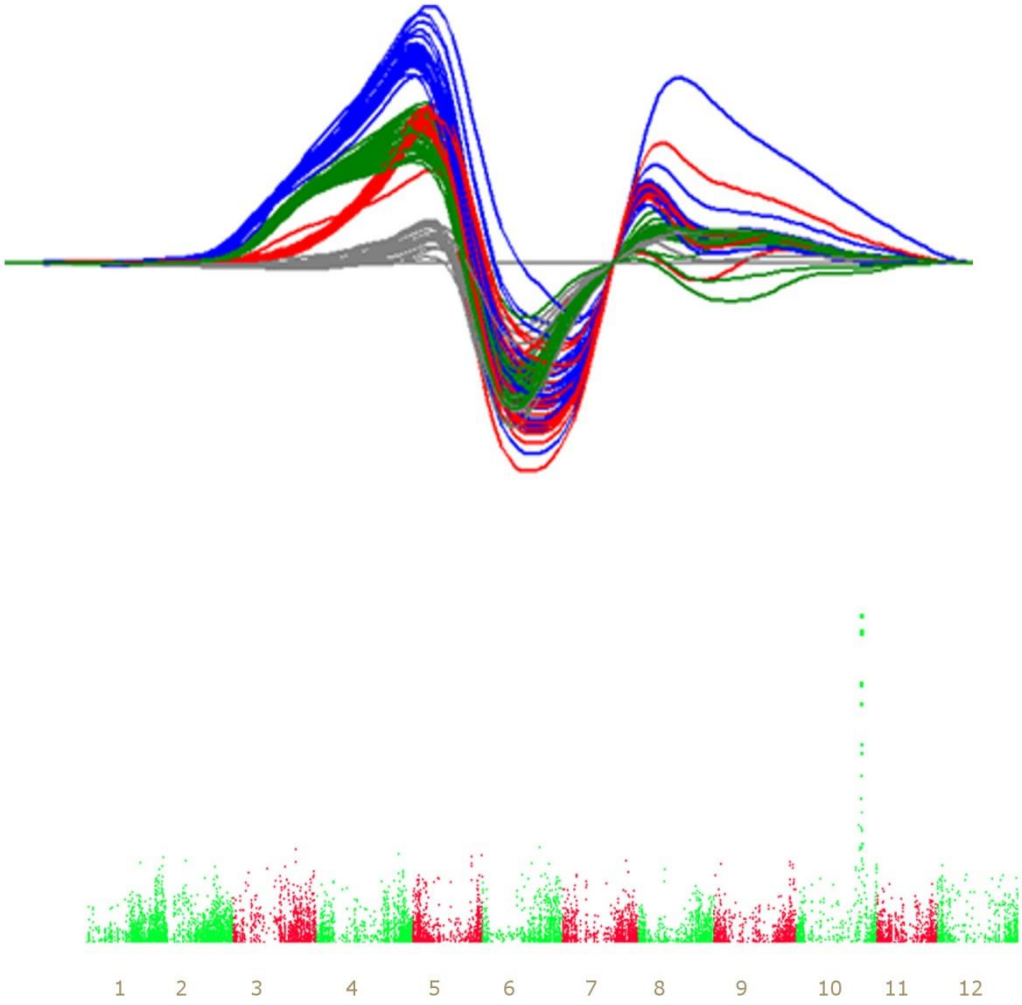
The identification of candidate genes responsible for variation in potato tuber shape

2014



Wageningen University- Department of Plant Breeding - Johan Willemsen- June 2014

Cover page illustration:

- Melt patterns of amplicons from HRM analysis of 96 CxE seedlings
- GWAS results show a significant peak on chromosome 10 for potato tuber shape

The identification of candidate genes responsible for variation in potato tuber shape

Johan Willemsen
Registration number: 910721959030
Major MSc Thesis Plant Breeding (PBR-80436)

Supervisors:
Dr. ir. H.J. van Eck
Ir. P. Vos

Examiners:
Dr. ir. H.J. van Eck
Dr. C.W.B. Bachem

Department: Plant breeding
Wageningen University
June 2014

Acknowledgements

This thesis would not have been possible without the help of other people. First I would like to thank my supervisors Herman van Eck and Peter Vos for allowing me to do research on this topic. Only after I did start with my MSc-thesis in September 2013, I realized that this topic was more or less a hobby of Herman and Peter. Nonetheless I have enjoyed working on this topic and have learnt a lot about QTL mapping, and about working in a genetically complex non-model species.

First I would like Herman for his critical feedback during work on my MSc-thesis and stimulating discussions about a wide range of topics. When I had a question, his office was always open and time was not an issue. Herman, thanks for that. I really learned a lot and I know that doing this thesis in potato has given me a strong feeling that I really enjoy puzzling on difficult challenges in (potato) genetics.

Secondly I would like to thank my other supervisor, Peter Vos who helped me with almost all practical things in the lab and with setting up the recombinant screen experiment. Although we did not work a lot together, your confidence in me, and support during the eight months really did help.

Another person that I want to thank is Ernest Aliche, who organized sowing plants for the second batch of seedlings, and helped with isolating DNA. Without this batch I would never have improve the QTL mapping of the *Ro* locus. Furthermore I want to thank all the students that were spending time in the 'breeders hole', because they helped with many things. Especially Jonathan Kalisvaart, who I would like to thank for all the discussions about this or other topics during the many coffee and lunch breaks that we spent together.

Summary

Potato tuber shape is an intriguing morphological trait which displays continuous trait variation ranging from compressed, round to oval and long. For potato tuber shape a single locus with multiple alleles on chromosome 10 has been proposed which explained a large proportion of the phenotypic variation (van Eck et al. 1994). This *Ro* locus was further fine mapped to a region of 900 kb in Witteveen (2013). In this study a recombinant screen has been conducted to identify the gene underlying the *Ro* locus.

In this recombinant screen of the C×E experimental population, the *Ro* locus is further confined to a region of 580 kb on scaffold DMB385,DMB546 and DMB446. In this region the so-far unanchored scaffold DMB773 was inserted to the reference genome by manual scaffolding based on BAC-end sequences of DM and RH BAC libraries. The insertion of this scaffold extended the preliminary target region of *Ro* by 133 kb.

The development of markers that are informative for meiosis of both parent proved to be difficult, therefore the sequence information from re-sequenced parental genotypes was analysed and subsequently large tracts of homozygosity were discovered in the target region, hindering successful marker development. From this analysis it was evident that using sequence information for marker development may improve the success rate of marker development, especially combined with haplotype phase information.

The recombinant screening as done in Witteveen (2013) identified co-segregation of a large region (~300 kb), but in a screening on 800 seedlings, six recombinants were identified within the co-segregating region. Phenotype analysis of these recombinants revealed that the *Ro* locus is confined between marker Asp6678 and marker Per20801, resulting in an region of approximately 280 kb.

In this region 14 candidate genes are located from which seven members of the peroxidase family. Based on expression analysis of all 14 positional candidate genes, the most likely candidate gene for studying potato tuber shape are members of the peroxidase gene family. Further fine mapping or candidate gene approaches will reveal more about how potato tuber shape is regulated.

Contents

Chapter 1 Introduction

Potato

Potato (*Solanum tuberosum)* is currently the world's fourth important food crop in terms of production (FAOSTAT, 2014). The genus *Solanum* consists out of many species, from which the European cultivated potato varieties belong to the *Solanum tuberosum* Group tuberosum. The geographic origin of the cultivated potato is the Andean region, where many wild relatives of *Solanum tuberosum* have been discovered (Hijmans and Spooner 2001). *Solanum tuberosum* is an outbreeding autotetraploid species with a heterozygous genome (Xu et al. 2011).

Traditionally, the focus of research in potato was the genetic basis of agronomical traits but the genetic basis of quantitative morphological traits, such as tuber shape is less investigated (van Eck 2007). In the cultivated potato, tuber shape is distinguished in the morphological classes round-oval-long as measure of the overall tuber shape (The European Cultivated Potato Database), but in wild relatives of *Solanum tuberosum* an enormous variation in tuber shape is revealed. As suggested by Uitdewilligen et al. (2010) during domestication, potatoes have been selected for uniform shape, because of demands of processing industry and consumers. This would result in less variation for potato tuber shape in cultivated potatoes opposed to more variation in tuber shape in wild species.

As such, potato tuber shape is an important characteristic for the processing industry, where cultivars producing long potatoes in general are used for the frying industry and cultivars producing round potatoes are used for crisps (van Eck et al. 1994). For marker-assisted plant breeding potato tuber shape is not important, because phenotypic selection is straightforward and therefore the application of markers for breeding is irrelevant.

Modulation of organ shape

The domestication of plants has resulted in many morphological and physiological changes that distinguish domesticated species from their wild progenitors (Alonso-Blanco et al. 2009). Whereas much research had as focus flowering time, flower colour, seed dormancy, germination and the primary metabolism of the plant, less is known about morphological traits, such as potato tuber shape. An example of a well-researched morphological trait is fruit size in *Solanum lycopersicum* where several genes have been identified that are involved fruit size and shape (Tanksley 2004). Other examples are the leaf shape of *Vitis Vinifera* (Chitwood et al. 2014) and grain shape of *Oryza sativa* (Fan et al. 2006). An extreme example of morphological modification is the *Brassica* genus where several morphotypes are observed in which different organ structures have enlarged (Lin et al. 2014).

In *Solanum tuberosum* much variation is observed for potato tuber shape. The potato tuber shape is a composite trait, consisting out of numerous characteristics that make up the potato tuber shape (Herman J. van Eck 2007). Commonly observed phenotypes are 'Kidney', which are compressed potatoes and long potatoes such as the cultivar *Asparges*. However in this study the potato tuber shape is defined as the length of the apex-heel relative to the tuber diameter (L/W

ratio), which can be seen as a measure of the overall tuber shape. Besides this quantitative measurement with the L/W ratio, the potato tuber shape can be easily visually assessed by distinguishing round from long tubers.

The developmental biology of tuber shape has been described extensively, but these studies describe the tuber organ development, while ignoring the genetic differences between genotypes or cultivars. Several studies describe changes that occur during tuber initiation and growth (Xu, Vreugdenhil, and Lammeren 1998; Ewing and Struik 1972; Leshem and Clowes 1972), but nonetheless these studies do not investigate the difference in cell shape, number and elongation between different potato genotypes.

Despite that these studies do not investigate differences in potato tuber shape within different genetic backgrounds, observations on tuber development suggest that potato tuber shape is formed early in the development of the tuber and visible after several weeks of growing (Xu, et al. 1998; Kloosterman et al. 2008). Moreover, Xu et al. (1998) observed that pith formation stops relatively fast after initiation of tuber swelling. Besides the early arrest of pith formation, a positive association was observed of tuber pith narrowness and longer potato tubers (Maj and Misjener, 1994). These observations imply that potato tuber shape is already formed in an early stage of the tuber development.

During tuber development from stolon to mature tuber, a total of 15,235 genes is expressed from which 1,217 transcripts are differentially expressed (Xu et al. 2011). In early stages of the tuber development (stolon to tuber transition) it is revealed that genes involved with cell cycle and translation are up-regulated (Kloosterman et al. 2008). These general patterns of gene expression are related to cell growth and cell division, which are expected to have a role in the formation of the potato tuber shape.

The identification of genes

In general there are two approaches for finding causative genes for traits under investigation. On one hand with reverse genetics approaches functions of genes are discovered by using sequence variations to find genetic basis of a phenotype or trait. On the opposite with forward genetic approaches associations are found on the basis of combining phenotype information with sequence information.

With forward genetics two approaches are possible. The first approach involves association of marker information with a given trait followed by QTL mapping within a segregating population (Pfieger et al. 2007). Genes are considered candidate genes on the basis of linkage with traits. The second approach is postulating candidate genes on the basis of biological information of the trait involved. This approach is often hindered by the limited understanding of the molecular pathway or the complexity of the trait (Kloosterman et al. 2010).

In *Solanum tuberosum* linkage mapping is in general more complex, because of the tetrasomic inheritance, which allows many different genetic models being postulated for each marker position. As a consequence only dominant markers are available in tetraploid genotypes for linkage analysis (i.e. simplex x nulliplex; duplex x nulliplex and simplex x simplex) (Gebhardt 2007). One major

advantage for doing linkage studies in potato is that the ploidy level can be reduced to the diploid level, which decreases the complexity of linkage mapping in potato. Whereas in crops often recombinant inbred lines are used for linkage studies, the creation of RILs in potato is not possible because of strong inbreeding depression when selfing. As a consequence, linkage studies in outbred heterozygous species have often been done with mapping populations derived from two non-inbred parents. The C×E experimental population as used in this study is an example of a pseudo backcross. Besides above mentioned characteristics, the extreme high heterozygosity allows the construction of two linkage maps in a single F1 population (Ritter, Gebhardt, and Salamini 1990).

The *Ro* locus

It is generally known that potato tuber shape is a quantitative trait, which displays continuous variation (Jong and Burns 1993; Masson 1985). As shown by Masson in 1985, diploid potato cultivars segregate in a simple Mendelian fashion, with round alleles being dominant over long potatoes. This locus was described as the *Ro* locus, with *Ro* dominant over *ro*. In a study by van Eck et al. (1994) a qualitative model was described to explain this continuous variation in potato tuber shape, where combinations of multiple alleles have different phenotypic effects. This study located the major tuber shape locus on chromosome 10 by means of RFLP linkage mapping (van Eck et al. 1994). They also observed a recombination event in the target region of the *Ro* locus that happened in meiosis leading to the male parent. Besides this initial linkage mapping, an association study showed significant associations with tuber shape on chromosome 10 at super-scaffold 385 (Vos et al. unpublished) With HRM analysis, Witteveen (2013) further mapped the *Ro* locus to an interval of 910 kb on chromosome 10.

Tuber shape is characterized by a high heritability, which ranges between 0.75 and 0.9 depending on the population and field design that is used (van Eck et al. 1994; D'hoop et al. 2014). As suggested by d'Hoop et al. (2014) the effect of GxE interactions is neglectable, because in a wide reference population the variance components associated with effects of location, year and trail are low. This results in a high overall heritability.

Use of sequence data

With the rise of next-generation sequencing technology, the use of sequence data for linkage mapping has steadily increased in the past decade. Traditional methods based on RFLP and SSR markers are less commonly used and SNP markers are becoming the most important marker type (Celton et al. 2010; Chagné et al. 2012). For linkage analysis sequence information has proven to be useful, as illustrated by the use of sequence information in marker-trait association of the *StCDF* gene involved in plant maturity (Kloosterman et al. 2010). Another approach is used in an investigation of the tuber flesh colour locus, using transcriptome sequencing. This approach is based on gene expression data and needs *a priori* knowledge about the trait under investigation, which is not feasible for less researched traits or more complex traits.

For the development of linkage maps, several low-throughput methods have been developed like CAPS marker analysis and High Resolution Melting (HRM). HRM analysis was first described by

Wittwer (2003) and has become more popular for use in linkage studies (Celton et al. 2010). This technology uses differences in melt temperature of amplicons to detect sequence variations (Wittwer 2003). Heterozygous and homozygous amplicons can be distinguished, as demonstrated by (Montgomery et al. 2010). The advantages of HRM are that it is a non-destructive method and that post-melting screening can be done (Vossen et al. 2009)**.** The disadvantages of HRM are that homozygous variants are difficult to distinguish, and the difference in melt patterns for some variants (A/T) SNPs is subtle and can often not be detected (Reed and Wittwer 2004)**.**

For HRM analysis multiple dyes are available, as well as multiple machines that provide accurate melt curve analysis (Wittwer, 2014). LCgreen™ is the most used fluorescent dyes, however other dyes such as EvaGreen™ or SYTO9 perform well in melt curve analysis. Fluorescent dyes like SYBR Green, which are commonly used in qPCR applications may also be suitable for HRM purposes as illustrated by the successful use of SYBR green by Price et al. (2007).

The utility for genotyping and variant scanning of diploid and tetraploid potato is already investigated in De Koeyer et al. (2009). In this study it was suggested that HRM could be used to do a HRM based haplotype diversity study at a given locus. Another example of the use of HRM in *Solanum tuberosum* was the fine mapping of the *StCDF* gene (Kloosterman et al. 2013).

Objectives

In this study it is tried to get more insight in the mechanism of potato tuber shape formation by the use of forward genetic screens. The primary goal was to identify the major effect QTL involved in potato tuber shape (using co-localization with genes) and derive positional candidate genes. To answer this question a fine mapping study was done using HRM on the experimental mapping population C×E which segregates for tuber shape. This population was previously used for mapping the *Ro* locus (Van Eck et al. 1994). Besides narrowing down the *Ro locus* to, it is tried to integrate sequence data as an integral part for HRM marker development. Furthermore *in silico* approaches have been used on publically available data to obtain more information on possible functional candidate genes.

Chapter 1. Material and methods

Plant material

About 2000 seedlings are grown from the C×E experimental population. The female parent C (US-W5337.3) originated from a cross between Solanum phureja PI225696.1 and the dihaploid US-W42, extracted from cv. Chippewa) and the male parent was a cross between VH3 4211 (*Solanum Vernei* tuberosum backcross) and the C parent (US-W5337.3). This population is previously used for QTL mapping purposes of the *Ro* locus (van Eck et al. 1994; Jacobs et al. 1995). It is known that both parents have round tubers and descendants segregate for tuber shape in a 3:1 ratio.

Seedlings were grown for three weeks and recombinants were distinguished from non-recombinant seedlings with using flanking markers of the *Ro* locus. From the initial batch of seedlings the crumpled genotypes were removed and not analysed for recombination. After selection of recombinants, plants were transplanted in pots of 1.1 L under 16h light conditions. Plants were grown until tuber maturation and shoot senescence.

DNA isolation

For HRM genotyping DNA was isolated with NAOH-based DNA isolation, adapted from the protocol of Wang et al. (1993) (see SI). This DNA was only used for the pre-screening of recombinants, whereas for the fine mapping Kingfisher DNA isolation was done to obtain more and pure DNA suitable for long-term storage.

SNP calling

C and E sequences for target region (Scaffold DMB385, DMB546 and DMB446) have been obtained from Theo Borm (unpublished data) as BAM files. Subsequently variant detection was done with GATK (DePristo et al. 2011) using the best practices as developed by the Broad Institute (Van der Auwera et al. 2013). Prior the SNP calling, quality control of read mapping is done with FASTQC and duplicate reads are flagged with Picard DuplicateReads script (http://picard.sourceforge.net). The visualization of input BAM files and output VCF files was done with the Integrative Genomics Viewer (IGV) using the potato DM genome release 4.03 as a reference. Overall nucleotide diversity for the target region was calculated with $\pi = \frac{number\ of\ pair-wise\ differences}{number\ of\ haplotypes \times length\ of\ sequence}$ (Nei and Li 1979).

Development of HRM markers

Initially markers are developed with a random approach by selecting intron-spanning amplicons with maximum length of 500bp. However when sequence information became available, markers were developed with using re-sequenced data from parental genotypes. The assumption underlying marker design, is that two out of four haplotypes are the same at any given position in the target region. With this assumption several genetic models are possible, namely WPxPV, WP x PP WPx PW:

Class I:	WP x PV	– Heterozygous in both parents	- 1:1:1:1
Class II:	PP x PV	– Heterozygous in E parent	- 1:1

Class III:	WP x PP	– Heterozygous in C parent	- 1:1
Class IV:	WP x PW	– Heterozygous in both parent	- 1:2:1
Class V:	PPxPP	– Homozygous in both parents	- No segregation

Class I is regarded as fully informative (i.e. detects meiosis in both father and mother), whereas class II and III are only informative in one parent. Class IV is informative but produces two homozygous amplicons, which are difficult to interpret. Class V are regions in which marker design is useless, because no segregation will occur. Although these are theoretical segregation patterns, HRM markers also allow the occurrence of 3 patterns in a 1:2:1 ratio, when a class I model is used, depending on the SNP type that is causative to the difference in haplotypes.

PCR-based screening with markers.

The development of HRM markers was done with primer3+ and different fluorescent dyes and polymerases were tested. To test a HRM assay eight progeny were screened using LCgreen and Phire. Assays were approved for further use when the four different allele combinations could be scored with melt patterns. Besides markers with four patterns, also markers with three distinct melt patterns were used. This situation can be explained when parental genotypes coincide (Aa), but show two three patterns in the CxE population. Although markers segregating in either the female or the male parent were less informative and gave only information about one parental meiosis, they were also used for screening.

A total of 34 primer sets have been developed with primer3+ and markers as used by Witteveen (2013) were also included in the HRM analysis. Assays were processed with standard PCR with addition of fluorescent dye. Both LCGreen as Evagreen were used as fluorescent dyes. In addition also different polymerases were used namely DreamTaq and the hot-start Phire DNA polymerase. Specific protocols for each assay can be found in the supplementary information (SI II).

Recombinant screening:

The recombinant screen was done in two steps. First to distinguish recombinant seedlings from non-recombinant seedlings, plants were screened with two flanking markers of the target region. In batch CE2013-okt the seedlings were screened with marker B446 and marker Amt241073, however in the CE2014-feb batch marker PhoTr31222 and Amt241073 are used to screen for recombinants. After selecting recombinant seedlings, DNA was isolated with the Kingfisher DNA isolation kit, which subsequently was used for saturation the region of interest with markers. This subset of recombinants was screened with HRM analysis for fine mapping of recombination events.

Recombination frequencies are calculated by dividing the number of recombinants by the total number of seedlings. The recombinant percentage was converted into a map distance with the Kosambi map function (Kosambi 1943), which is defined as: $d = \frac{1}{4} ln\frac{1+2r}{1-2r}$. Genetic maps were visualized with Mapchart (Voorrips 2002).

Phenotype analysis of tuber shape

A total number of three representative tubers were quantitatively evaluated for the length (longitudinal) and tuber width (transversal) of recombinants in target region. Subsequently phenotypes were defined as the length/width ratio (L/W). Distinct classes of tuber morphology are defined as flat tubers when the L/W ratio is somewhat less than one, round/oval tubers when the L/W ratio is around one, and long tubers when the L/W ration is substantial higher than one (Tai and Misener 1994). Besides a quantitative evaluation, also a qualitative evaluation of tuber shape was done by assigning tubers visually to long or round tuber categories.

Statistical analysis

Segregation distortion: For testing of segregation distortion the Pearson chi-squared test was used to investigate the influence of gametic or zygotic selection on the target region. Gametic selection was investigated within the gametic classes from each parent separately and assumed significant with χ^2 (df=1) > 3.84 (α: 0.05). Zygotic selection was assumed to be significant if χ^2 (df=1) > 3.84 (α: 0.05) over all genotypic classes.

Effect of genotypic classes: To investigate whether each of the four genotypic classes has a different effect on the phenotype, an analysis of variance was done on all recombinants of each parental genotype combination with corresponding quantitative phenotypes. This was implemented using the LMM in Genstat 14. Subsequently a Tukey post-hoc test to compute pair-wise differences between groups was performed.

Variance components: Linear mixed models were used to estimate variance components. Several other experimental factors were introduced in the model: 1) the qualitative phenotypic classes for long or round, 2) the genotype classes within the morphological classes, 3) the among clone effect, 4) the between clone effect and 5) the within tuber effect. These models are compared with the Likelihood Ratio Test to compare models and determine whether using an extra term is necessary. All terms are included as random to estimate variance components. The model is formulated as following: $y_{ijklm} = \underline{Morpho_i} + \underline{Geno_{ij}} + \underline{Cl_{ijk}} + \underline{Cl_{ijkl}} + \underline{Tu_{ijklm}} + \underline{e_{ij}}$

Heritability estimates: For heritability estimates the following model was used: The broad-sense heritability was estimated by using variance components as estimated with using a linear mixed model. The broad-sense heritability is defined as: $h^2 = \frac{\sigma^2_{gen}}{\sigma^2_{gen} + \sigma^2_{env}}$. The proportion of the variance explained by the visual classification was calculated by: $h^2 = \frac{\sigma^2_{Ge}}{\sigma^2_{Ge}\, \sigma^2_{Cl} + \sigma^2_{Tu} + \sigma^2_{e}}$. The proportion of the variance due to the *Ro* locus was estimated by: $h^2 = \frac{\sigma^2_{Ro}}{\sigma^2_{Ro}\, \sigma^2_{Cl} + \sigma^2_{Tu} + \sigma^2_{e}}$.

Integration of scaffold DMB773

Previously unanchored scaffold DMB773 was integrated in the DM reference genome V4.03 (Sharma et al. 2013) by using several approaches. The first approach was manually checking if two BAC-ends, which had been aligned to the reference genome by Sharma et al. (2013) belong to the same clone. Subsequently BAC clones were selected which connected scaffold DMB385 and DMB773 or BAC clones which connected scaffold DMB773 and DMB546.

The second approach was done by aligning BAC-ends of RH and DM to the target scaffolds (DMB385, DMB773, DMB546 and DMB446) using the standard BLAST+ suite (Camacho et al. 2009; Morgulis et al. 2008). To detect significant hits two criteria were defined: 1) The BAC-end hit to the reference genome should be specific, 2) the length of the region between forward and reverse BAC-end hits should be between 30 kb and 200 kb.

A stepwise approach was employed. First, all BAC-sequences were aligned to the target region and significant hits were selected on coverage and identity parameters. Significant hits were then checked for specificity against the whole reference genome 4.03. The final step was to calculate the distance between forward and reverse hits, and select hits that were located between 30 kb and 200 kb from each other. The same analysis has been done with DM fosmid ends, although the length of fosmid clones was defined as between 10 and 30 kb.

Synteny comparison with *Solanum lycopersicum*

To investigate the amount of similarity between *Solanum lycopersicum* and *Solanum tuberosum* a synteny analysis was performed. Preliminary synteny analysis was performed by using dotplot analysis of target scaffold against the *Solanum lycopersicum* chromosome 10. These dotplots were made by using a word-size of 30 bp.

Subsequently a more detailed synteny analysis was done with the Symap 4.0 synteny explorer using the potato genome reference sequence V4.03. In this analysis, scaffold DMB773 was inserted between scaffold DMB485 and DMB546. From the synteny explorer, synteny hits of genes were exported and cross-referenced with the list of target genes in *Solanum tuberosum*.

Expression analysis of RH and DM RNA-seq datasets

To get more information about potential candidate genes RNA-seq datasets as provided by (Xun Xu et al. 2011) for both RH89-038-16 and DM1-3 516 R44 tissues were used to investigate expression profiles for genes located in the target region. An overview of used subset of tissue samples for both DM as well as RH89-038-16 can be found in table one. Other tissue samples are described in Xu et al. (2009). Gene expression was analysed with the edgeR package for R and fold changes were calculated from FKPM values compared to stolon tissue and visualized.

Genotype	Tissue	Description	Time of sampling
DM1-3 516 R44	Stolons	Some stolons were above ground and were green	-
DM1-3 516 R44	Tubers, Sample 1	Whole tubers	5-5-09
DM1-3 516 R44	Tubers, Sample 2	Whole tubers	6-1-09
RH89-039-16	Mature Tuber	Large tubers with average size range between 2.5 and 3.5 cm (tuber filling stage).	Four weeks after visible swelling
RH89-039-16	Stolon	Non-tuberizing stolon tips (2-6 cm in length).	After several weeks of growth.
RH89-039-16	Young Tuber	Small developing tuber with size of around 1cm	After one week of the observation of first swelling, during anthesis

Table 1 Description of informative tissues for tuber development, from RNA-seq datasets as provided by Xu et al. (2011). These tissue samples were used for expression profiling of candidate genes.

For phylogenetic analysis of peroxidase genes a multiple alignment was done on amino-acid sequences with the Clustal algorithm and tree construction was performed using the Neighbor-joining method. Reliability of the dendrogram was assessed with using a bootstrap test with 500 replicates.

Chapter 2. Results

Pre-screening results

About 2500 plants are grown in batch CE2013-okt) and after removal of crumpled genotypes (crcr), a total of 2000 seedlings were screened, from which 1217 were informative for observing recombination events. The second batch (CE2014-feb) included in total 1500 plants, and after removal of crumpled genotypes, 960 plants remained. From these, 798 progeny were informative and 29 recombinants were selected and screened for recombination with HRM.

Recombination events in the target region were identified by using flanking HRM markers on three week old seedlings. As already indicated by van Eck et al. (1994) the target region of the Ro locus has an ancestral recombination event in the E parent. The C×E population can be regarded as a backcross, which implies that three haplotypes are available in the parents. Each parent has four possible gametes which results in a total of 16 different combinations in the progeny.

	Crossing scheme
Paternal genotypes	$\frac{AA}{BB} \times \frac{AB}{CC} \rightarrow \frac{AA}{CC} : \frac{BB}{CC} : \frac{BB}{AB} : \frac{AA}{AB} \quad (1:1:1:1\ ratio)$
Recombinants in female parent	$\frac{\boldsymbol{AB}}{\boldsymbol{BA}} \times \frac{AB}{CC} \rightarrow \frac{AB}{CC} : \frac{BA}{CC} : \frac{AB}{AB} : \frac{BA}{AB} \quad (1:1:1:1\ ratio)$
Recombinants in male parent	$\frac{AA}{BB} \times \frac{\boldsymbol{AC}}{\boldsymbol{CB}} \rightarrow \frac{AA}{AC} : \frac{AA}{CB} : \frac{BB}{AC} : \frac{BB}{CB} \quad (1:1:1:1\ ratio)$
Double recombinants	$\frac{\boldsymbol{AB}}{\boldsymbol{BA}} \times \frac{\boldsymbol{AC}}{\boldsymbol{CB}} \rightarrow \frac{AB}{AC} : \frac{BA}{CB} : \frac{AB}{CB} : \frac{BA}{AC} \quad (1:1:1:1\ ratio)$

Table 2 Crossing scheme of female parent C with male parent E. Recombinants were selected on the basis of two markers. Each parent has four possible gametes, with two recombinant gametes (in bold). This results in a total of 16 different haplotype combinations in the C×E progeny

For the CE2013-okt experiment markers B449/F and Amt241073 were used, resulting in a target region of 2.1 Mb. In this region 175 recombinants were identified in the female parent and 74 in the male parent. After swapping the order of rows on 96-wells plates with selecting recombinants, two recombinants were recovered and paternal genotypes in a 1:1:1:1 segregation (see table 3).

The second batch (CE2014-feb) was screened for recombination with marker PhoTr31222 and Amt241073 resulting in a target region of 1.1 Mb. Based on the pre-screening 52 recombinants were identified in the mother and 11 recombinants in the father, but also 11 double recombinants were identified. After re-analysis of recombinant seedlings on high-quality DNA isolated with the Kingfisher, a high number of false-positives were observed. Final numbers of recombinants are 19 in the female parent and 10 in the male parent. In contrast to Witteveen (2013) a total of six recombinants are identified in scaffold DMB546. This region was previously found to co-segregate with tuber shape.

In batch CE2013-apr 142 maternal recombinants were found, and 31 recombinants in the father (Witteveen, 2013). Given the observation that the C×E population segregates for maturity in a 3:1

ratio, for approximately 25% of the recombinants no phenotypes were available, The final numbers of recombinants are 105 in the female parent and 27 in the male parent (table 1).

	P					♀					♂					♀♂					
	$\frac{AA}{AB}$	$\frac{AA}{CC}$	$\frac{BB}{AB}$	$\frac{BB}{CC}$		$\frac{BA}{AB}$	$\frac{AB}{CC}$	$\frac{AB}{AB}$	$\frac{BA}{CC}$		$\frac{AA}{CB}$	$\frac{AA}{AC}$	$\frac{BB}{CB}$	$\frac{BB}{AC}$		$\frac{BA}{CB}$	$\frac{BA}{AC}$	$\frac{AB}{CB}$	$\frac{AB}{AC}$		
CE2013-apr	217	210	234	229	**890**	50	29	39	24	**142**	9	8	5	9	**31**	4	2	1	0	**7**	**1070**
CE2013-okt	212	281	218	240	**951**	47	51	37	40	**175**	15	12	29	18	**74**	8	4	3	2	**17**	**1217**
CE2014 - feb	176	179	183	186	**724**	11	19	10	12	**52**	2	4	2	3	**11**	1	0	2	8	**11**	**798**
																					3085

Table 3 Results from pre-screening of recombinants from seedlings with flanking markers. Numbers represent initial data including false-positives that were identified later. A total number of 3085 seedlings were screened, although only batch CE2013-apr and batch CE2014-feb were used for recombinant analysis.

Slight deviation from Mendelian expectations were found near the *Ro* locus, between marker B449/F and marker Amt241073. After statistical analysis with Pearson's Chi-square on parental genotypes of batch CE2013-okt, significant gametic segregation distortion was found in the E parent, where haplotype CC was overrepresented ($p < 0.01$). Zygotic selection was not found using a 2x2 Chi-square contingency test. The observation of gametic selection in the male parent was not confirmed in CE2014-feb (P-value > 0.05) and in data from Witteveen (2013) with P-value > 0.05. This deviations were not observed in other batches.

	♀		♂	
	AA	*BB*	*AB*	*CC*
CE2013-apr	0.52	0.48	0.49	0.51
CE2013-okt	0.52	0.48159831	0.45	0.55**
CE2013-feb	0.51	0.49	0.50	0.50

Table 4 Testing for zygotic selection revealed slight deviations from mendelian expectations in batch CE2013-okt with p-value < 0.01. This observation was not found in other batches.

Recombinant screen results

Further fine mapping of the *Ro* locus was done by saturating the target region with HRM markers. Recombinants as identified in batch CE2013-apr revealed cosegregation of tuber shape with markers Asp6678, CatPer20798 and Per20801 (280 kb). Previously Witteveen (2013) reported the fine mapping of tuber shape between markers PhoTr31222 and Cons15896, resulting in a region of approximately 900 kb. However the inclusion of paternal markers Solcap_1, Nslp31336, Nslp11953 and UPF11859 in combination with phenotype information, further reduced the location of the *Ro* locus to 580 kb.

In order to get a high resolution for linkage mapping another set of seedlings was screened for recombination. The CE2014-feb batch was screened with markers PhoTr31222, Asp6678, Per20801 and Amt241073. In this batch no cosegregation between markers Asp6678 and Per20801 was observed, thus providing room for further fine mapping. Between markers Asp6678 and Per20801 five recombination events were identified in the female parent, but only one in the male parent. Phenotypic analysis of tuber shape for these recombinants revealed that the *Ro* locus is located downstream of marker Asp6678 and upstream of marker Per20801. The physical distance between these markers is approximately 280 Kb.

Markers	C			E			CE2014-23	CE2014-57	CE2014-34	CE2014-36	CE2014-74		CE2014-25
PhoTr31222 (10)	a	b		a	c		a	a	b	a	a		b
Asp6678 (19)	a	b		b	c		a	a	b	a	a		b
	R	L		L	R		R	R	R	R	R		L
Per20801 (22)	a	b		b	c		b	b	a	b	b		b
Amt241073 (12)	a	b		b	c		b	b	a	b	b		b
													R
							L	L	R	R	L		
PhoTr31222 (10)	a	b		a	c		a	a	c	c	a		c
Asp6678 (19)	a	b		b	c		b	b	c	c	b		c
	R	L		L	R		L	L	R	R	L		R
Per20801 (22)	a	b		b	c		b	b	c	c	b		b
Amt241073 (12)	a	b		b	c		b	b	c	c	b		b
							Rr	**Rr**	**RR**	**RR**	**rr**		**rR**
							♀	♀	♀	♀	♀		♂
							↓	↓	↓	↑	↓		↑
Phenotypes							Round	Round	Plat	Plat	Long		Round

Table 5 Five informative recombinants were observed in the female parent, whereas only one informative recombinant was observed in the male parent. These recombinants were all found in batch CE2013-feb.

For both batches of recombinants the recombination frequencies are calculated and subsequently the map distances in cM are used to visualize the genetic map with Mapchart (Figure 2). In batch CE2013-apr the genetic distance in the female parent (C) was 9.8 cM and 2.5 in cM in the male parent (E) between markers LS_B446 and 495_499. Because different flanking markers are used for batch CE2014-feb a total genetic distance of 2.5 cM was observed in the female parent and 1.4 cM in the male parent between marker PhoTr31222 and Amt241073. In batch CE2013-apr this corresponds to a genetic distance of 1.87 and 0.74 cM in the female and male parent respectively. Based on analysis of batch CE2014-feb the *Ro* locus is located between marker Asp6678 and Per20801, which resulted in a genetic distance of 0.6 cM in the female parent and 0.125 cM in the male parent.

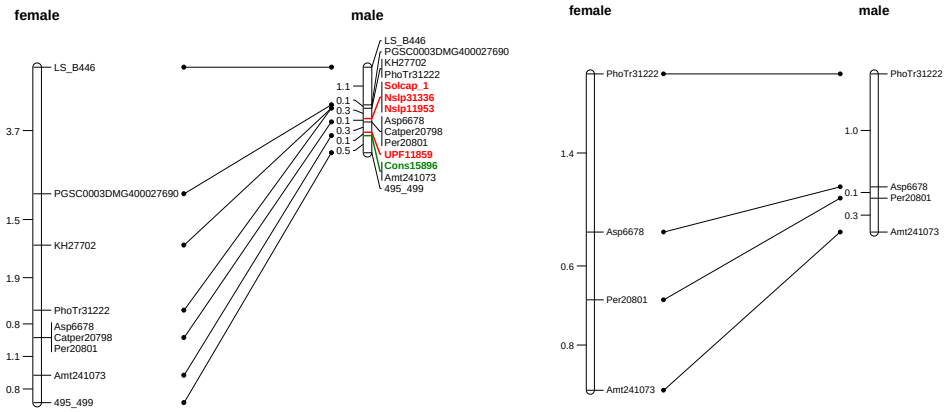


Figure 1 Genetic maps based on batch CE2013-apr (left) and batch CE2014-feb (right) for both female and male parent. The batch CE2014-feb is screened with a limited number of markers, but shows recombination between markers Asp6678 and Per20801.

Potential candidate genes

Results from the recombinant analysis revealed that the *Ro* locus is located in a region of 280 kb between markers Asp6678 and Per20801. Based on the reference genome annotation V4.03, 14 genes reside in this region on chromosome 10, from which seven members of the peroxidase gene family. Besides these peroxidases, four genes with unknown functions are located in this region, one polyprotein was observed and one non-specific lipid transfer protein.

To investigate the involvement of these candidate genes in the regulation of tuber shape gene, RNA-seq datasets as provided by Xu et al. (2011) for both RH89-038-16 and DM1-3-516-R44 were used to get insight into gene expression of these candidate genes. Genotype RH89-38-16 produces oval tubers and genotype DM1-3-516-R44 produces long tubers (Figure 2). Based on expression levels from all DM and RH tissues no RNA-seq support was found for several genes (PGSC0003DMG400045482, PGSC0003DMG400040544, PGSC0003DMG400035649 and PGSC0003DMG400039458).

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	✓

Table 6 List of candidate genes and description, between marker Asp6678 and Per20801. In this region 14 genes were located. RNA-seq support based on all tissue samples from DM and RH genotypes is reported. RNA-seq support was defined as having expression in more than one sample of genotype DM and RH.

Three of the genes were annotated as 'gene with unknown function' and one was annotated as polyprotein. Another gene with unknown function (PGSC0003DMG400020797) was constitutively expressed among all tissue samples in both RH and DM. All peroxidases are only expressed in tuber-specific tissues from both genotypes (data not shown). The non-specific lipid transfer protein was constitutively expressed in most developmental tissues and did only provide some expression in one tuber sample from DM13 516 R44.

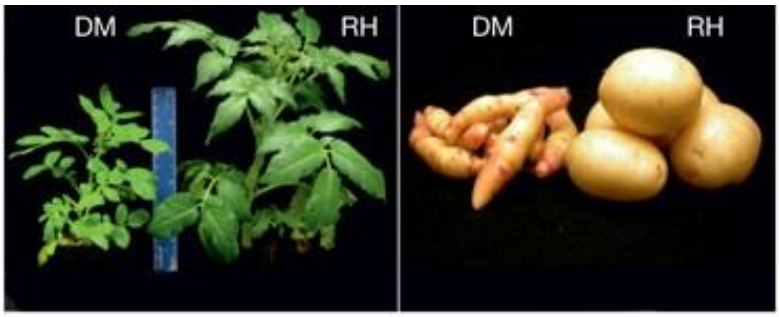


Figure 2 On the left plants from genotype RH and genotype DM are shown. On the right DM long tubers and oval tubers from genotype RH are showed. This image is taken from Xu et al. (2011).

The formation of tuber shape is regulated early in the tuber development, which implies that tissue samples taken in the early stages of tuber development are the most informative. These subsets of informative tissue samples were evaluated from both DM1-3-516-R44 and RH89-039-16 for gene expression. Gene expression libraries from genotype DM1-3-516 R44 did provide information about expression in stolons compared with expression in two tuber samples isolated at different time points from tubers. Gene expression during tuber development was also studied in the genotype

RH89-039-16 in tissue samples consisting out: 1) Stolons without tuberization 2) Tuber tissue isolated after one week of swelling 3) Tuber tissue isolated after four weeks of swelling).

The comparison between expression of genes of RH and DM revealed differences in gene expression in several genes (Figure 3). In these specific subsets the non-specific lipid transfer protein PGSC0003DMG400040954 was up-regulated in only one sample of DM tuber tissue, but not in RH, Another gene, the aspartate aminotransferase PGSC0003DMG400006678 was up-regulated in during stolon-tuber transition in both RH and DM. Among peroxidases two expression profiles were observed: The peroxidase genes in DM were all down-regulated during stolon to tuber development, but in RH peroxidase expression of genes PGSC0003DMG4000020795 and PGSC0003DMG4000020801 was down-regulated and expression was up-regulated of genes PGSC0003DMG400006679, PGSC0003DMG400006680, PGSC0003DMG400006681 and PGSC0003DMG400002079. Two peroxidase genes were slightly upregulated, namely PGSC0003DMG4000020800 and PGSC0003DMG4000020799.

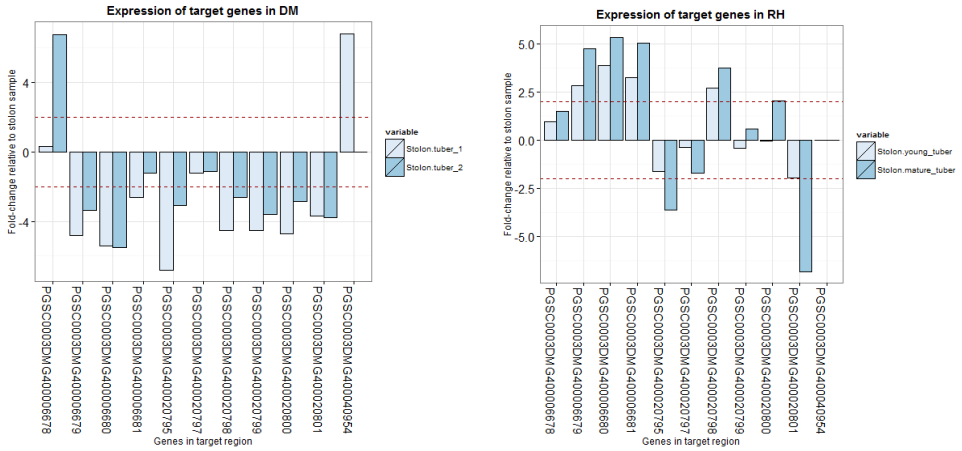


Figure 3 Expression levels of genes located in the target region. For tuber specific tissues expression is visualized for genotype DM (left) and genotype RH (right). Bar plots were made with the R package ggplot2.

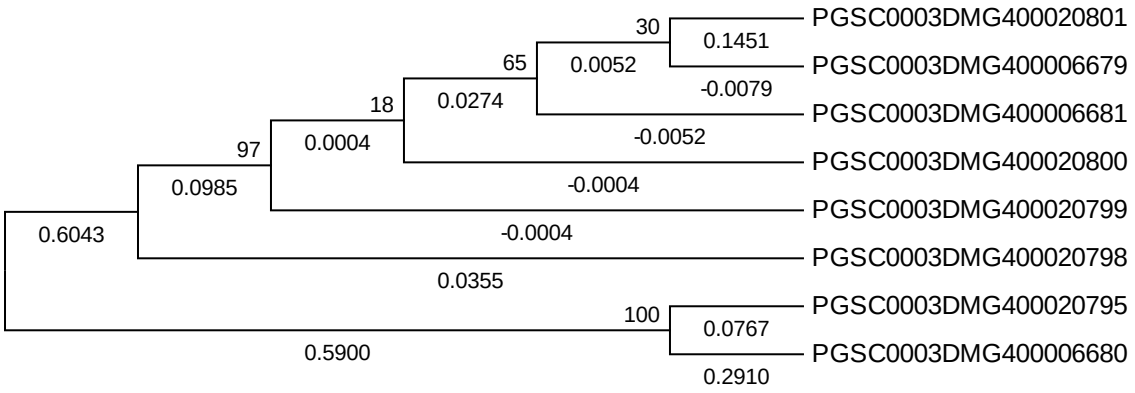


Figure 4 Peroxidases show a limited sequence divergence based on amino-acid sequence. This dendrogram was inferred using the Neighbor-Joining method. The branch length represents the number of amino acid substitutions per site. The reliability of the dendrogram was assessed using a bootstrap test with 500 replicates.

The comparison of peroxidase gene expression among DM and RH resulted in striking differences between expression in DM and RH. To investigate the sequence similarity between the peroxidases, a Neighbor-joining dendrogram was constructed using protein sequences of peroxidases (Figure 4). From this alignment, sequence divergence was calculated as the number of amino acid substitutions per site. This showed that sequence divergence among peroxidases is low.

A search with Gene ontology term GO: 0050789 (peroxidase activity) resulted in 144 peroxidase genes in *Solanum tuberosum*. Expression analysis of these genes related to stolon-tuber transition revealed that many more peroxidase genes were down-regulated in DM stolon to tuber transition (Figure 5).

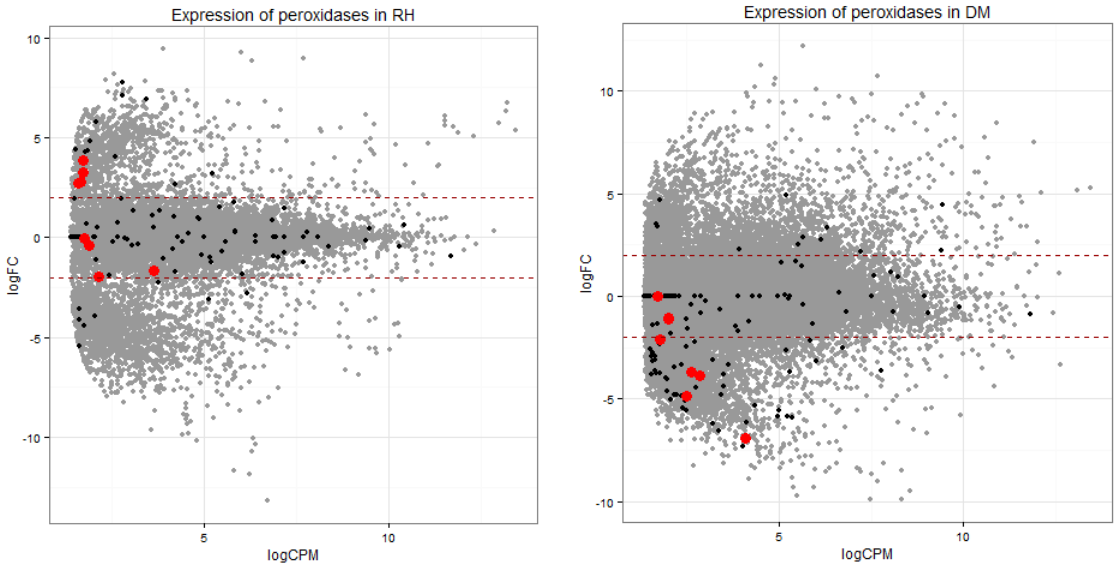


Figure 5 Expression analysis of all 144 peroxidase genes in *Solanum tuberosum* compared from stolon to tuber transition. Light-grey dots are expression values of all genes, whereas black dots represent peroxidase genes. Peroxidase genes located in the target region are visualized as red dots. On the **left** expression differences in RH tissues are represented, which is during transition of non-tuberizing stolons to young tuber tissue. On the **right** expression of DM is compared in non-tuberizing stolons to young/mature tuber tissue.

Phenotypic analysis and heritability analysis

Heritability estimation and phenotypic analysis was done for both phenotypic data from batch CE2013-apr and for the pot experiment with seven clones for each genotype catergory. In both experiment per clone three tubers were evaluated qualitively for tuber shape and quantitatively for their L/W ratio. From batch CE2013-apr 324 genotypes have been phenotyped and for the pot experiment seven clones with three replicate clones have been phenotyped.

The quantitative assesment of tuber shape with the L/W ratio allowed a separation based of two morphological classes (long and round), but also differences were observed within the round tuber morphological class. Previously within the C×E population three haplotypes were identified causing a 1:1:1:1 segregation ratio (van Eck, 1994). However within this category phenotypes could not be unambigiously assigned to each genotype class. This can be seen in figure 6B where a clear difference between long and round tubers is observed. However, when all classes are used an overlap of different classes was observed (Figure 6A).

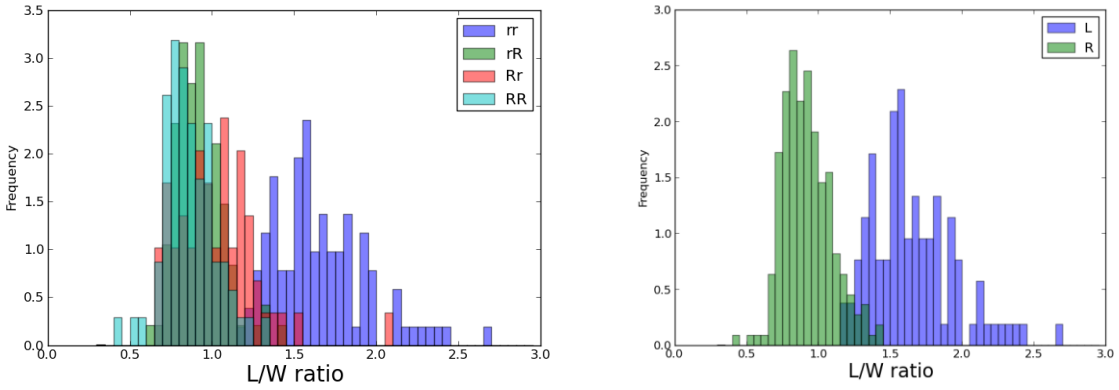


Figure 6 Histogram of tuber shape as represented by length/width ratio for batch CE2013-apr per genotypic classes (left) and per morhpological class (right).

	Genotypic class	Numbers	Mean	Variance		S.E.M.
ro♀ro♂	1	102	1.658	0.0839		0.028685
ro♀Ro	2	95	0.924	0.0206		0.014723
Ro♀ro♂	3	58	1.003	0.0407		0.026498
Ro♀Ro♂	4	69	0.861	0.02465		0.018901
	Morfo	Numbers	Mean	Variance		S.E.M.
ro♀ro♂	1	102	1.652	0.08456		0.028515
Ro/-	2	222	0.921	0.0281		0.0113

Table 7 Mean L/W values for tuber shape based on data from batch CE2013-apr.

	Genotype	observations	Mean	Variance	S.E.M.
ro♀ro♂	1	80	1.702	0.16368	0.045236
Ro♀ro♂	2	49	1.068	0.03801	0.027857
ro♀Ro♂	3	50	0.881	0.04045	0.02844
Ro♀Ro♂	4	63	0.863	0.04197	0.025815
	Visual	observations	Mean	Variance	S.E.M.
ro♀ro♂	1	80	1.702	0.1637	0.045236
Ro/-	2	162	0.931	0.0482	0.017246

Table 8 Mean L/W values for tuber shape based on data from experiment with seven tuber genotypes.

In the C×E the *Ro* locus does segregate in four genotypic classes, therefore a multiple comparison was done of the mean of different genotype classes. With an one-way AOV significant differences were found between all four classes. Subsequently a post-hoc Tukey test revealed that in batch CE2013-apr the mean of rR is not significantly different from the mean of class RR (compressed) and Rr. The same analysis was done for the pot experiment and in this experiment rR was not significantly different from RR (compressed), but different from Rr. In both experiments the rr tubers (long) are different from all other genotype classes.

Pot experiment				Batch CE2013-apr				Van Eck (1994)		
	Genotype	Mean	Tukey		Genotype	Mean	Tukey	Genotype	Mean	LSD
4	Ro♀Ro♂	0.863	**a**	4	Ro♀Ro♂	0.861	**a**	Ro♀Ro♂	**0.8057**	**ab**
3	ro♀Ro♂	0.881	**a**	2	ro♀Ro♂	0.924	**ab**	ro♀Ro♂	**0.8898**	**a**
2	Ro♀ro♂	1.068	**b**	3	Ro♀ro♂	1.003	**b**	Ro♀ro♂	**0.7530**	**b**
1	ro♀ro♂	1.702	**c**	1	ro♀ro♂	1.658	**c**	ro♀ro♂	**1.5112**	**c**

Table 9 Pair-wise comparison of means of different genotypic classes for all experiments with one-way ANOVA using the four genotype classes followed by a Tukey pair-wise comparison.

Variances as used in heritability estimation are calculated by using linear mixed models on phenotypic data from batch CE2013-apr and the pot experiment. This design is an unbalanced design in which all factors were taken as random. The broad sense heritability (h^2) is defined as the proportion of the variance that is due to tuber shape. Based on estimated variance compontents the h^2 based on between clone and within clone variance of batch CE2013-apr was estimated to be 0.92. The proportion of variance explained by the morphological classes is 0.85 and the proportion of variance explained by the genotypic classes is 0.74 (table 3).

Heritability estimation from L/W ratios of the pot experiment resulted in a broad sense heritability of 0.77 and the explained variance by the morphological classes is 0.77. When genotypic classes are used as qualitive factor the explained variance is 64% (table 4). The model for the pot experiment as used for estimating variances was not improved when using the within clone effect and when using the within tuber effect, implying that using replicates of tuber clones does not incease the predictive value of the model. This suggests that the between among clone variance is neglectable.

Model	Component	Variance components	Total variance	h^2
$y_{kl} = \mu + \underline{Cl_{ijk}} + \underline{Tu_{ijkl}} + \underline{e_{ij}}$	σ^2_{Gen} σ^2_e	0.16157 0.0146	0.1763	0.92
$y_{jkl} = \mu + \underline{Morpho_i} + \underline{Cl_{ik}} + \underline{e_{ij}}$	σ^2_{Morpho} σ^2_e	0.27996 0.0146+0.03564	0.33033	0.85
$y_{jkl} = \mu + \underline{Geno_i} + \underline{Cl_{ik}} + \underline{e_{ij}}$	σ^2_{Geno} σ^2_e	0.14248 0.0146+0.03524	0.19242	0.74

Table 10 Linear mixed models as used for the estimation of variance components for estimating heritability values of batch CE2013-apr.

Model	Component	Variance components	Total variance	h^2
$y_{kl} = \mu + \underline{Cl_{ijk}} + \underline{e_{ij}}$	σ^2_{Gen} σ^2_e	0.16995 0.0505	0.22045	0.77
$y_{jkl} = \mu + \underline{Morpho_i} + \underline{Cl_{ik}} + \underline{e_{ij}}$	σ^2_{Morpho} σ^2_e	0.29420 0.0504 + 0.03712	0.38172	0.77
$y_{jkl} = \mu + \underline{Geno_i} + \underline{Cl_{ik}} + \underline{e_{ij}}$	σ^2_{Geno} σ^2_e	0.15107 0.03446 + 0.0504	0.23593	0.64

Table 11 Linear mixed models as used for the estimation of variance components for estimating heritability values of the pot experiment.

Variant detection

Sequence variants in the target region were identified using the GATK haplotype caller. Before SNP calling, sequence reads were processed and checked for quality with FASTAQC. Based the quality assessment, subsequently low quality and duplicate reads are removed. After pre-processing sequence reads, variant detection was performed using the GATK haplotype caller with local realignment round Indels. The variant detection from the target region of 1,040,00 bp, resulted in a total of 7652 variants in parent C and 14793 variants were observed in parent E (table 8).

The calculation of nucleotide diversity revealed that nucleotide diversity (Π) was lower in the C parent than in the E parent. Nucleotide diversity was 2×10^{-3} in the C parent, but 5×10^{-3} in the E parent. This lower diversity in the C parent is also seen in figure 7 where large tracts of homozygosity are observed in the C parent. Local tracts of homozygosity were identified at the end of scaffold DMB385 and also in the beginning in DMB446. However based on this analysis, it seems that DMB546 is heterozygous compared to other regions.

	indels	Homozygous indels	Heterozygous indels	SNPs	Homozygous SNPs	Heterozygous SNPs	Total
C	*769*	*429*	*340*	*6883*	*2996*	*3887*	*7652*
E	*1401*	*363*	*1038*	*13392*	*8297*	*9629*	*14793*

Table 12 Results from the variant detection using GATK haplotypecaller. For both the C parent as well as the E parent sequence reads were processed and used for variant detection. In the E parent more variants were identified than in the C parent.

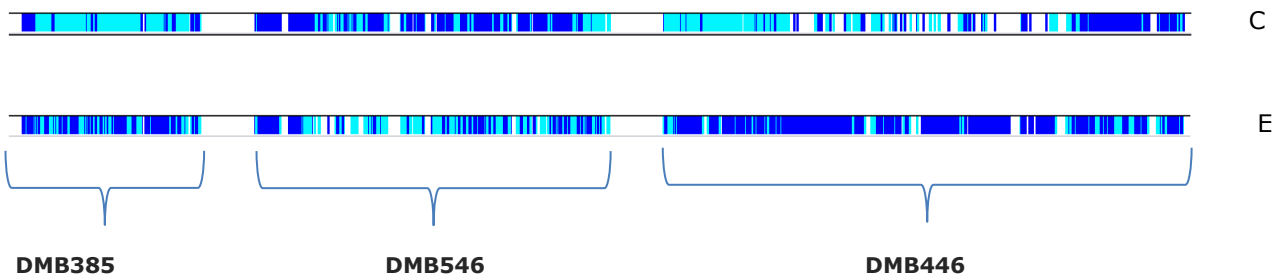


Figure 7 Results of SNP analysis visualized in IGV by using a VCF file. Local tracts of homozygosity are visualized as light-blue in the upper bar (C) and in the lower bar (E). In this figure can be seen that SNP density in scaffold DMB546 is higher than in scaffold DMB385 and DMB546.

Marker design

For optimal marker design the use of haplotype information could be useful. For existing markers that identify the three haplotypes that occur in the C×E backcross, sequence information was used to check whether it would be possible to better predict whether markers are polymorphic in both parents. An example of the combination of haplotype information with HRM analysis output is shown in Figure 8 where only one SNP makes the difference between two different haplotypes. More examples are given in Supplementary information (IV). A list with all markers and location on the genome combined with segregation data, are given in Supplementary information V . The use of haplotype data is necessary to predict whether markers are polymorphic in both parents in situations where a larger number of variants occur in one amplicon, although when only a limited amount of SNPs are used (as in figure 8) this information is less useful.

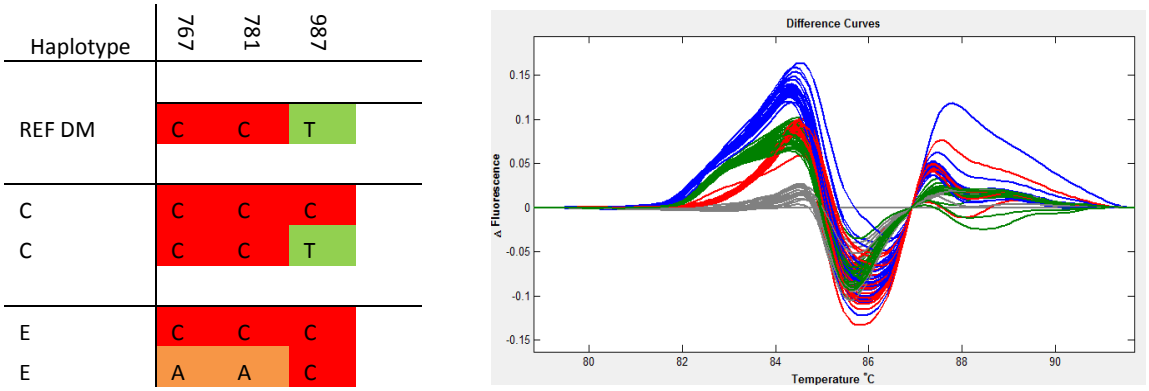


Figure 8 Comparison of haplotype information with lightscanner output. Four patterns are distinghuished in the lightscanner output (right). In the sequence information three haplotypes are identified (left), resulting in four possible progeny classes.

Application of HRM technology

For HRM analysis several different compounds were available in the laboratory which were compared for performance. Common practise is using LCgreen in combination with the high fidelity DNA polymerase Phire. However other fluorescent dyes (EvaGreen™) and DNA polymerases (Dreamtaq) were also available in the laboratory for testing.

Several markers were tested for the performance with these fluorescent dyes, as well as the effect of using the two different DNA polymerases. In general it is noted that EvaGreen™ has limited effect on the melting temperature than LCgreen and assays using this dye were performed at a lower PCR melting temperature. The comparison of different fluorescent dyes while using Phire DNA polymerase shows that subtle differences were observed between the two analyses. However in both cases the four groups can clearly be distinguished (Figure 9).

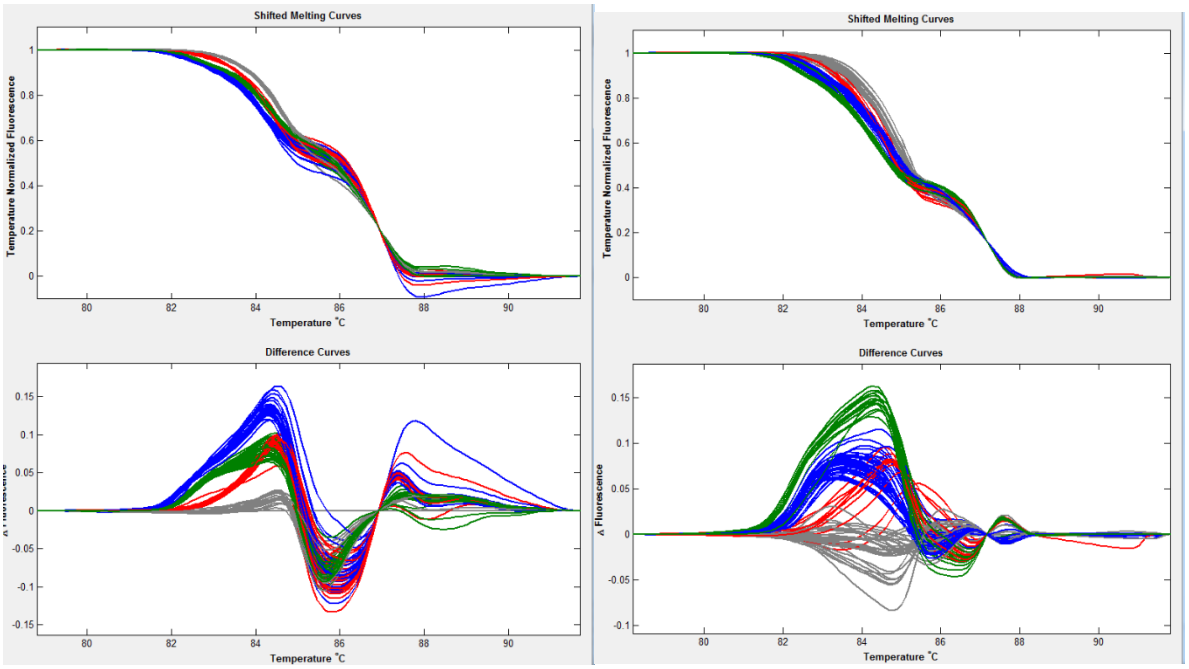
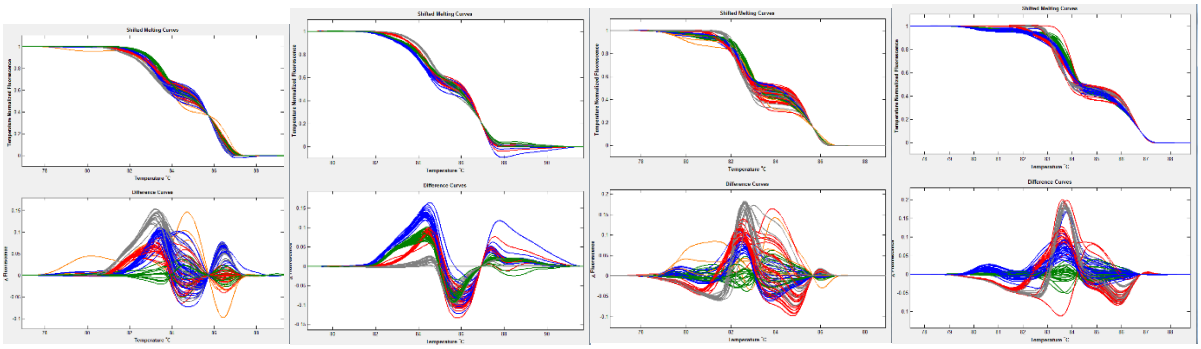


Figure 9 Comparison of lightscanner patterns of 96 C×E seedlings with marker Amt241374 with EvaGreen (left) and on the right patterns of marker Amt241374 with LCGreen (right).

Secondly the different DNA polymerases were compared with using EvaGreen as fluorescent dye. This showed that using Dreamtaq provided less robust and reliable assays as compared to the hot start Phire enzyme (Figure 10). Based on earlier experience the use of a non-hot start DNA polymerase causes in not well-optimized problems with the occurrence of primer-dimer. Assays with using marker Amt241374 and PhoTr31222 were not optimized for non-hot start polymerase assays.



A B C D

Figure 10 A) Assay with marker Amt241073 and Dreamtaq polymerase and EvaGreen, B) assay with marker Amt241073 and Phire polymerase and EvaGreen, C) assay with marker PhoTr31222 and Dreamtaq polymerase and EvaGreen D) assay with marker PhoTr31222 with Phire polymerase and EvaGreen.

Integration of scaffold DMB773 and synteny analysis

A genome-wide association study (GWAS) revealed a significant association with marker solcap_snp_c2_56344 located on unanchored scaffold DMB773 (LOD of 9.1). This suggested that scaffold DMB773 should be inserted between scaffold DMB385 and DMB546. The length of this scaffold is approximately 133 kb. To investigate whether scaffold DMB773 was located in this region, several approaches were used.

The first approach that was used was by considering function of genes on the scaffolds. Within scaffold DMB385 a cluster of non-specific lipid transfer proteins was found, which extended to the first 70 kb of DMB773 suggesting that this scaffold is located downstream of DMB385. Another approach was to manual align BAC-end sequences which were located in the target region consisting out of scaffolds DMB385, DMB773 and DMB546. From this analysis five BAC clones were found in RH that connect DMB385 to DMB773 (table 12). However no BAC clones were found that connect scaffold DMB773 to DMB546 allowing for a gap between these two scaffolds.

BAC clones	*Start scaffold*	End scaffold
RHPOTKEY043_D20	*385*	773
RHPOTKEY136_F15	*385*	773
RHPOTKEY075_M12	*385*	773
RHPOTKEY201_I24	*385*	773
RHPOTKEY125_L06	*385*	773
RHPOTKEY079_K19	*546*	446
RHPOTKEY039_J21	*546*	446
RHPOTKEY048_H12	*546*	446
RHPOTKEY026_H09	*546*	446
RHPOTKEY100_N15	*546*	446
RHPOTKEY045_I05	*546*	446
RHPOTKEY084_G23	*546*	446

Table 13 Gap-spanning BAC clones between scaffolds identified with manual alignment using BAC ends annotated by Sharma et al. (2013).

In order to explore if the gap between DMB546 and DMB773 was significant, BAC-ends of RH and DM as well as fosmid ends from DM are aligned to the target scaffolds using the BLAST+ suite. Significance of these hits was defined on two criteria: First the BAC-end hit to the reference genome had to be specific, and secondly the length of the region between forward and reverse BAC-ends hits, had to be between 30 kb and 200 kb.

This analysis was done with a stepwise approach to save computation time. First, all BAC-sequences were aligned to the target region and significant hits were selected on coverage and identify. After selection of significant hits to the target region, these hits were checked for specificity with the whole reference genome 4.03. The final step in the analysis was to check if hits were located between 30 kb and 200 kb on the genome. After this selection a total of 60 unique BAC clones remained (table 13).

Based on this selection no BAC clones could be found that connect DMB773 to DMB546. From the total of 60 BAC clones only five BAC clones were found that connect DMB385 to DMB773 and two BAC clones were found that connect DMB546 to DMB446. For DM fosmid-ends the same analysis has been done, but the final selection was based on the criteria that fosmid clones should have a length between 15 and 30 kb. This resulted in 65 fosmid clones aligning to the target region, but none of these was found spanning different scaffolds.

	385	773	546	446	385-773	773-546	546-446
RH BAC	10	X	1	14	5	X	x
DM BAC	13	X	3	22	x	X	2
Fosmids	25	3	10	27	x	X	x

Table 14 BAC and fosmid clones aligned to reference scaffold. A total of 60 BAC clones were found unique for the target region. From RH BACs five clones were found that connect DMB385 and DMB385. In the DM BAC library no BAC clones were identified that connect DMB773 to other scaffolds. Besides BAC clones also 65 Fosmid clones were found that align to the target region. None of these fosmid clones connect different scaffolds.

The BAC and fosmid -end scaffolding did not reveal any clones that connected DMB773 to DMB546, however these analysis revealed a significant connection of DMB385 to DMB773. In order to explore whether no extensive gap was located between these scaffolds, a synteny analysis was done with *Solanum lycopersicum*. This synteny analysis showed extensive macro-collinearity between chromosome 10 of *Solanum lycopersicum* and chromosome 10 of *Solanum tuberosum*. Hence, when looking at the target region, extensive micro-collinearity was found between the whole target region implying that the gap between DMB773 and DMB546 is neglectable (Figure 11).

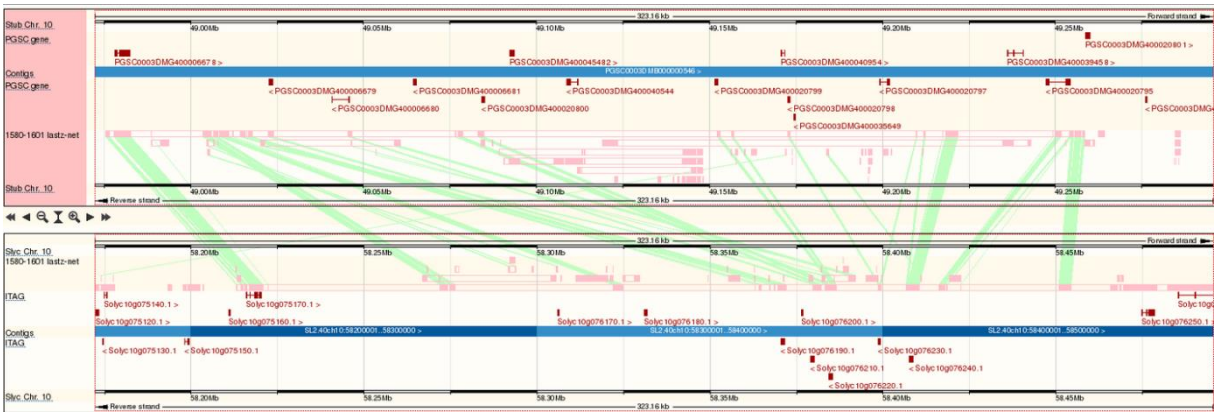


Figure 11 Micro-collinearity of candidate genes between potato and tomato as shown in the Ensembl genome browser (http://plants.ensembl.org/index.html).

Potatochr 10 Start	Potatochr10 End	Gene description	Gene name	Tomato chr10 Gene	Tomato chr10 Description	Tomatochr10 Start	Tomatochr10 End
49061577	49066134	Aspartate aminotransferase	PGSC0003DMG400006678	Solyc10g075170.1;Solyc10g075170.1.1	Aspartate aminotransferase	58215723	58220610
49090229	49095871	-	-	-	-	58321208	58324732
49101873	49109892	Peroxidase 2	PGSC0003DMG400006679	Solyc10g076220.1.1;Solyc10g076220.1	Peroxidase 1	58380340	58386549
49122699	49123004	-	-	-	-	58391098	58391394
49159542	49162237	-	-	Solyc10g076220.1.1;Solyc10g076220.1	Peroxidase 1	58380242	58386624
49235199	49236530	Cationic peroxidase 1	PGSC0003DMG400020799	Solyc10g076190.1.1;Solyc10g076190.1	Peroxidase 1	58370523	58371838
49251385	49252734	-	-	-	-	58175885	58177778
49252784	49253068	-	-	Solyc10g075140.1.1;Solyc10g075140.1	Unknown Protein (AHRD V1)	58175574	58175858
49254074	49254508	Non-specific lipid-transfer protein 2	PGSC0003DMG400040954	Solyc10g076200.1;Solyc10g076200.1.1	Non-specific lipid-transfer protein	58376230	58376658
49262652	49264176	-	-	Solyc10g076240.1.1;Solyc10g076240.1	Peroxidase 1	58407295	58408809
49330791	49338088	Cationic peroxidase 1	PGSC0003DMG400020795	Solyc10g076240.1.1;Solyc10g076240.1	Peroxidase 1	58407399	58408786
49331840	49338019	Cationic peroxidase 1	PGSC0003DMG400020795	-	-	58416639	58418547

Table 15 List of genes and corresponding synteny hits. Micro-collinearity was found between potato candidate genes and tomato genes. Where a dash is shown no annotation was available, but synteny hits are discovered in regions without genes.

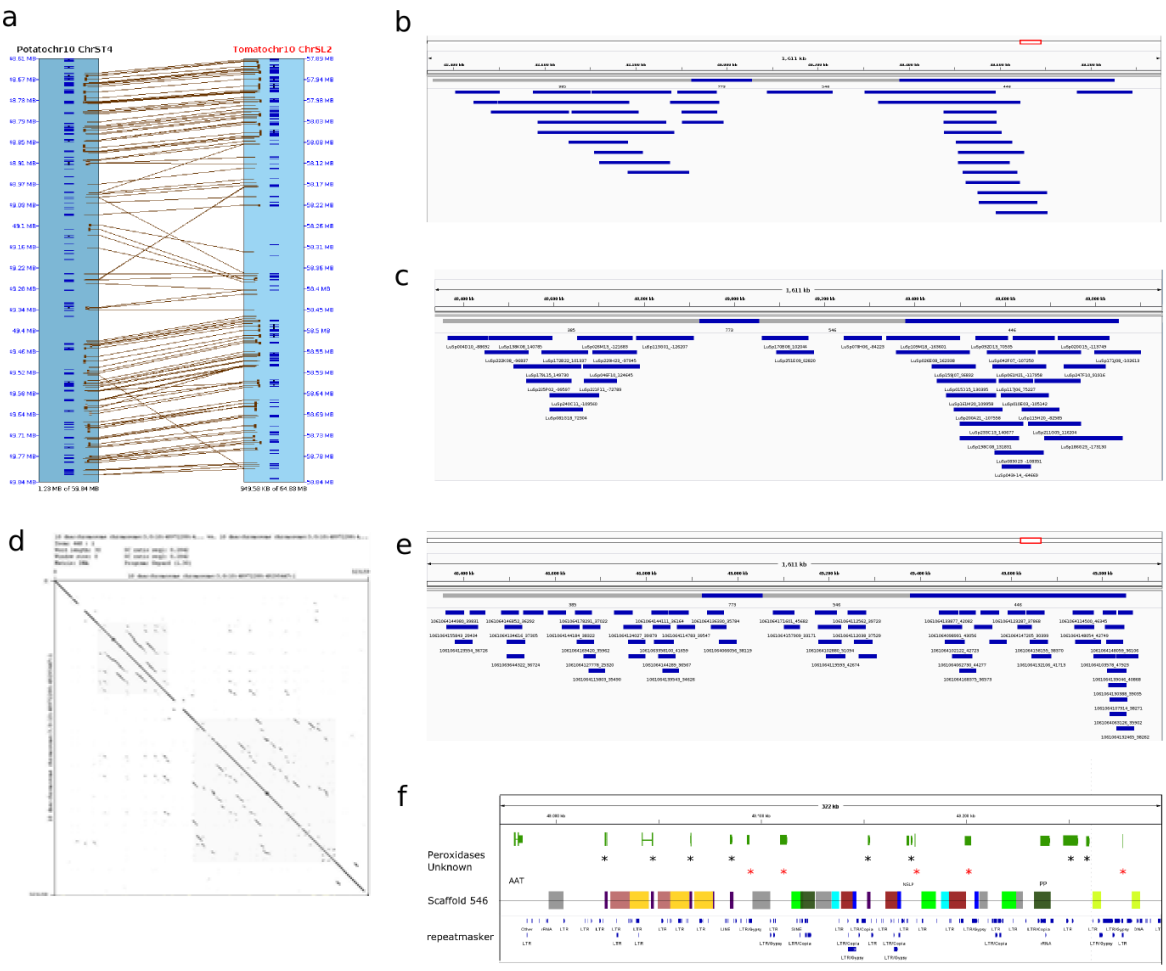


Figure 12 a) Synteny reconstruction of target region (DMB385, DMB773, DMB546, DMB446). This reconstruction shows that micro-collinearity connects DMB773 to DMB385 and DMB546. **b)** DM BAC clones aligned to the target region. Most BAC clones align to the same scaffold, although several BAC clones connect DMB546 and DMB446. **c)** RH BAC clones aligned to the target region. Again most BAC clones align to the same scaffold, but several clones connect DMB773 and DMB385. **d)** Dotplot comparison of DMB546 against DMB546, showing repeated regions. **e)** Fosmid clones aligned to the target region. All fosmid clones align to one scaffold and do not connect different scaffolds. **f)** Repeat annotation of DMB546. Repeats of peroxidases are shown and also repeatmasker annotation was shown.

Chapter 3. Discussion

Potato tuber shape is mediated by many factors, but the overall tuber shape is to a large extend mediated by a single locus which was found to be located on chromosome 10 (van Eck, 1994). In the past decade, several studies have identified QTLs for potato tuber shape, but none of these studies have detected the major QTL on chromosome 10 for tuber shape (D'hoop et al. 2014; Śliwka et al. 2008). This absence of association may be due to the extent of linkage disequilibrium on chromosome 10 combined with the limited marker density of these studies.

Nevertheless, an association study with the use of a 18k SNP-genotyping array (Vos et al. 2014) has revealed SNPs with a significant association with tuber shape to DMB385 (Vos, unpublished data). As already demonstrated by d'Hoop (2014) and Kloosterman (2008) the combined effort of QTL mapping with linkage studies gives more confidence and reliability in order to identify candidate trait loci involved in complex traits.

Preliminary QTL mapping showed that the location of the *Ro* locus was at least one Mb downstream the most significant association. Further refinement in Witteveen (2013) confined the *Ro* locus to a region of 910 kb on chromosome 10, between scaffold DMB385 and DMB446. In this study the Ro locus was mapped to a target region with length of approximately 280 kb, which was based on recombination events between marker Asp6678 and marker Per20801.

Recombinant screen

Several batches of seedlings from the C×E diploid experimental mapping population have been analysed for recombination events and scored for tuber shape after senescence. In order to distinguish recombinant seedlings from non-recombinant seedlings a pre-screening was done using flanking markers of the target region. After screening only recombinant individuals were grown until senescence of plants. Initial screening on recombinant seedlings from batch CE2013-apr with additional markers mapped the *Ro* locus between DMG11953 and DMG11859. This refinement was based on only four paternal recombinants, though these recombinants were not contradictory.

In this batch a large region of 280 kb on DMB546 is cosegregating with potato tuber shape, leaving no opportunities for further refinement of the *Ro* locus. Subsequently another batch of 2000 seedlings were grown and analysed for recombination, but no recombinants were discovered because of human error. The third batch of recombinants was sown in February 2014 and showed recombination events between cosegregating markers of Witteveen (2013). After phenotyping a total of six informative recombinants were found that all suggested that the *Ro* locus was located between marker Asp6678 and Per20801.

Phenotypes of informative recombinants showed that only one long tuber recombinant was observed, which genotype implied that the *Ro* locus was located downstream of marker Asp6678. Besides these recombinants, five recombinant genotypes were observed with a round tuber phenotype. Previous analysis of phenotype information of batch CE2013-apr showed that the difference between Ro♀ro♂, ro♂Ro♀ and ro♀ro♂ progeny cannot be scored unambiguously (Figure 6).

This ambiguity in phenotyping suggests that clones within the round tuber class may be less informative for linkage mapping than clones within the long tuber class.

However, despite the higher change of ambiguous phenotypes, critical recombinants within the round tuber class are informative for linkage mapping. Recombinants were observed, which certainly have inherited the Ro♂ and clearly have a compressed phenotype. This coupled with the observation that they had either the maternal Ro♀ allele at marker Per20801 (recombinant CE2014-34) or the maternal Ro♀ allele at marker Asp6678 (recombinant CE2014-36) implies that these recombinants are fully informative for mapping the *Ro* locus to this region. This line of reasoning can be used to show that recombinants which certainly inherited the long ro♂ allele and have a round phenotype, should have the maternal Ro♀ in their genotype (e.g. recombinants CE2014-23, CE2014-57 and CE2014-25).

Sequence divergence

As demonstrated by Uitdewilligen et al. (2013) nucleotide diversity estimates are low on chromosome 10 when compared to other chromosomes of potato, because of possible selection for overall tuber shape. This would suggest that in the vicinity of the *Ro* locus haplotypes are more conserved and nucleotide diversity would be lower than normal. This low level of nucleotide diversity was confirmed based on a SNP analysis of re-sequencing data from the C and E parents. In this analysis large tracts of almost complete homozygosity were found at the end of scaffold DMB385 and the beginning of scaffold DMB446 in the female parent (Figure 7). It is striking that those haplotypes are similar in DMB385, but are not in DMB546. The higher sequence divergence on scaffold DMB546 may be explained with the observation of repetitive regions within scaffold DMB546 with a high sequence similarity (Figure 11), which will result in a higher amount of miss-aligned reads. A higher amount of misaligned reads will subsequently bias the amount of sequence variations that is observed.

Development of markers

High-resolution melting (HRM) markers were developed to saturate the genetic map to locate recombination events. Initially marker development was done with a random approach using intron-spanning amplicons with a maximum length of 500 bp. In general this method is simple and reliable to detect the four haplotype combinations that occur within the C×E mapping population. Nonetheless the random development of markers proved to be difficult on scaffold DMB385, DMB773 and the beginning of DMB446 because of homozygosity in the C parent.

As a consequence of these homozygous regions, problems occurred with marker development as only recombinants in the paternal genotype could be scored and no markers could be developed that allowed distinguishing the two haplotypes in the C parent. To develop markers which are polymorphic in both parents and allow detection of the three alleles in the C×E population, the use of sequence information was explored. Based on this sequence information it is theoretically possible to predict how markers will perform, but in practice this was not easy. Raw sequence information does not always correctly display significant SNPs, because of indels located in the reference genome. Therefore identifying regions, with raw sequence information, which are suitable

for marker development is cumbersome. In order to circumvent that problem, a SNP calling was done on raw reads with GATK, which allowed selection of high-confidence SNPs for marker development. Despite the usefulness of variant information for the development of markers, in order to fully predict whether markers will detect three alleles, haplotype information is needed, which in diploid potato may be done with a read-backed phasing approach using GATK haplotypecaller (DePristo et al. 2011).

From the analysis of sequence variation several advantages could be noted: 1) It is possible to detect homozygous tracts of sequence, where no markers should be developed, and 2) it is possible to find regions which have a low density of SNP, but polymorphic in both parents to allow successful marker design. Disadvantages are that this approach relies on the quality of sequencing and alignment, where it may be possible that the C×E sequence information may be informative for SNP calling, however, because of lower coverage not suited to haplotype phasing.

Baldwin et al. (2012) discussed automated HRM marker design in onion which is a highly heterozygous species like potato and reported a validation rate of 56% when looking at polymorphism in the F1 originating from two double haploid lines. However when the F1 is inter-crossed, then only 9% of the markers was polymorphic in the F2. This is due to the fact that that homozygous products are not easily distinguished. The C×E population can be regarded as a backcross, originating from two non-inbred parents. Hence, within progeny of this mapping population in general three alleles will be available, and only one homozygous product will be generated. This allows a better separation of HRM melt patterns.

Integration of scaffold DMB773 and synteny analysis

Previously unanchored scaffold DMB773 was inserted between DMB385 and DMB546 with manual BAC-end scaffolding. From this analysis it was revealed that scaffold DMB385 was connected with scaffold DMB773, but no link was observed between either scaffold DMB385 and DMB546 or DMB773 and DMB546. Subsequently in order to investigate whether scaffold spanning BAC clones could be retrieved, all BAC-ends from DM and RH were aligned to the target region. But again no hits were observed which span DMB773 and DMB546. The lack of BAC clones connecting these scaffolds together can be due to that either no BACs have been sampled for this target region, or that BAC-ends were not aligned because of repetitive end sequences (Thesis Theo Borm). This is supported by the sequence analysis where large repetitive regions have been observed in scaffold DMB546.

Because the BAC-end analysis did not reveal BAC clones that connect scaffold DMB773 and DMB546 an synteny analysis was done with *Solanum lycopersicum.* This synteny analysis showed micro-collinearity between the target region *including* scaffold DMB773 and the tomato genome. This suggests that the gap between DMB773 and DMB546 is either small or not existing. However because the recombinant screen has identified the location of the *Ro* locus within scaffold DMB546 this gap is not important.

From the synteny analysis, information about the target genes on scaffold DMB546 is revealed. In the corresponding syntenic region on *S. lycopersicum* eight genes are located. From these genes

six of them are peroxidases and one is known as an Aspartate aminotransferase. This suggests that during the divergence of *Solanum tuberosum* and *Solanum lycopersicum* this gene cluster has been conserved.

Tuber shape from a molecular perspective

From a developmental perspective the formation of tuber shape consists out of two processes as demonstrated by Xu et al. (1998a, 1998b). First stolon elongation is observed, followed by radial expansion causing tuberization at stolon tips. This sequential development of potato tuber development does provide clues about where and in what developmental stage potato tuber shape is determined. Observations during the growth of potato tubers indicate that potato tuber shape initiated early in the tuber development and that after a few weeks of growing the difference between long and round tubers is already visible. Other studies did reveal that longer potato tuber genotypes have a more narrow pith as compared to genotypes with round tubers (Taj and miserner, 1994).

Previous studies have identified several mutants which display elongated tuber phenotypes. The knock-down of a steroid dehydrogenase transcript (GANGLY) resulted in elongated phenotypes (Bachem et al. 2001), and also (Morris et al. 2006) modified the isoprenoid biosynthesis pathway which resulted in elongated tubers. Activation-tagging of cv. Bintje did also show an elongated phenotype, probably due to pleiotropic effects of inactivating a major gene (Aulakh et al. 2014).

In the latter report however a two fold increase was seen in the number of eyes per tuber in elongated tubers and also a decrease in the tuber eye depth. As shown by Li (2005) the locus for eye depth is located on chromosome 10 closely linked to the Ro. The increase in the number of eyes for longer tubers suggests that in this case tuber shape is formed early in the development of the tubers, because tuber eyes are already formed during stolon elongation. An laborious, but potential useful experiment could be to count eyes for long and round tubers in the C×E population to see whether there is a difference in the number of eyes.

From literature about cell elongation and organ shape modulation, it is known that two processes influence organ growth and size, namely cell proliferation and cell expansion (Powell and Lenhard 2012) and that in general growth of plant organs is more due to cell expansion than cell division. These processes are responsible for the size and growth of plant organs, but to achieve directional growth, cells undergo anisotropic expansion. Anisotropic expansion is due to turgor pressure, and non-uniform turgor pressure will cause elongation of cells. This anisotropic expansion is closely related to cell wall synthesis and cell wall loosening (Wolf et al. 2012). When the direction of cell growth is longitudinal, cellulose microfibrils are placed at a right angle of the growth axis and cause non-uniform turgor, which result in cell elongation (McFarlane et al. 2014). This process in characterized by an interplay of many different factors which is described in detail in Wolf et al. 2012.

Several reports illustrate the potential of H_2O_2 and peroxidases to increase cross-linking of cell wall components which increases rigidity of the cell wall and will prevent radial swelling (Pereira et al. 2011; Vanholme et al. 2010; Oosterveld et al. 2000; Lamport et al. 2011). On the contrary other

reports suggest that peroxidases could convert O_2 and H_2O_2 and then by peroxidase activity cause cell-wall loosening (Chen and Schopfer 1999; Schopfer 2002) leading to elongation of cells.

Recently a report suggested that a repression of peroxidases under the influence of a MYB-like transcription factor (KUA1) which affected cell size in *A. thaliana* (Lu et al. 2014). Under the influence of KUA1 a down-regulation of peroxidases resulted in higher cell expansion rates and up-regulation resulted in inhibition of cell expansion. Moreover Lin et al (2014) suggests that peroxidases play a role in turnip shape formation in the genus of the *Brassicaceae*, where peroxidases are associated with the phenylpropanoid biosynthesis. However because the prenylpropanoid synthesis pathway produces a wide range of metabolites, involved in taste, flavour and several other plant characteristics (Vogt 2010), this could also be explained by a difference in taste, rather than shape.

In the target region of the *Ro* locus seven peroxidase genes were identified, which suggest that peroxidase genes may influence the difference between round and long potato tubers. Gene expression from tissue samples datasets provided by Xu et al. (2009) was analysed for expression for genotype DM which produces long tubers, and for genotype RH which produces oval tubers (Figure 2). During stolon to tuber transition in DM all peroxidases are down-regulated, where in RH some peroxidases were down-regulated and some were up-regulated. Other genes in the target region (PGSCDMG40954 and PGSCDMG6678) are also expressed in other developmental tissues than tuber-specific tissues, which given that tuber shape is mediated in an early stage of the tuber development are not likely candidate genes. Despite this striking difference in peroxidase gene activity between these genotypes, it should be noted that expression was not analysed within the same genetic background. Moreover it is not exactly known whether these tissue samples are comparable in terms of developmental time points.

Repeats in DMB546

Peroxidase genes are known for their tendency to cluster in the genome. In both rice and *A. thaliana* reports show that peroxidase genes are not homogenously distributed among the genome (Welinder et al. 2002). Based on GO term annotation 144 genes were found with a peroxidase activity. As described above, within scaffold DMB546 a cluster of peroxidases is located. In this cluster limited sequence divergence was observed based on amino-acid sequences. Based on a dotplot comparison of DMB546 two regions are observed that are repeats of each other (Figure 11). The first repeat region consists out of four repeats of peroxidases. The second region consists out of a region where also four repeats of peroxidases are located. This dotplot image is made with conservative word size of 30, meaning that these regions are quite similar. These repeated regions may cause problems with development of markers in this target region, but the availability of SNP information for DMB546 could support the development of markers.

Because peroxidases are clustered within this region and clear signs of repeats are found, this may suggest that copy-number-variations (CNVs) are causative for the elongated tuber shape. This was also suggested in Lin et al. 2014, where CNVs in peroxidase genes were associated with turnip shape formation. Although a clear cosegregation is observed with the peroxidase cluster, this does not imply that CNVs of peroxidases in potato are causative to the elongated tuber phenotype. To

investigate which (peroxidase) gene(s) may be responsible for the elongated tuber shape, further fine mapping is needed. It could be argued that candidate gene approaches are also well-suited to find this gene, but gene expression profiling will be more difficult because of repeats and the little sequence divergence between genes.

As of now it is not entirely clear whether up-regulation or repression of peroxidase gene activity will result in longer or rounder tuber phenotypes. Another simple candidate gene approach to test whether peroxidases have an effect on tuber phenotypes would be to transform recessive long tubers with an over-expression vector of a peroxidase gene in order to see whether this will result in round tubers. Likewise, homozygous round tuber genotypes could be transformed using an over-expression vector of a peroxidase gene to see whether this will result in a longer tuber shape.

Heritability of tuber shape

Potato tuber shape is regarded as a qualitative trait, in which one locus with multiple alleles explains the majority of the phenotypic information. To investigate what the heritability of tuber shape is, variance components associated with different factors were estimated. For estimation of heritability phenotypic values of batch CE2013-apr and the pot experiment were used.

First the broad sense heritability was calculated by using clone effect and within clone effect. Based on this analysis the h^2 for tuber shape 0.92 in batch CE2013-apr and 0.77 in the pot experiment. When h^2 of batch CE2013-apr (0.92) is compared with an estimate (0.8) by Van Eck (1994), a significant difference in value is observed. This difference can be explained by the number of replicates that is used in estimating variance components from batch CE2013-apr, where with more replicates the variance between samples is decreased. This h2 of 0.9 is also reported for tuber shape based on data from one growing season in d'Hoop et al. (2014).

Subsequently the proportion of variance explained by the visual classification was calculated for both experiments. For batch CE2013-apr it is shown that 85% of the phenotypic variance can be explained by overall tuber shape, and that 4% of the variation was explained by minor genetic factors. Based on four genotypic classes of the *Ro* locus 75% of the phenotypic variation was explained, and 7% by minor genetic factors. The difference between these estimations based on morphological classes and genotypic classes is probably due to phenotype miss-classifications in the 'round' phenotypic class. These miss-classifications are possible giving the overlap between tuber shape distributions of different round genotypic classes within the 'round' phenotypic classes.

In the pot experiment 77% of the variation in tuber shape was explained by the visual classification, and 64% was explained by the four different genotypic classes. The reason that estimates from the pot experiment were lower than from data on recombinants from the CE2013-apr, is because in this experiment a low number of clones are used for each genotype category. These heritability estimates are less reliable, because only a limited number of clones are used.

Finally these heritability estimates were compared with estimated from Van Eck et al. (1994). In the latter study a higher proportion of the phenotypic variation was explained by genotypic classes compared with using morphological classes. However, this was not confirmed in both batch CE2013-apr and in the pot experiment.

Multiple alleles for tuber shape

Phenotypic assessment of L/W ratio did confirm the conclusion of van Eck et al. (1994) which demonstrated the occurrence of multiple alleles at the Ro locus. Conversely, when looking at phenotypic data from batch CE2013-apr the conclusion is that ro♀Ro♂ is not significantly different from Ro♀ro♂ and RR, but in data from the pot experiment and data from Van Eck (1994) the conclusion is reached that only ro♀Ro♂ is not significantly different from Ro♀Ro♂. When looking at distribution of phenotypic effects (Figure 6) this suggests that there is a small difference between ro♀Ro♂ and Ro♀Ro♂, which is confirmed in all three experiments. In both experiments the ro♀ro♂ is significantly different from all other genotypic classes. The difference between Ro♀ro♂ and ro♀Ro♂ coupled with the fact that both parents have the long allele ro in common, is explained by the occurrence of two different Ro alleles in the C×E progeny. This implies that the maternal allele (Ro♀) has a more elongating effect as compared to the paternal R allele (Ro♂), which is opposite to the conclusion of Van Eck (1994). This difference may be caused with artefacts of the RFLP mapping used by Van Eck (1994).

Although the effect of different alleles of the *Ro* locus is ambiguous over different studies, the observation of multiple alleles for tuber shape within the experimental population C×E is still valid.

In both diploid as tetraploid potato genotypes, potato tuber shape is shown to be a quantitative trait displaying continuous trait variation. This suggests a polygenic trait, however by multiple combinations of alleles most of the phenotypic variance is explained. This may be the reason that early studies into the inheritance of potato tuber shape (Jong and Burns 1993) did not postulate a clear genetic model.

The *Ro* locus is an example of a trait that is for a large extent qualitatively mediated by a single gene, however shows a quantitative phenotype within cultivated potato germplasm. Despite the observation of multiple significant regions in the potato genome by means of association mapping, this trait is a clear example where the model of (Sirks 1929) is applicable, showing that multiple alleles of a particular gene could explain quantitative trait variation. When looking in literature it is found that the distribution of allelic effects follows an exponential curve, where most additive genetic variation is caused by a minimum amount of loci (Robertson 1967), which clearly is not in line with the popular infinitesimal model as developed by Fisher (1930).

The infinitesimal model suggests that multiple loci with minor effects explain quantitative traits, however other models predict that in most cases a few loci would explain the majority of the genotypic variation, while the majority of the loci will have minor genetic effects (Orr, 2003). The last model may be more applicable for potato tuber shape, which clearly is a polygenic trait, but potato tuber shape variation is caused by a few loci. Other examples that seem to fit this model are the *ZEP* gene (Wolters et al. 2010) and the *StCDF* gene (Kloosterman et al. 2013).

Conclusion

Multiple alleles of the *Ro* locus explain a large proportion of phenotypic variation of potato tuber shape. In this study the location of the *Ro* locus was refined to a region of 280 Kb on chromosome

10 between markers Asp6678 and Per20801. Within this region 14 positional candidate genes are located, from which six are members of the peroxidase gene family.

It is demonstrated that using sequence information for marker development is useful, albeit as described above, both advantages and disadvantages can be noted. In general sequence information is useful to investigate whether markers could potentially distinguish between four allele combinations as occurring in the C×E mapping population. Despite the usefulness of sequence information, the use of haplotype information would further improve the prediction of marker performance.

Within the target region between marker Asp6678 and marker Per20801, 14 candidate genes are located. Based on recombination, a distance of 0.6 cM was observed in the female parent and 0.125 cM in the male parent. This genetic distance allows further fine mapping of the *Ro* gene. However candidate gene approaches may also be well-suited to find the causative gene or sequence variations for the *Ro* locus.

When potato tuber shape is seen within the perspective of organ shape modulation, studying potato tuber shape may reveal fascinating biology into the context of shape modulation. Potato tuber shape can be regarded as a composite trait, with multiple characteristics mediating the final tuber shape, although the overall tuber shape is mediated by a single gene.

From all 14 candidate genes that reside in the target region, members of the peroxidase gene family are the most likely candidate genes for mediating potato tuber shape. As of now no reports are available which suggest the activity of peroxidases in mediating organ shape. Hence, studying the formation of potato tuber shape may reveal a novel function for the peroxidase gene family.

Chapter 4. References

Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D., & Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? The Plant Cell, 21(7), 1877–96.

Aulakh, S. S., Veilleux, R. E., Dickerman, A. W., Tang, G., & Flinn, B. S. (2014). Characterization and RNA-seq analysis of underperformer, an activation-tagged potato mutant. Plant Molecular Biology, 84(6), 635–58.

Bachem, C. W., Horvath, B., Trindade, L., Claassens, M., Davelaar, E., Jordi, W., & Visser, R. G. (2001). A potato tuber-expressed mRNA with homology to steroid dehydrogenases affects gibberellin levels and plant development. The Plant Journal : For Cell and Molecular Biology, 25(6), 595–604.

Baldwin, S., Revanna, R., Thomson, S., Pither-Joyce, M., Wright, K., Crowhurst, R., … McCallum, J. A. (2012). A toolkit for bulk PCR-based marker design from next-generation sequence data: application for development of a framework linkage map in bulb onion (Allium cepa L.). BMC Genomics, 13(1), 637.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. BMC Bioinformatics, 10, 421

Celton, J.-M., Christoffels, A., Sargent, D. J., Xu, X., & Rees, D. J. G. (2010). Genome-wide SNP identification by high-throughput sequencing and selective mapping allows sequence assembly positioning using a framework genetic linkage map. BMC Biology, 8(1), 155.

Chagné, D., Crowhurst, R. N., Troggio, M., Davey, M. W., Gilmore, B., Lawley, C., Peace, C. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PloS One, 7(2), e31745.

Chen, S. X., & Schopfer, P. (1999). Hydroxyl-radical production in physiological reactions. A novel function of peroxidase. European Journal of Biochemistry / FEBS, 260(3), 726–35.

Chitwood, D. H., Ranjan, A., Martinez, C. C., Headland, L. R., Thiem, T., Kumar, R., … Sinha, N. R. (2014). A modern ampelography: a genetic basis for leaf shape and venation patterning in grape. Plant Physiology, 164(1), 259–72.

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.

De Koeyer, D., Douglass, K., Murphy, A., Whitney, S., Nolan, L., Song, Y., & De Jong, W. (2009). Application of high-resolution DNA melting for genotyping and variant scanning of diploid and autotetraploid potato. Molecular Breeding, 25(1), 67–90.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics, 43(5), 491–8.

Ewing, E. E., & Struik, P. C. (1972). Tuber formation in Potato:Induction, Initiation and Growth. (J. Janick, Ed.). Oxford, UK: John Wiley & Sons, Inc.

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., … Zhang, Q. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and Applied Genetics. 112(6), 1164–71.

FAOSTAT (2013). FAOSTAT. Retrieved February 13, 2014, from http://faostat3.fao.org/faostat-gateway/go/to/home/E

Fisher, R. (1930). The genetical theory of natural selection. At The Clarendon Press.

Gebhardt, C. (2007). Molecular Markers , Maps and Population Genetics.

Hijmans, R. J., & Spooner, D. M. (2001). Geographic distribution of wild potato species. Am. J. Botany, 88(11), 2101–2112.

Jacobs, J. M., Van Eck, H. J., Arens, P., Verkerk-Bakker, B., Te Lintel Hekkert, B., Bastiaanssen, H. J., … Stiekema, W. J. (1995). A genetic map of potato (Solanum tuberosum) integrating molecular

markers, including transposons, and classical markers. Theoretical and Applied Genetics. 91(2), 289–300.

Jong, H., & Burns, V. J. (1993). Inheritance of tuber shape in cultivated diploid potatoes. American Potato Journal, 70(3), 267–284.

Kloosterman, B., Abelenda, J. A., Gomez, M. del M. C., Oortwijn, M., de Boer, J. M., Kowitwanich, K., … Bachem, C. W. B. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature.

Kloosterman, B., De Koeyer, D., Griffiths, R., Flinn, B., Steuernagel, B., Scholz, U., … Bachem, C. W. B. (2008). Genes driving potato tuber initiation and growth: Identification based on transcriptional changes using the POCI array. Functional and Integrative Genomics, 8(4), 329–340.

Kloosterman, B., Oortwijn, M., uitdeWilligen, J., America, T., de Vos, R., Visser, R. G. F., & Bachem, C. W. B. (2010). From QTL to candidate gene: genetical genomics of simple and complex traits in potato using a pooling strategy. BMC Genomics, 11(1), 158.

Kosambi, D.D., (1943). THE ESTIMATION OF MAP DISTANCES FROM RECOMBINATION VALUES. Annals of Eugenics, 12(1), 172–175.

Lamport, D. T. A., Kieliszewski, M. J., Chen, Y., & Cannon, M. C. (2011). Role of the extensin superfamily in primary cell wall architecture. Plant Physiology, 156(1), 11–9.

Leshem, B., & Clowes, F. A. L. (1972). Rates of Mitosis in Shoot Apices of Potatoes at the Beginning and End of Dormancy. Ann. Bot., 36(4), 687–691.

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

Lu, D., Wang, T., Persson, S., Mueller-Roeber, B., & Schippers, J. H. M. (2014). Transcriptional control of ROS homeostasis by KUODA1 regulates cell expansion during leaf development. Nature Communications, 5, 3767.

Masson, M. F. (1985). Mapping, combining abilities, heritabilities and heterosis with 4x X 2x crosses in potato. University of Wisconsin--Madison.

McFarlane, H. E., Döring, A., & Persson, S. (2014). The Cell Biology of Cellulose Synthesis. Annual Review of Plant Biology.

Montgomery, J. L., Sanford, L. N., & Wittwer, C. T. (2010). High-resolution DNA melting analysis in clinical research and diagnostics. Expert Review of Molecular Diagnostics, 10(2), 219–40.

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. Bioinformatics (Oxford, England), 24(16), 1757–64.

Morris, W. L., Ducreux, L. J. M., Hedden, P., Millam, S., & Taylor, M. A. (2006). Overexpression of a bacterial 1-deoxy-D-xylulose 5-phosphate synthase gene in potato tubers perturbs the isoprenoid metabolic network: implications for the control of the tuber life cycle. Journal of Experimental Botany, 57(12), 3007–18.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America, 76(10), 5269–73.

Oosterveld, A., Beldman, G., & Voragen, A. G. J. (2000). Oxidative cross-linking of pectic polysaccharides from sugar beet pulp. Carbohydrate Research, 328(2), 199–207.

Pereira, C. S., Ribeiro, J. M. L., Vatulescu, A. D., Findlay, K., MacDougall, A. J., & Jackson, P. A. P. (2011). Extensin network formation in Vitis vinifera callus cells is an essential and causal event in rapid and H(2)O(2)-induced reduction in primary cell wall hydration. BMC Plant Biology, 11, 106.

Powell, A. E., & Lenhard, M. (2012). Control of organ size in plants. Current Biology : CB, 22(9), R360–7.

Price, E. P., Smith, H., Huygens, F., & Giffard, P. M. (2007). High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of Campylobacter jejuni. Applied and Environmental Microbiology, 73(10), 3431–6.

Reed, G. H., & Wittwer, C. T. (2004). Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. Clinical Chemistry, 50(10), 1748–54.

Ritter, E., Gebhardt, C., & Salamini, F. (1990). Estimation of Recombination Frequencies and Construction of RFLP Linkage Maps in Plants From Crosses Between Heterozygous Parents. Genetics, 125(3), 645–654.

Robertson, A. (1967). The nature of quantitative genetic variation. In R. A. Brink) (Ed.), In Heritage from Mendel (pp. 265–280). Madison: University of Wisconsin Press.

Schopfer, P. (2002). Hydroxyl radical-induced cell-wall loosening in vitro and in vivo: implications for the control of elongation growth. The Plant Journal, 28(6), 679–688.

Sirks, M. J. (1929). Multiple Allelomorphs Versus Multiple Factors. Proceedings of the International Congress of Plant Sciences, 803–814.

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.

Tai, G. C. C., & Misener, G. C. (1994). A comparison of tuber shape and tissue composition of potato genotypes. Potato Research, 37(4), 353–364.

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. The Plant Cell, 16 Suppl, S181–9.

The European Cultivated Potato Database. (n.d.). The European Cultivated Potato Database. Retrieved from http://www.europotato.org/display_character.php?char_no=32&character=Tuber shape

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics.

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.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., & Boerjan, W. (2010). Lignin biosynthesis and structure. Plant Physiology, 153(3), 895–905.

Vogt, T. (2010). Phenylpropanoid biosynthesis. Molecular Plant, 3(1), 2–20

Voorrips, R. E. (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. Journal of Heredity, 93(1), 77–78.

Vos, P., Paulo MJ, van Eeuwijk, F. A., Visser, R. G., & Van Eck, H. (2014). Development and application of a 20K SNP array to characterise the tetraploid gene pool of potato (Solanum tuberosum).

Vossen, R. H. A. M., Aten, E., Roos, A., & den Dunnen, J. T. (2009). High-resolution melting analysis (HRMA): more than just sequence variant screening. Human Mutation, 30(6), 860–6.

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.

Wittwer, C. T. (2003). High-Resolution Genotyping by Amplicon Melting Analysis Using LCGreen. Clinical Chemistry, 49(6), 853–860.

Wolf, S., Hématy, K., & Höfte, H. (2012). Growth control and cell wall signaling in plants. Annual Review of Plant Biology, 63, 381–407.

Wolters, A.-M. A., Uitdewilligen, J. G. A. M. L., Kloosterman, B. A., Hutten, R. C. B., Visser, R. G. F., & van Eck, H. J. (2010). Identification of alleles of carotenoid pathway genes important for zeaxanthin accumulation in potato tubers. Plant Molecular Biology, 73(6), 659–71.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., … Visser, R. G. F. (2011). Genome sequence and analysis of the tuber crop potato. Nature, 475(7355), 189–95.

Xu, X., Vreugdenhil, D., & Lammeren, A. A. M. v. (1998). Cell division and cell enlargement during potato tuber formation. Journal of Experimental Botany, 49(320), 573–582.

Supplementary information I: Protocol DNA extraction with NaOH

Method according to Wang H *et al.* modified by Koen Pelgrom

Preparation:

- Bucket with ice
- Tweezers
- '8 in a row'- tubes and a box to store them.
- NaOH 0.5M (breaks down the plant material, DNA will be released)
- Tris 100mM (buffer, slows down the degradation of the plant material by NaOH)
- Pipets for 20 µL (repetitive), 5 µL (8-canal) and 100 µL (8-canal)

Greenhouse: collect leaf material

- Mark the tubes in order to prevent confounding of the tubes
- Keep the box with tubes on ice.
- Collect a small piece of leaf (0.5 cm^2) per sample, preferably the first leaf and not the cotyledon.

Lab

- Spin down at full speed (4500 rpm).
- Check whether all your samples are at the bottom of the tubes, if not, use your tweezers to do so and spin down again.
- Add 20 µl NaOH 0,5M per sample. There is no need to spin down the sample again: the NaOH will go down in the tissue striker. Don't waste much time between adding the NaOH and the next step (the NaOH is breaking down your sample).
- Place your box with tubes in the Tissue striker™.
- Check whether the striker is clean, otherwise rinse with water.
- In case you sampled less than 96 plants, make sure that the amount of tubes is equal for the left side and the right side of the box or fill the box with empty tubes. Keep balance.
- Strike for 5 minutes at full speed.
- Rinse the striker with water after use.
- Check whether all your samples are grinded, if not, grind manually by using a pipet point.
- Immediately add 20 µl Tris 100mM 7.5 pH per sample.
- Spin down
- Check whether your samples are at the bottom of the tubes.
- Prepare a plate by filling the tubes with 200 µl Tris.
- Add 5 µl liquid of the leaf sample to the prepared plate with 200 µL Tris. Don't pipet up the grinded leaf, but the liquid!
- Mix your leaf sample with the Tris by pipetting up and down with a volume of 100 µL.

Storage

- When you store the DNA in the -20^{0}C freezer, the DNA quality is good for at least 4 weeks.
- When you store the DNA in the refrigerator, the DNA quality will be degraded after a week (the NaOH will break down the DNA as well).

Use in PCR

- use 1 µL DNA isolate per PCR reaction

Reference

Wang H., Qi Meiqing and Cutler A. (1993): A simple method of preparing plant samples for PCR. – Nucleic Acids Research Vol. 21, No. 17: 4153-4154.

Supplementary information II: Procedure lightscanner

For optimal product amplification the procedure has been adapted from the lightscanner manual (REF). Reactions have been optimized for specific enzymes and fluorescent dyes.

Lightscanner PCR procedure Phire

Step	Temperature (°C)	Time (min)
1	98	0:30
2	98	0:10
3	Tm	0:10
4	72	0:30
5	go to step 2, **40x**	
6	94	0:30
7	25	0:30
8	10	∞

Preparation of lightscanner PCR

Step	What
1	Prepare mastermix (vortex!)
2	Pipette mastermix into each well
3	Add the DNA template into each well
4	Pipette oil overlay into each well
5	Seal
6	Centrifuge plate at 1500 rpm for one minute
7	Perform PCR
8	Remove seal
9	Run plate through lightscanner machine.

The melting temperature is changed according the DNA polymerase and fluorescent dyes. For reactions with LCGreen & Phire a general temperature of 60 °C is used, where reactions with EvaGreen & Phire a lower temperature of 58 °C because Evagreen has less influence than LCgreen on the melting temperature. The PCR procedure with the use of Dreamtaq is different, because Dreamtaq is not a hot-start DNA polymerase. The Tm is set around 57 °C when LCgreen is used, where with EvaGreen the Tm -5 is used.

Lightscanner PCR procedure Dreamtaq

Step	Temperature (°C)	Time (min)
1	95	3:00
2	94	0:30
3	Tm - 5	0:10
4	72	0:30
5	go to step 2, **40x**	
6	94	0:30
7	25	0:30
8	10	∞

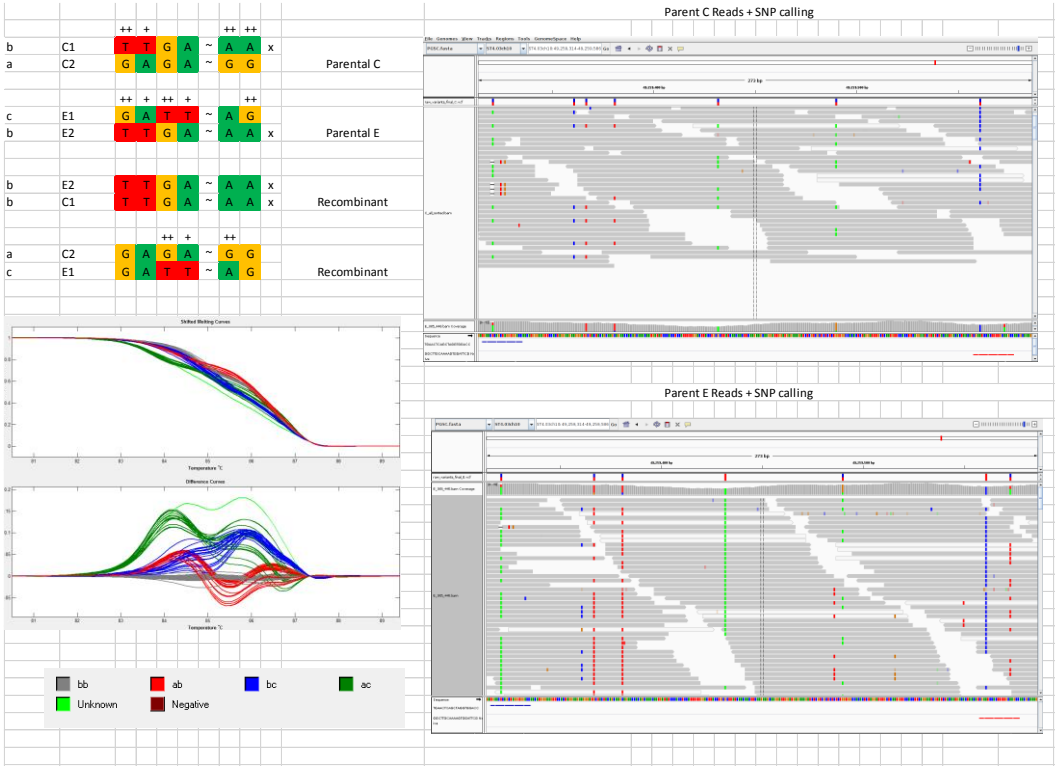
Because predicted melt temperature in primer3+ (bioinformatics wur) deviates from melt temperature as delivered by the company the melt temperature of primer3 is kept as a baseline.

Supplementary information III: Recombination frequencies and genetic distance

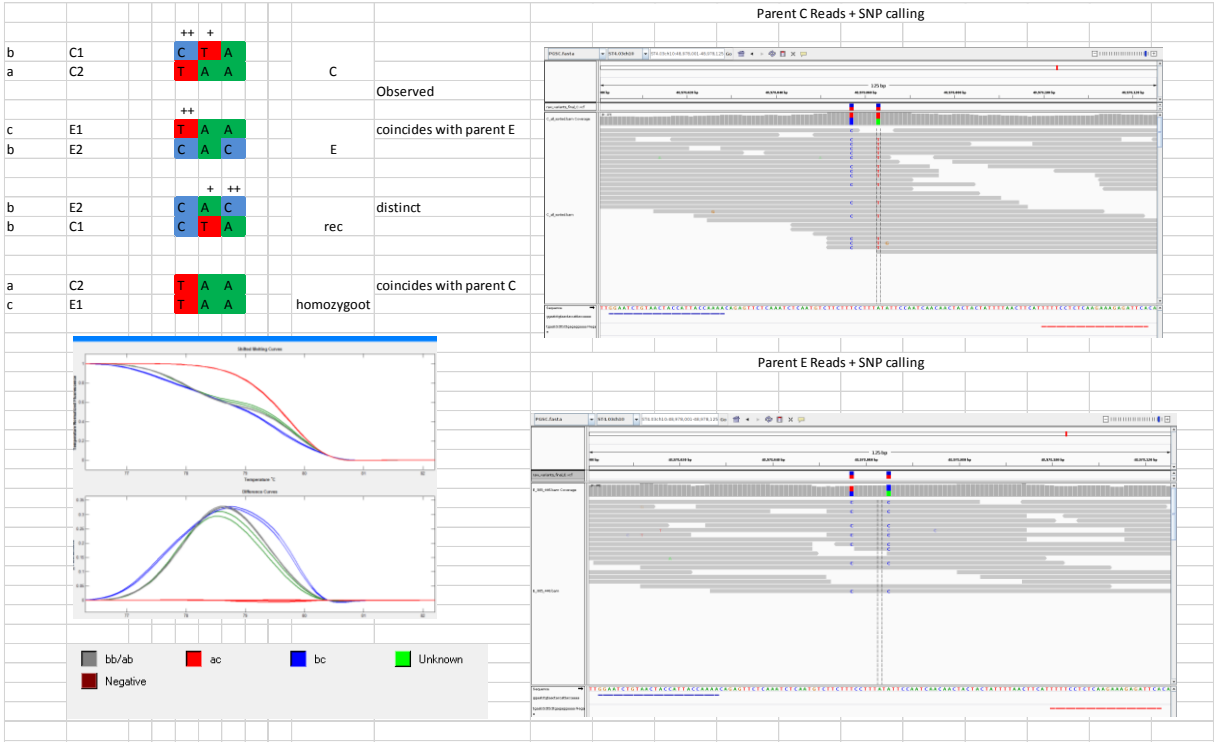
Witteveen (2013)				
Marker			Female parent C	
pair		# rec	r	d (cM)
LS_B446	PGSC0003DMG400027690	39	0.036	3.651
PGSC0003DMG400027690	KH27702	17	0.016	1.589
KH27702	PhoTr31222	20	0.019	1.870
PhoTr31222	Asp6678	8	0.007	0.748
Asp6678	Catper20798	0	0.000	0.000
Catper20798	Per20801	0	0.000	0.000
Per20801	Amt241073	12	0.011	1.122
Amt241073	495_499	9	0.008	0.841
				9.821
			Male parent E	
			r	d(cM)
LS_B446	PGSC0003DMG400027690	12	0.011	1.122
PGSC0003DMG400027690	KH27702	1	0.001	0.093
KH27702	PhoTr31222	0	0.000	0.000
PhoTr31222	**Solcap_1**	3	0.003	0.280
Solcap_1	**Nslp31336**	0	0.000	0.000
Nslp31336	**Nslp11953**	0	0.000	0.000
Nslp11953	Asp6678	1	0.001	0.093
Asp6678	Catper20798	0	0.000	0.000
Catper20798	Per20801	0	0.000	0.000
Per20801	**UPF11859**	3	0.003	0.280
UPF11859	Cons15896	1	0.001	0.093
Cons15896	Amt241073	0	0.000	0.000
Amt241073	495_499	6	0.006	0.561
				2.524
		798		
CE2013-feb				
Marker			Female parent C	
pair		# rec	r	d (cM)
PhoTr31222	Asp6678	11	0.014	1.379
Asp6678	Per20801	5	0.006	0.627
Per20801	Amt241073	6	0.008	0.752
				2.757
			Male parent E	
			r	d(cM)
PhoTr31222	Asp6678	8	0.010	1.003
Asp6678	Per20801	1	0.001	0.125
Per20801	Amt241073	2	0.003	0.251
				1.379

Supplementary information IV: Examples of haplotype information coupled with lightscanner outputs.

Per20801_F22: ST4.03ch10:49259316-49259577



JW_48978003:ST4.03ch10:48978003-48978123



Supplementary information V: List of markers and segregation data

In this table all positions for markers are shown. The IGV coordinates can be used with the potato reference genome version 4.03 to find locations of markers. The segregation of these markers is scored:

1) X means that no amplification was observed.
2) ♂ means segregation in the female parent.
3) ♂♀ means segregation of markers in both parents.
4) A dash means that no information is available.
5) no seq means that amplicon is homozygous in both parents.

If NT is added to the segregation column, this means that the marker is Not Tested (NT)

Name	Primer #	Primer name	Primer Fwd	Primer Rev	Segregation	chromosome	start_location	end_location	length	IGV coordinates
JW	4	JW_22059723b	ATGTGGGGTAGATCTCATGCTT	GGACCTATCACAAGGAACCAAG	x	ST4.03ch00	220089 20	220091 56	236	ST4.03ch00:220089 20-22009156
JW	3	JW_22059723a	AGGGTCAAAATTGCTTAGGTCA	CCACAGATGTTCTTGATTTTGC	x	ST4.03ch00	220095 32	220097 20	188	ST4.03ch00:220095 32-22009720
JW	18	JW_11951	TAAGGGTCTTTTGGGTGCAG	CACCCATCGATGTAATTTCTCA	♂	ST4.03ch00	220135 29	220142 28	699	ST4.03ch00:220135 29-22014228
JW	7	JW_11951	TCCCTGGAGTTTGTGGAGTC	AACACAAGCAACCACCAATG	♂	ST4.03ch00	220136 96	220141 15	419	ST4.03ch00:220136 96-22014115
		JW_11951	TCCCTGGAGTTTGTGGAGTC	AACACAAGCAACCACCAATG	-	ST4.03ch00	220359 80	220364 00	420	ST4.03ch00:220359 80-22036400
JW		F_11953	TCAGCCCCTCTACTGACTGC	AGAAACAATGGATCCACAAGC	♂	ST4.03ch00	220570 81	220575 42	461	ST4.03ch00:220570 81-22057542
JW	9	JW_22127238	GTTTGCAAATGCAGTGGAATTA	ATGCATTAAATTTCACCCCAAC	x	ST4.03ch00	220832 05	220822 34		ST4.03ch00:220832 05-22082234
JW		F_11955	CCCTTCAAGATTAGCCTCTCC	TTTTTCATCGGTCTCACTGG	♂	ST4.03ch00	220991 85	221000 15	830	ST4.03ch00:220991 85-22100015
JW		R_11955_intron	CCCTTCAAGATTAGCCTCTCC	AGGTAAGTGTTTAAGGCTTATTCG	♂	ST4.03ch00	220991 85	220996 47	462	ST4.03ch00:220991 85-22099647
JW		F_37286_exon	TGACCGAGCCCTAGAAACC	GGGTAAGCAACAAAAGTGAGC	-	ST4.03ch00	221263 80	221265 47	167	ST4.03ch00:221263 80-22126547
JW		R_37286_intron	TGACCGAGCCCTAGAAACC	GCAAGGGATCTGATTGATCG	-	ST4.03ch00	221263 80	221269 94	614	ST4.03ch00:221263 80-22126994
JW	8	JW_11950	GGCCACTATTCCTAGCACCA	CAACGCAAGTGAGCACAAAT	♂	ST4.03ch00	221294 14	221297 87	373	ST4.03ch00:221294 14-22129787
JW	15	JW_11950	CTTGCTCATCTTGTGCTGGA	CCAAGGCAATGGGAATAAAG	♂	ST4.03ch00	221296 80	221302 53	573	ST4.03ch00:221296 80-22130253
AW	1	Un13216_F1	ATCCCACATTCGAGAACCAA	TGGTCCAAGTTTGCGTGTT	-	ST4.03ch10 h10	469861 67	469861 86	19	ST4.03ch10:469861 67-46986186
AW	2	Kak11946_F2	GCAAAAATTTGTTCATTAGAGGTC	GTGGCGTGCTGCTGTATTTA	-	ST4.03ch10 h10	471076 40	471079 49	309	ST4.03ch10:471076 40-47107949
AW	3	NTGP3119 45_F3	ACCATCCTGGTGCAGTACCT	CCCTAATAGCCGCATCAAAA	-	ST4.03ch10 h10	471681 14	471681 33	19	ST4.03ch10:471681 14-47168133
AW	4	Axi1p16825_F4	CATTGGGTCTCCCATGAAAC	AAGACGTTGGCTCCTCTCAA	-	ST4.03ch10 h10	480615 86	480620 11	425	ST4.03ch10:480615 86-48062011
AW	5	Un16829_F5	GAATCTTGAAGGCGCACAAC	GCCAATGTTTCTGCCTTTGT	-	ST4.03ch10 h10	481556 04	481556 23	19	ST4.03ch10:481556 04-48155623
AW	6	Sca16833_F6	TCCCAAGCATGAATTCAACA	AGTGCAGCAACAGCAACAAC	-	ST4.03ch10 h10	482528 53	482532 73	420	ST4.03ch10:482528 53-48253273
AW	7	APK1B2768 9_F7	CAAGGTGATGAGACCCATGT	TCCTTGCTTGGTCTGGTCTT	-	ST4.03ch10 h10	484226 01	484230 53	452	ST4.03ch10:484226 01-48423053
AW	8	Un42461_F8	CTGATGTGATTCCATCTGGTG	TGTTCTCCATTCCAATTGATCC	-	ST4.03ch10 h10	485676 71	485680 70	399	ST4.03ch10:485676 71-48568070
AW	9	KH27702_F9	ATTGAGACGGGCACAAGAAA	CGTTTCAATCCACCGCTATT	-	ST4.03ch10 h10	486716 47	486721 77	530	ST4.03ch10:486716 47-48672177

A		PhoTr3122	CTAGTTGCTGCTGC	AGCCCCTTTGACCA		ST4.03c	487656	487660		ST4.03ch10:487656
W	10	2_F10	TTCTGC	AAACTT	♂♀	h10	29	94	465	29-48766094
			AGGGTTCAGTGAAT	CATGGTCGTTCCAC		ST4.03c	488450	488453		ST4.03ch10:488450
JW	10	JW_31233	GGGTTG	CTTTCT	♂	h10	16	68	352	16-48845368
			AAGCGTCCCATCAA			ST4.03c	488453	488453		ST4.03ch10:488453
JW	13	JW_31233	AATTCA		♂	h10	65	45	-20	65-48845345
A		PKA13235_	CATACCCCGAGAGC	GGCTTCTTTGCAAC		ST4.03c	488780	488785		ST4.03ch10:488780
W	11	F11	TTAGCA	ATTCTTC	-	h10	27	00	473	27-48878500
		F_4889198	gctcttcacgagttccaat	caatcaggacgcatgtgg		ST4.03c	488920	488921		ST4.03ch10:488920
JW		0	ga	ta	♀ NT	h10	33	98	165	33-48892198
			TGTTGCGGTGGTGT	TTGTTGAAGAAGCA		ST4.03c	489043	489049		ST4.03ch10:489043
JW	14	JW_31236	TAAGAG	ACAACCA	♂	h10	57	51	594	57-48904951
			CTCCGTTGATTGCT	TGGATCAGCTTTCA		ST4.03c	489045	489047		ST4.03ch10:489045
JW	2	JW_31236	CCAAGT	TCAACCT	♂	h10	32	81	249	32-48904781
		F_4891460	aaatttcaccaaaatgtg	tgtggtgatttcagctaac		ST4.03c	489146	489148		ST4.03ch10:489146
JW		0	agca	aaaa	♀ NT	h10	38	46	208	38-48914846
		JW_48978	ggaatctgtaactaccatt	tgaatctctttcttgagag		ST4.03c	489780	489781		ST4.03ch10:489780
JW		003	accaaaa	gaaaa	♂♀	h10	03	23	120	03-48978123
A		Asp6678_F	CCCCAACTATGAGC	GCACAGCCCAATAG		ST4.03c	489807	489809		ST4.03ch10:489807
W	19	19	AGTGGA	ATGTCC	♂♀	h10	45	95	250	45-48980995
A		Cons45482	CTGTCCCAGTTGCA	CGGCAAGCTGAATC		ST4.03c	490927	490929		ST4.03ch10:490927
W	20	_F20	CAAGTC	CTAACC	♂♀	h10	29	97	268	29-49092997
A		Per20801_	TGAACTCAGCTAGG	GGCTTGCAAAAGTG		ST4.03c	492593	492595		ST4.03ch10:492593
W	22	F22	TGGACC	GATTCG	♂♀	h10	16	77	261	16-49259577
A		Ala11868_	AACCAAACCAAACG	TCAGTCCCGGAATT		ST4.03c	493557	493561		ST4.03ch10:493557
W	16	F16	CTTACAG	GGTGAA	-	h10	43	65	422	43-49356165
			TGGTTCCCGACATA	TCCTCTTCCCCTCAC		ST4.03c	493571	493576		ST4.03ch10:493571
JW	5	JW_11868	GTGACA	ATCAC	♂	h10	05	52	547	05-49357652
			GTGATGTGAGGGG	CGTACGAGGCGTGT		ST4.03c	493576	493581		ST4.03ch10:493576
JW	16	JW_11868	AAGAGGA	AGTGAA	♂	h10	33	10	477	33-49358110
		JW_49360	TCGCTCTGGCGGTG	ACTTTAAGTCGCGT		ST4.03c	493605	493606		ST4.03ch10:493605
JW	?	573fwd1	CCGCTGTA	CTGGCC	-	h10	74	38	64	74-49360638
		JW_49360	TCGCTCTGGCGGTG	ACTTTAAGTCGCGT		ST4.03c	493605	493606		ST4.03ch10:493605
JW		573fwd1	CCGCTGTA	CTGGCC	♂ NT	h10	74	38	64	74-49360638
		F_11867_e	GTGCTCAGATGCAG	GACAGTGGACGAC		ST4.03c	493709	493714		ST4.03ch10:493709
JW		xon	TGTTGG	ACAAACC	♂	h10	75	05	430	75-49371405
A		Trans1186	TGCTCTTTTAGTCCG	TGAAGCCAGTTATC		ST4.03c	494414	494417		ST4.03ch10:494414
W	15	5_F15	TGCAG	ACCGGA	-	h10	44	93	349	44-49441793
			GGTGTGCTTTCATC	TTATCACCGGAGCG		ST4.03c	494415	494417		ST4.03ch10:494415
JW		F_11865_1	TCTTGC	AATAGC	♂	h10	32	84	252	32-49441784
			TACCTTGCCTTTCTG	AAGCGTAGCAGTCA		ST4.03c	494420	494424		ST4.03ch10:494420
JW		F_11865_2	CATGG	TCAACG	♂	h10	70	12	342	70-49442412
		JW_11864	TACCAGCTGTGCTG	TCTGCGACAGCAGG		ST4.03c	494699	494700		ST4.03ch10:494699
JW		_Fwd3	GATGTC	TGTAAC	♂	h10	40	91	151	40-49470091
		JW_11864	TATCCACATTGGCA	AGTCGGAAATTCCA		ST4.03c	494706	494708		ST4.03ch10:494706
JW		_Fwd2	TGTGCT	GCAGAC	♂	h10	31	70	239	31-49470870
		JW_11864	GGGGTTGTGTTTTT	TTGTCCAGCTGAAT		ST4.03c	494714	494718		ST4.03ch10:494714
JW		_Fwd1	CAGTGG	TTTCCA	♂	h10	43	11	368	43-49471811
		JW_11859	GCATGGACTTGCTT	AACCTGATCACCAA		ST4.03c	494736	494740		ST4.03ch10:494736
JW	19	_fwd1	TGGTTT	CGAAGC	♂	h10	18	81	463	18-49474081
			CCAATGTGATCAGT	TGTCATCCCTGGTG		ST4.03c	494785	494788		ST4.03ch10:494785
JW	6	JW_11863	GCCAAC	CAGATA	x	h10	23	80	357	23-49478880
		JW_15887	CCCCAAATCTCGAT	GATCCATTCCAGCC		ST4.03c	495525	495529		ST4.03ch10:495525
JW		_Fwd	TTTTCC	AAAGAA	♂	h10	51	15	364	51-49552915
			CCCCAAATCTCGAT	GATCCATTCCAGCC		ST4.03c	495525	495529		ST4.03ch10:495525
JW		F_15887	TTTTCC	AAAGAA	-	h10	51	15	364	51-49552915
			TTCCAGAGATTCGT	ATCTGGCAAATCCT		ST4.03c	495870	495879		ST4.03ch10:495870
JW	12	JW_15895	GCAGTG	TGGTTG	no seg	h10	07	76	969	07-49587976
			CGGTCGTTAATGAG	CCAAATGCAGCTCT		ST4.03c	495876	495881		ST4.03ch10:495876
JW	17	JW_15895	TGTTGC	TTTTCC	no seg	h10	58	74	516	58-49588174
		JW_15890	TAGGCAGAGGACTT	TGTCACAACAATTG		ST4.03c	495985	495987		ST4.03ch10:495985
JW		_Fwd1	GGGGTA	GGGAAA	♂	h10	49	44	195	49-49598744
			TAGGCAGAGGACTT	TGTCACAACAATTG		ST4.03c	495985	495987		ST4.03ch10:495985
JW		F_15890	GGGGTA	GGGAAA	-	h10	49	44	195	49-49598744
		JW_49624	TACTAAAGTAAATA	AGACCGATGAAAAT		ST4.03c	496244	496245		ST4.03ch10:496244
JW	?	488fwd1	CTAGAGAT	AAATA	-	h10	92	63	71	92-49624563
		JW_49624	TACTAAAGTAAATA	AGACCGATGAAAAT		ST4.03c	496244	496245		ST4.03ch10:496244
JW		488fwd1	CTAGAGAT	AAATA	♂♀ NT	h10	92	63	71	92-49624563
			GGATGGCACGTTG	TGGCAGGTCTTTCA		ST4.03c	496275	496281		ST4.03ch10:496275
JW	11	JW_41163	AAGAAGA	CTGGAT	♂	h10	45	04	559	45-49628104
A	13	Cons15896	AGGCCATCTCATAA	TCAGAGTTGAGGCC	-	ST4.03c	496306	496309	338	ST4.03ch10:496306

W		_F13	CGTCGT	ATTGAAA		h10	57	95		57-49630995
A		Amt24107	CACCGCTGCTTATT	TCCTCCATTGAAAC		ST4.03c	497232	497238		ST4.03ch10:497232
W	12	3_F12	GGGTAT	CTGACC	♂♀	h10	73	23	550	73-49723823
		JW_22067	AGGCTAACCACCCT	AGTTTGCAAAGCGA		ST4.03c	497712	497714		ST4.03ch10:497712
JW		_Fwd	TGACCT	TCTGGT	no seg	h10	63	70	207	63-49771470
		JW_Sotub1	GCATCGACCTGGAT	AGCTCCCTTGTTGC		ST4.03c		488755		ST4.03ch10:NA-
JW	1	8470	GTTCTTA	TCAGA	♂	h10	NA	86		48875586
A		Per25492_	AGTTACTAGTTGGA	GCTGATATTGTCGC		ST4.03c				
W	17	F17	GCTGGGA	GTTGGT	-	h10	NF	NF		ST4.03ch10:NF-NF
A		Glu25500_	GGGCTACTATGACA	GCACGAAAGCTTGT		ST4.03c				
W	18	F18	AGGCCT	ACCTTATG	-	h10	NF	NF		ST4.03ch10:NF-NF
A		KEAP11186	ATCGGCCTTGGACT	TGTCCTTCAGAGTA		ST4.03c			#WAA	
W	14	1_F14	TGTACT	GCTCCAG	-	h10	NS	NS	RDE!	ST4.03ch10:NS-NS
A		Catper207	GGCAGTGGTTCTAC	CTGCTAGAGGATGT		ST4.03c				
W	21	98_F21	TTGCTG	GAGGTGA	♂♀	h10	NS	NS		ST4.03ch10:NS-NS
		JW_11948	TCAGCCAGAGGATT	AGTTCCACTTCCCAT		ST4.03c				
JW	?	_fwd1	TGAGGT	GTTGG		h10	NS	NS		ST4.03ch10:NS-NS
		JW_11948	AAGGGTTGTTGAG	TTTGCTTGACGTTG		ST4.03c			#WAA	
JW	?	_fwd1	GATGCAG	ACTTGC	-	h10	NS	NS	RDE!	ST4.03ch10:NS-NS

Supplementary information VI: Comparison of models

Pot experiment					
	Deviance	DF	LRT	DF	P-value
Visual	-338.35	239			
Visual + clone_1	-415.26	238	76.91	1	**0**
Visual + clone_1 + clone_2	-416.15	237	0.89	1	**0.3455**
Visual + clone_1 + clone_2 + tubers	-418.07	236	1.92	1	**0.1659**
Visual + clone_1	-415.26	238			
Visual + clone_1 + tubers	-417.1	237	1.84	1	**0.175**
Visual	-338.35	239			
Visual + clone_2	-338.43	238	0.08	1	**0.7773**
Genotype	-344.29	239			
Genotype + clone_1	-411.54	238	67.25	1	0
Genotype + clone_1 + clone_2	-412.51	237	0.97	1	0.3247
Genotype + clone_1 + clone_2 + tubers	-414.43	236	1.92	1	0.1659
Witteveen (2013)					
Morpho	-657.92	321			
Morpho + clone_1	-795.19	320	137.27	1	0
Morpho + clone_1 + tuber	-795.74	319	0.55	1	0.4583
Genotype	-656.32	321			
Genotype + clone_1	-787.97	320	131.65	1	0
Genotype + clone_1 + tuber	-788.29	319	0.32	1	0.5716

Supplementary information VII: Dotplots

On the left a dotplot of scaffold DMB385 is shown, but on the right side a dotplot of scaffold DMB385, DMB546 and DMB446 is showed.