the PCR-products as result of the qPCR on *Solanum tuberosum* CxE dormant tubers cDNA. A: Peroxidase B, B: Peroxidase C, C: 678 and D: 797