Engineering of tuber shape in potato (Solanum tuberosum) with marker assisted breeding or genetic modification using StOFP20

Herman J. van Eck (herman.vaneck@wur.nl)
Wageningen University & Research https://orcid.org/0000-0002-6530-0616
Maria E.P. Oortwijn
Wageningen University & Research
Inez R. Terpstra
Meijer Potato B.V.
Natascha H.M. van Lieshout
Wageningen University
Esther van der Knaap
University of Georgia
Johan H. Willemsen
Wageningen University & Research
Christian W.B. Bachem
wageningen university & Research

Research Article

Keywords: Ovate Family Protein, Tuber Shape, Potato, Solanum tuberosum, haplotypes, allele mining, natural diversity

Posted Date: July 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1807189/v1

Abstract

Potato *OFP20* is associated with the regulation of tuber shape. To further characterise the role of this gene in tuber shape, a panel of 136 potato varieties was re-sequenced to identify variants in the coding region of StOFP20. These SNPs were assembled into haplotypes and their allelic dosage was determined. Haplotype *StOFP20.1* is the most common allele (65%) which in quadruplex condition results in long tubers. *StOFP20.3* represents the second most common haplotype (22%) and is recognized as a dominant allele that is associated with round tubers in a dosage dependent manner. *StOFP20.4* represents the third common haplotype (5%) and encodes a non-functional gene. The remaining haplotypes represent rare alleles and their phenotypic effect is unclear. We developed reliable DNA markers that distinguish between the long and round alleles for marker-assisted selection of tuber shape in a diverse collection of varieties. We also demonstrate that we can genetically engineer the tuber shape of two commercial tetraploid varieties and the diploid clone DM. Knock down of *StOFP20* in the variety Atlantic using the RNAi strategy changed the tubers from round to oval and long-oval shapes. Conversely, overexpression of *StOFP20* in Spunta changed tuber shape from long to round. The most dramatic change from very long into round tubers was achieved by overexpression of *StOFP20* in DM. Our results demonstrate that engineering potato tuber shape is readily achieved by modulating *StOFP20* gene expression in a dose dependent manner, either by traditional breeding or by using genetic engineering methods.

Key Message

Knockout and complementation with *StOFP20* changes potato tuber shape, and for non GMO applications the breeders can select natural allelic variation.

Introduction

Potato is one of the major crops of increasing importance in the rapid developing economies in the world. Unique among staple crops, potato forms tubers that arise from underground adapted stems called stolons. In recent years, the pathway controlling the initiation of tuber formation has been elucidated (Abelenda et al. 2011). Initiation of tuberization involves the synthesis of a mobile FT-like signal peptide called a tuberigen that travels from the leaves to the stolon tip. Tuberigen forms a tuber-initiation complex that regulates the expression of downstream genes involved in tuber identity (Navarro et al. 2011). These downstream genes are thought to induce the reorientation of the cytoskeleton from predominantly transverse to longitudinal giving rise for the swelling of the stolon tip (Sanz et al. 1996).

After its initiation, organ growth occurs along three axes (Van der Knaap et al 2014). Differently shaped organs such as round versus elongated fruits and seeds are likely resulting from enhanced cell division along one of the axes of growth. A recently identified genetic mechanism that might explain differential cell division involves members of the OFP and the TRM families. In tomato certain members of these two families regulate the direction of cell division and thereby fruit shape (Wu et al, 2018; Lazzaro et al, 2018).

A subset of the TRMs are associated with microtubules, and TRMs are part of a protein complex that regulates the organization of cortical microtubules and the formation of the preprophase band. The control of cell division and cell growth are thought to generate the majority of the shape variation in plant organs (Snouffer et al, 2020).

In tomato, a knock-out mutation in *OVATE* and downregulation of *SlOFP20* together lead to an elongated and obovoid tomato fruit (Wu et al. 2018). Orthologs of tomato *OFP20* have effects on organ shape in a variety of plant species and organs including melon and Arabidopsis. OVATE and *SlOFP20* proteins interact with Tonneau1 Recruitment Motif (TRM) proteins in Yeast 2-Hybrid and in tobacco leaf epidermal cells. Interestingly, mapping studies have shown that the *Ro* locus involved in potato tuber shape maps to a region on chromosome *10* (Van Eck et al. 1994). Tuber shape was mapped with to PGSC coordinates chr10:48218826..51926832 by Endelman and Janski (2016) in an F2 population derived from a cross between *Solanum tuberosum* group Phureja DM1-3 516 R44 (hereafter referred to as DM) and *Solanum chacoense* M6. High resolution mapping delimited the physical interval of the tuber shape locus to a ~ 280kb region (chr10:48978066..49260070) within scaffold DMB546 (Willemsen et al. 2018b). No obvious candidate gene could be identified in the DM reference genome, as scaffold DMB546 is enriched for peroxidase genes. When the genome sequence of *S. chacoense* M6 was released (Leisner et al. 2018) it appeared that the potato ortholog of *SlOFP20* was deleted in the DM reference genome. Based on the genetic and physical co-localisation of *SlOFP20* with the *Ro* locus involved it was inferred that the potato orthologue of *SlOFP20* should be involved in tuber shape (Wu et al. 2018).

Genetic modulation of tuber shape is of great commercial value, because the market value of potato varieties high depends on tuber shape. Varieties suitable for French fries should be long, whereas varieties suitable for crisps should be round. When successful crisp varieties can be edited in varieties for French fries and vice versa, breeding success for one market niche can be exploited in the other market niche as well.

Here, we report that modulation of *StOFP20* gene expression affects tuber shape. The down regulation in a genotype with round tubers leads to elongated tubers whereas over expression in a genotype with long tubers leads to the wild type round tuber shape. We also describe allelic variation in tetraploid potato varieties, offering a prediction model to explain potato tuber shapes. These results shed further light on the central role of StOFP20 protein in determining tuber shape as well as opening the possibility to engineer tuber shape rapidly both by marker assisted breeding or by genome editing techniques.

Material And Methods

Tetraploid potato *S. tuberosum* cultivars Spunta (long), Atlantic (round) and *S. tuberosum* group *Phureja* DM1-3 516 R44 (long; hereafter referred to as DM) were used as targets for transformation. The round tuber tetraploid potato cultivar Lady Rosetta was used to isolate the *OFP20.2* allele. The *OFP20* coding region was PCR amplified from cDNA using Phusion High-Fidelity DNA Polymerase (Thermo Scientific™) with the forward primer 5’- caccATGGGGAATTATAGGTTTAGATTATCAGA annealing at the start codon

and the reverse primer 5'-TTACTTTAGTCGAATTTCAGTGATGTC annealing downstream from the stop codon. The amplification product was recombined into the pENTR™/D-TOPO® (Invitrogen) vector according to the manufacturer recommendations. For over expression in long tuber shape genotypes, the complete *OFP20.2* cDNA cloned by LR recombination into the binary vector pk7GW2 behind the CaMV 35S promoter using Gateway® cloning. For RNAi knock-down of the *OFP20* gene, a 500 bp section coding region was amplified with RNAi primers f-CACCAACAGCAGAGGCCATAGTCA and r-TTGTGACTGCAGAACTGATTACAT and cloned by LR recombination into the vector pK7GWIWG2(II) as described previously (Navarro et al. 2011). All constructs were sequenced to verify integrity, prior to transformation into potato.

Transformation of all potato genotypes was done using Agrobacterium-mediated transformation of in vitro grown internodes as described previously (Visser 1991). *In vitro* shoots were regenerated on selection with kanamycin (100 mg l − 1) Murashige and Skoog (MS) medium.

Untransformed controls and plants transformed with different constructs were propagated *in vitro*, on standard MS medium with 2% (w/v) sucrose for three weeks. After reaching the 5–10 leaf stage regenerants were transferred from *in vitro* medium to soil and after hardening off were transferred to 2 litre pots and grown in the greenhouse.

Tuber phenotyping

After 8 weeks (Spunta) or 10 weeks (DM and Atlantic), plants were harvested, and tuber shape was assessed. Tuber length (L) was measured (mm) and two width measurements were taken; maximal central width (MaxW) and minimal central width (MinW). A length/width ratio (R) was calculated as R = L / (MaxW + MinW)/2.

Statistical analysis was performed using in MS Excel. Specific details of the statistical test used, number of biological and technical replicates, and the description of error bars are provided in figure legends.

RNA Isolation

Total RNA from swelling stolons was obtained using RNeasy Mini Spin Columns including DNase treatment (QIAGEN). An aliquot of 1 µg total RNA was used to synthesize cDNA with iScript™ cDNA Synthesis Kit (Bio-Rad), following the manufacturer specifications. Quantitative RT-PCR was performed in reaction mixtures with IQTM SYBR® Green Supermix (Bio-Rad) and carried out with the BioRad CFX 96 system. Expression of Eukaryotic initiation factor 3E subunit (IF3E PGSC0003DMT400076704) was used in qRT-PCR assays to normalize target gene expression was used to normalize target gene expression (f-AACCATGACAATGGCAGGAC, r-TAGCCAAGTATCGCAGCAAG). The primers to monitor *StOFP20* gene expression were f-CAGAGAGCTTTGCAGTGGTG and r-GCAAATCTGGTTCGACATCA.

DNA sequence data

Three sources of sequencing data were available:

Firstly, two potato varieties, Altus and Colomba, have been sequenced to create *de novo* assemblies by NRGene (Ness Ziona, IL). These assemblies have lengths of 50–250 kb (Finkers, 2018, Hoopes et al. 2022) and were analysed with BLAST to find contigs with homology to the M6 version of *OFP20* (*ScOFP20*; Data S1). Four contigs were identified, each representing a unique haplotype.

Secondly, we performed PCR reactions with the same primers listed above on 19 tetraploid varieties (Data S2). PCR fragments were directly cloned into pGEM-T EASY (Clonetech) and individual colonies were sent for Sanger sequencing. In view of the high nucleotide diversity in potato the PCR primers may not amplify all alleles with equal efficiency.

Thirdly, we explored this region with WGS 150 bp PE sequences in a panel of 137 potato varieties (manuscript in preparation). The variety names of this panel are shown in Data S3 along with the phenotypic values for tuber shape of these varieties. Most tuber shape phenotypic values in this panel were retrieved from D'hoop et al. (2011) and complemented with data from breeders.

Bioinformatics

WGS reads of 137 tetraploid varieties were mapped to a reference composed of the open reading frame of the OFP20 gene from *S. chacoense* clone M6 (scaffold 814; Leisner et al. 2018) including 150 bp flanking sequences upstream of the promoter and downstream of the stop codon using the Bowtie2 algorithm (Langmead and Salzberg 2012) with default alignment parameters. Alignment data was processed with SAMtools and Picard (Durbin et al. 2009) to mark duplicate reads and estimate the average insert size of the paired-end reads. The Genome Analysis Toolkit (GATK) (McKenna et al. 2010) was used for indel realignment, base-score recalibration, and extraction of read depth information. Read depth and coverage data were processed with custom Perl scripts and BEDTools (Quilan and Hall, 2010).

Sequence variants were identified simultaneously among the aligned reads from all 83 tetraploids and the single monoploid using the FreeBayes polymorphism discovery algorithm (Garrison and Marth 2012). In order to include an alternate allele as a variant, supporting bases required a minimum base quality (BQ) of at least 13, and at least one supporting alignment was required to have BQ $\geq$ 20.

For genotype calling, zygosities of all alternative sequence variants were resolved by allele-specific read depths for all non-duplicated reads with MQ $\geq$ 13 using FreeBayes. This resulted in nulliplex (*aaaa*), simplex (*aaab*), duplex (*aabb*), triplex (*abbb*) and quadruplex (*bbbb*) genotype calls, recorded as 0, 1, 2, 3, 4 relative to the dosage of alternative variant, where M6 is the reference allele.

To verify the presence of DM-like haplotypes in the genepool of cultivated potato requires an experiment that aims to show the presence of an absence, because StOFP20 is part of a 42 kb deletion in DM. We used the M6 scaffold_814:197131–257245 as reference to map reads from our collection of tetraploid varieties. When both read pairs mapped to this 60 kb reference with an apparent insert size of 42kb between the read pairs, we conclude the presence of a DM-like haplotype in such a tetraploid variety.

To describe specific SNPs or haplotypes in StOFP20 haplotypes, the coordinates of the M6 genome were not used, but we rather used the nucleotide positions of ScOFP20 (1050 bp, SEQ. ID No.1) as coordinates where the A of the start codon is on coordinate 1 and the stop codon is at 1050.

Dot plot images were created with Genome Pair Rapid Dotter (GEPARD) software (Krumsiek et al. 2007) downloaded from http://cube.univie.ac.at/gepard.

Evolutionary history of PCR amplicon sequences was inferred by using the Maximum Likelihood method as implemented in MEGA7 (Kumar et al 2016). The tree with the highest log likelihood (-490.3710) was retained. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 88 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 279 positions in the final dataset.

Association Analysis

Phenotype means for tuber shape in each of the 137 tetraploid varieties, measured previously over a period of five years (D'hoop et al., 2011), were used for conducting association analysis between sequence variants and tuber shape. Additive and dominant genotype models were each tested both with and without correction for population structure using GWASpoly (Rosyara et al. 2016). The genotype clusters identified by PCA analysis were used as the adjustment factors for population structure. For the additive tetraploid genotype models, we applied linear regression models implemented in Genstat. For each trait and each marker, the model applied was as follows: response = allele copy number (+ structure) + error.

Haplotype inference

Unphased bi-allelic sequence variants were used to compute longer haplotypes. The procedure "Happy-haplotype-interference" is available, implemented in a set of Python scripts, whereas both pairwise phasing (EM) as subsequent joining procedure are used to reconstruct haplotypes. As input a matrix with marker scores on rows (0-k ploidy), and columns corresponding to varieties. The scripts can be found on the Gitlab repository located at:

https://git.wageningenur.nl/wille094/Happy-haplotype-inference/tags/0.8.2 . Detailed description of the method can be found in Willemsen et al. (2018).

Results

The tuber shape gene is absent in the DM reference genome

In order to determine the origin of tuber shape diversity, we studied the collinearity of the genome sequences of two diploid potato clones DM and M6 at the tuber shape locus. Specifically we aligned scaffold_814 of M6 (https://datadryad.org//resource/doi:10.5061/dryad.kc835) with scaffold PGSC0003DMB000000546 of DM which is located at chr10:48972288..49295447 (http://spuddb.uga.edu/) of the potato reference genome (v4.03; PGSC, 2011). We observed collinearity between the second part of scaffold_814 which aligns with the first part of DMB546 (Fig. 1). At a 40 kb distance from the start of DMB546 a larger interruption of the diagonal in the dotplot indicates a 42 kb gap in the DM sequence relative to the M6 sequence. Thereafter, the remainder of the DMB546 sequence shows a complex homology pattern with multiple diagonals representing the large number of tandemly duplicated sequences annotated as peroxidase genes. Hereafter we refer to this DM-haplotype comprising the 42 kb deletion as the DM null-allele.

Assessment of the allelic composition at the *StOFP20* locus of tetraploid varieties would be severely limited in the case that haplotypes similar to the DM null allele (missing the *StOFP20* gene) are present. Therefore, we aimed to verify the presence of DM-like haplotypes in the gene pool of cultivated potato. We initially used 15 genotypes with a long phenotype or a genotype tentatively identified as quadruplex for the “long” allele (i.e. Annabelle, Caesar, Clearwater Russet, Cmk1997007004, Cmk2008622009, Cmk2008071009, Crisps4all, Dinky, Jurata, Lady Terra, Leonardo, Mondial, Memphis, Monalisa and Spunta), as these varieties may have an increased probability to contain the DM-like null alleles. In four of these varieties (Crisps4all, Dinky, Leonardo and Spunta) read pairs were identified with an apparent insert size of 42 kb (mapping quality > 20; >10% of the read coverage). Thus, we conclude that the DM-like haplotype is also present in the commercial potato gene pool, estimated at an allele frequency of approximately 5%. It appeared that this analysis strongly depends on thresholds for read depth and mapping quality. Therefore, in the remainder of this manuscript we ignore the null-alleles due to their low estimated frequency and, since such null alleles only marginally interfere with the characterisation of functional (“round”) alleles.

Analysis of DNA sequence variants in the coding sequence of *StOFP20*

In scaffold 814 of M6 a gene encoding OFP20 was identified and named *ScOFP20* where the first two letters refer to *Solanum chacoense* (Data S1). *ScOFP20* encodes a gene of 1050 bp without introns. Using the sequence of *ScOFP20* four additional potato haplotypes of OFP20 could be identified in *de novo* assembled scaffolds of the tetraploid varieties Altus and Colomba (Hoopes et al. 2022). These four haplotypes have been assigned the names *StOFP20.1* to *StOFP20.4* (Data S2), where the first two letters refer to *Solanum tuberosum*. *StOFP20.1*, *StOFP20.2* and *StOFP20.3* show high sequence homology (96–99%) and their coding regions are 1041 bp, 1074 bp and 1053 bp long, respectively. The main difference between these three haplotypes are three in frame indels of 3 bp, 9 bp and 21 bp (translating to P, HDV and KLVNYSP indels) as well as a (AGC)5 or (AGC)6 microsatellite. Haplotype *StOFP20.4* has a 241 bp insertion, where a premature stop renders a truncated protein. Details are shown in Data S3 where protein sequence alignments of the alleles can be found.

PCR amplification products of *StOFP20* from 19 tetraploid varieties were cloned and individually sequenced colonies resulted in 82 *StOFP20* sequences, representing 11 unique haplotypes. The sequences were clustered into a Neighbor Joining tree using MEGA7 (Kumar et al. 2016) and two haplotypes *StOFP20.1* and *StOFP20.3* could be identified in the majority of varieties, along with minor clusters of sequences representing other rarer haplotypes, as well as haplotypes observed only once (Fig. 2). These singleton haplotypes require confirmation to exclude possible sequencing errors. The distribution of the haplotypes in 137 tetraploid varieties is shown in Data S4 and Data S5.

Comparison between the sequences of the PCR amplicons and *ScOFP20* of the M6 genome, identified five private SNPs uniquely observed in *S. chacoense* but not in *S. tuberosum*. These were SNPs on coordinates 342 [C/T], 392 [C/T], 516 [C/T], 546 [G/A] and 771 [T/A]. Within *S. tuberosum* germplasm, 31 sequence variants were observed in the ORF of *StOFP20* gene, which is an average of 1 SNP per ~ 35 bp. This includes the P, HDV and KLVNYSP indels described above and 28 additional SNPs.

PCR amplicon sequencing may be affected by unequal amplification or primer annealing efficiency and may therefore not reveal all haplotypes. As a complementation we explored WGS read pairs from a panel of 136 potato varieties. This allowed us to confirm the polymorphic sites across PCR identified haplotypes and also to identify new polymorphisms or haplotypes.

SNP calling across the 136 potato varieties added another eight, albeit very rare variants, with an allele frequency below 1%. The SNPs at coordinates 132 [A/T], 184 [C/T], 595 [C/T] were observed in only four potato clones: the progenitors AM 78-3736 and AM 78-3787 and their offspring clones; Mercury and Vectra (haplo_9). Other SNPs were observed only once or twice in genotypes AM 78-3704 (haplo_30) or Lady Amelia and Lady Jo (haplo_64, haplo_75). In view of the low frequency of these SNPs and their haplotypes we need to be cautious, but “AM 78” clones are known to carry introgressed alleles from wild progenitors. Data S5 gives an overview of the SNPs, their population allele frequency, and their haplotype specificity.

Analysis of DNA sequence variants in regions flanking the coding sequence of *StOFP20*

Apart from sequence variation affecting protein function, other upstream sequence variants may affect the promoter region, the binding of transcription factors and thus affect gene expression. As clone M6 has round tubers (Jansky et al, 2014) and most likely homozygous (Marand et al. 2019), we can assume that *ScOFP20* not only encodes a functional protein, but also has a functional promoter. A peculiar feature is a C-rich region at -110 to -163 nt upstream the start codon with several mononucleotide motifs (poly-C tracks) up to $(C)_{11}$ with one or two dispersed A residues (Fig. 3). In contrast, this feature is much more pronounced in DNA sequences derived from *S. tuberosum*, where the C-rich region is at least 100 bp long and the mononucleotide motif $(C)_n$ assumes lengths of exceeding 50 bp C’s. A more precise description of this region is not feasible, because the *de novo* assemblies by NRGene are intervened by long stretches of N, suggesting that the C-rich tract may be even longer. Haplotype specific analysis of RNA transcript levels are required to study the impact of these mononucleotide motifs on gene

expression. If these features hamper gene expression, then CrisprCas9 excision of these tracts may restore normal transcription levels, and may result in a more round tuber shape.

Statistical phasing of sequence variants into haplotypes

The sequence variants were joined into haplotypes using the “Happy-haplotype-inference” approach (Willemsen et al., 2018). In total 92 haplotypes were inferred from 137 potato varieties (note that 137 tetraploid varieties represent 548 alleles). However, 70 haplotypes were observed only once and 11 haplotypes twice (Data S5). These could be discarded as artefacts, potentially due to erroneous SNP allele dosage estimation, spurious chimeric haplotypes, or rare alleles representing remnants of introgression breeding with wild species. For the remaining 11 haplotypes four were in perfect agreement with the haplotypes already obtained from the *de novo* assemblies by NRGene and PCR amplicon sequencing (haplotypes 71, 54, 20 and 55). The remaining 7 haplotypes are rare in the gene pool (haplotypes 79, 21, 22, 9, 31, 65 and 89). They were observed in our panel between 3 and 6 times (0.55% <MAF < 1.1%) but could not be matched with our PCR haplotypes for two reasons: Firstly, not all varieties used for PCR belong to the variety panel. Secondly, some haplotype specific variants observed by PCR were not detected in WGS data, probably because these rare alleles were below the threshold for variant calling.

Identification of *StOFP20* allele compositions in potato varieties

The four WGS inferred haplotypes, with complete sequence identity with PCR derived sequences, were also compared for their relative abundance in both datasets. In both cases highly similar allele frequencies were obtained. We conclude that these four alleles are well confirmed as they could be retrieved from all three sources (de novo assemblies, WGS and PCR).

It appeared that, across the panel of 137 potato varieties, *StOFP20.1* was the most frequent haplotype (scaffold 2173, haplo_71) with an allele frequency of 65% and represents an allele associated with long tubers. The second most frequent haplotype was *StOFP20.3* (scaffold 3525, haplo_54) with a frequency of 22% and is mostly found in varieties with round tubers. The third most frequent haplotype was *StOFP20.4* (scaffold 10792, haplo_20) with a frequency of 5% and also represents a long allele. *StOFP20.4* is a typical allele as it includes a 241 bp insert at coordinate 356 of the M6 *ScOFP20* reference allele. The insert causes a frame-shift mutation and the truncated protein is likely not to be functional. Sequence comparison of the insert using BLAST, identifies AB496980.1 with 99% sequence homology to the dTstu1 MITE transposon (Momose et al. 2010). The least frequent haplotype was *StOFP20.2* (scaffold 216, haplo_55) with a frequency of 1.5% and is mostly found in varieties with round tubers. These four alleles comprise 94% of the total alleles.

Phenotypic effect of sequence variants or haplotypes

In the panel of 137 varieties, we associated the sequence variants with trait values for tuber shape. Three SNPs were most significantly associated with a $-\log_{10}(p)$ value of approx. 7.0, as shown by the

Manhattan plot (Fig. 4). These were SNP_57_G/A [ref/alt], SNP_156_A/G and SNP_359_G/A which contrast the “round” *StOFP20.2* and *StOFP20.3* haplotypes having the M6 reference allele from the “long” *StOFP20.1* and *StOFP20.4* haplotypes which have the alternative SNP allele. The effect of these SNPs on the trait value is 1.02, which is not spectacular on a scale ranging from 2 to 9. However, when analysing the effect of zero, one or two round alleles, a clear dosage effect was observed: LLLL, RLLL, RRLL have tuber shape values of 3.9, 5.7 and 7.1 respectively (N = 18, 76 and 9 varieties).

The effects of the indels P, HDV and KLVNYSP was also evaluated. The P indel at coordinate 92 ($-\log_{10}(p)$ value = 5.4) is largely in LD with the three SNPs mentioned above and also separates the two round alleles from the two long, albeit with a few differences in data points. The presence of the HDV motif at the indel at coordinate ($-\log_{10}(p)$ value = 3.1) is only indicative for one round allele *StOFP20.3*, and therefore lacks power as it does not include the other round allele. The effect of the LVNYSP indel, which is only present in *StOFP20.2* insert is not reported in the Manhattan plot, because this sequence variant was not called from the WGS data. Probably the reads differed too much from *ScOFP20* reference due to the 21bp insert, and with MAF = 1.5% the allele is quite rare. Because the *StOFP20.2* allele was used for complementation, we know that the presence of motif LVNYSP is associated with a functional protein. Most of the other *ScOFP20* haplotypes lack the LVNYSP motif, but *StOFP20.1* and *StOFP20.2* alleles both contribute to round tuber shape. Hence we can conclude that the absence or presence of this LVNYSP motif does not influence the functionality of the protein.

Transformation events using StOFP20

A transgenic approach was used to test whether silencing or over expression of StOFP20 would allow changes in tuber shape in potato including in commercial tetraploid varieties. From the genetic analysis described above, the “round” *StOFP20.2* gene would be suitable to complement long varieties without round alleles and change their long phenotype into a transformant with round tubers. The results of our complementation experiments using the variety Spunta and clone DM is shown in Figs. 5 and 6. The insertion of a dominant round allele *StOFP20.2* indeed changed genotypes with a long tuber shape into plants production of round or oval tubers. Conversely, the variety Atlantic with round tubers could be modified into a plant with oval to long-oval tubers by knock down the expression of *StOFP20* using an RNAi approach.

Untransformed Spunta, Atlantic and DM plants were grown in five biological replicates.

Seven Spunta transformation events and ten from Atlantic and DM were planted in three biological replicates, grown from *in vitro* cuttings. The L/W ratio was determined for individual tubers and averaged across tubers per plant. The L/W ratio and the error-bars shown in the histogram represent the L/W ratio and st.dev. across replicated plants per transformation event.

While transformants with highly significant changes of the length/width (L/W) ratio were generated, considerable variation was observed across the different transgenic events. All seven Spunta transformants analysed, showed a round L/W ratio, where Spunta_35S_2 produced nearly perfect round

tubers. However, Spunta_35S_14 produced mostly oval to long-oval tubers. The most dramatic change was found for DM (L/W ratio = 3), where the extreme long tuber phenotype was modified to almost round tubers (L/W ratio = 1.27 for DM_35S_3). Four DM transformants (DM_35S_1, DM_35S_3, DM_35S_6 and DM_35S_7) did not show a significant effect.

In order to assess the effects of over expression and RNAi knock-down of the *StOFP20* in the three transgenic potato backgrounds, RNA was isolated from swelling stolon tips and *StOFP20* transcript was detected using qRT/PCR. As expected, the 35S over expression showed high levels of the introduced gene. RNAi knock-down of expression of the *StOFP20* gene showed a significant reduction in expression of the resident gene (Fig. 7C). In both cases of gene expression modulation, the resulting gene expression was largely in accordance with the severity of the change in shape phenotype of the transgenic tubers. Our data underlines the direct link between gene expression levels and the function of the *StOPF20* gene and its effects on tuber shape.

Discussion

The functional name assigned to the various haplotypes, being a “round allele” or a “long allele” was based on its association with trait values or in the case of *StOFP20.4* based on the transposon insert. Except for *StOFP20.4* and the 42kb deletion of DM we have not studied the sequence variants causal to maintain or abolish its molecular function. Perhaps individual amino acids, but more likely the three or seven amino acid indels HDV or KLVNYSP at position 98 and 124 in the protein of clone M6 could affect molecular function. Besides of sequence variants in functional *StOFP20* proteins, expression variation caused by extensive length variation in mononucleotide motifs $(C)_n$ in the promoter region may play a significant role in the phenotype.

The most dramatic effect on tuber shape was observed in clone DM where the extreme long shaped tubers were modified to round ones. This suggests that the long tuber shape of DM is due to the 42 kb deletion completely eliminating the *StOFP20* locus. The complementation study also shows that expression of *StOFP20* alone, is sufficient to explain the variation in tuber shapes, ranging from round to oval and long. This is in agreement with literature where tuber shape was described as a monogenic trait with Mendelian inheritance of round being dominant over long (Van Eck et al, 1994).

Spunta is among the longest potato varieties from our panel, but hardly exceeds an L/W ratio larger than 2.0. Spunta still retains alleles that produce OFP20 protein, albeit an isoform that differs from the OFP20 proteins from “round” alleles. This may suggest that the “longest” alleles from the commercial gene pool have a residual effect on the phenotype that prevents the extreme long shape found in DM. Future breeding work may reveal the importance of the 42 kb deletion allele for the development of varieties with tubers with a L/W ratio exceeding 2.0.

The results presented in this paper allows breeders to modulate an economically important quality trait of potato. Existing varieties for processing industry can be edited to improve the length of French fries or to

create a round variety suitable for crisps out of a long tuber variety. It may simplify potato breeding when new material is developed by the breeding dept. from yellow fleshed, late, round parents, whereas the editing department can diversify these into combinations of white fleshed, early and long clones by knock downs of BCH, CDF and OFP20. However, marker assisted breeding would be more straightforward and we show that three SNP markers SNP_57_G/A, SNP_156_A/G and SNP_359_G/A are sufficient to contrast the “round” StOFP20.2 and StOFP20.3 haplotypes from the “long” StOFP20.1 and StOFP20.4 haplotypes.

Declarations

Author contributions

HvE, JHW, CB designed research; MO performed lab and transformation experiments; IRT, NHMvL and JHW performed bioinformatic analysis; HvE wrote the paper with input from EvdK and CB.

Acknowledgements

We thank Dirk-Jan Huigen and our Unifarm colleagues for taking care of plant experiments in the greenhouse.

Compliance with Ethical Standards

Not applicable

Funding Statement

JHW was financially supported by NWO project 831.13.002 of the Graduate School T&U program and HvE is grateful to the EU project H2020 INVITE under grand agreement Number 817970.

Conflict of Interest Disclosure

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this publication. Authors HJvE and EvdK have no ownership nor rights related to the patent application “WO2020008078A1” submitted by HZPC, in spite of being listed as inventor on these patent applications.

Data availability

The datasets used and/or analysed during the current study are available in the published manuscripts, or available as supplementary files.

References

1. Abelenda JA, Navarro C, Prat S (2011) From the model to the crop: genes controlling tuber formation in potato. Curr Opin Biotechnol 22(2):287–292. doi:10.1016/j.copbio.2010.11.013
2. D'hoop BB, Paulo MJ, Visser RG, van Eck HJ, van Eeuwijk FA (2011) Phenotypic analyses of multi-environment data for two diverse tetraploid potato collections: comparing an academic panel with an industrial panel. Potato Res 54(2):157
3. Durbin R, Li H, Handsaker B, Wysoker A, Fennell T et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079
4. Endelman JB, Jansky SH (2016) Genetic mapping with an inbred line-derived F2 population in potato. Theor Appl Genet 129(5):935–943
5. Finkers R (2018) De Novo Haplotype Assembly of Two Potato Genomes Using Denovomagic 3.0. Conference Abstract, PAG, San Diego, USA https://pag.confex.com/pag/xxvi/meetingapp.cgi/Paper/32010
6. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing.arXiv preprint arXiv:1207.3907
7. Hoopes G, Meng X, Hamilton JP, Achakkagari SR, de Alves Freitas Guesdes F, Bolger ME, Finkers R (2022) Phased, chromosome-scale genome assemblies of tetraploid potato reveals a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. Mol Plant 15(3):520–536
8. Jansky SH, Chung YS, Kittipadukal P (2014) M6: A diploid potato inbred line for use in breeding and genetics research. J Plant Registrations 8(2):195–199
9. Krumsiek J, Arnold R, Rattei T (2007) Gepard: A rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23(8):1026–1028
10. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874
11. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat methods Apr 9(4):357
12. Leisner CP, Hamilton JP, Crisovan E, Manrique-Carpintero NC, Marand AP, Newton L, Pham GM, Jiang J, Douches DS, Jansky SH, Buell CR (2018) Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species Solanum chacoense, reveals residual heterozygosity. Plant J 94(3):562–570
13. Marand AP, Jansky SH, Gage JL, Hamernik AJ, de Leon N, Jiang J (2019) Residual Heterozygosity and Epistatic Interactions Underlie the Complex Genetic Architecture of Yield in Diploid Potato. Genetics 212(1):317–332
14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K et al (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

15. Momose M, Abe Y, Ozeki Y (2010) Miniature inverted-repeat transposable elements of Stowaway are active in potato. Genetics 186(1):59–66
16. Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, Shimamoto K, Prat S (2011) Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature 478(7367):119–122
17. PGSC Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355):189
18. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842
19. Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato.The plant genome, *9*(2)
20. Sanz MJ, Mingo-Castel A, van Lammeren AA, Vreugdenhil D (1996) Changes in the microtubular cytoskeleton precede in vitro tuber formation in potato. Protoplasma 191(1):46–54
21. Snouffer A, Kraus C, van der Knaap E (2020) The shape of things to come: ovate family proteins regulate plant organ shape. Curr Opin Plant Biol 53:98–105
22. Van Eck HJ, Jacobs JM, Stam P, Ton J, Stiekema WJ, 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
23. Visser RGF (1991) Regeneration and transformation of potato by Agrobacterium tumefaciens. Plant tissue culture manual. Springer, Dordrecht, pp 301–309
24. Willemsen JH, Vos PG, Witteveen A, Visser RGF, van Eck HJ (2018a) Chap. 2 – The *Ro* locus involved in potato tuber shape is located in a ~ 280kb region enriched for peroxidases. In: The Identification of Allelic Variation in Potato. PhD thesis, Willemsen H.J., Wageningen University. ISBN 9789463435130, https://library.wur.nl/WebQuery/wda/2247530
25. Willemsen J.H., Visser, R.G.F., van Eck, H.J. (2018b) Chapter 4 - Haplotype inference in polyploid species and application to genetic analysis in potato. In: The Identification of Allelic Variation in Potato. PhD thesis, Willemsen H.J., Wageningen University. ISBN 9789463435130, https://library.wur.nl/WebQuery/wda/2247530

Figures

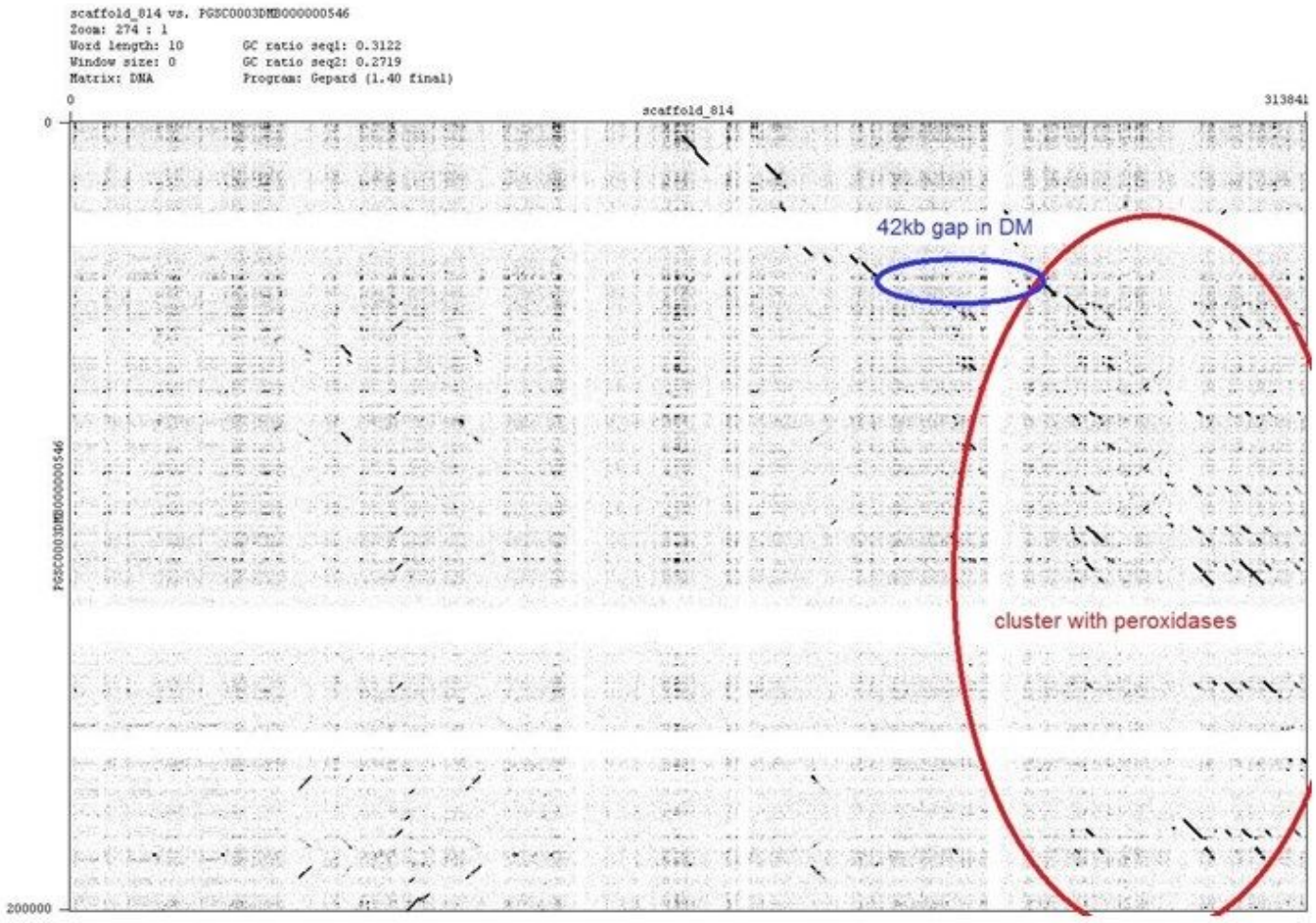

Figure 1

Dot plot analysis of sequence collinearity between *S. chacoense* clone M6 (scfld814) and the potato reference genome DM (v4.03 scfld DMB546) identifies a 42 kb deletion in DM (blue oval), which explains the absence of *StOFP20* from the reference genome DM. The red oval indicates the location of tandemly repeated sequences annotated as peroxidases.

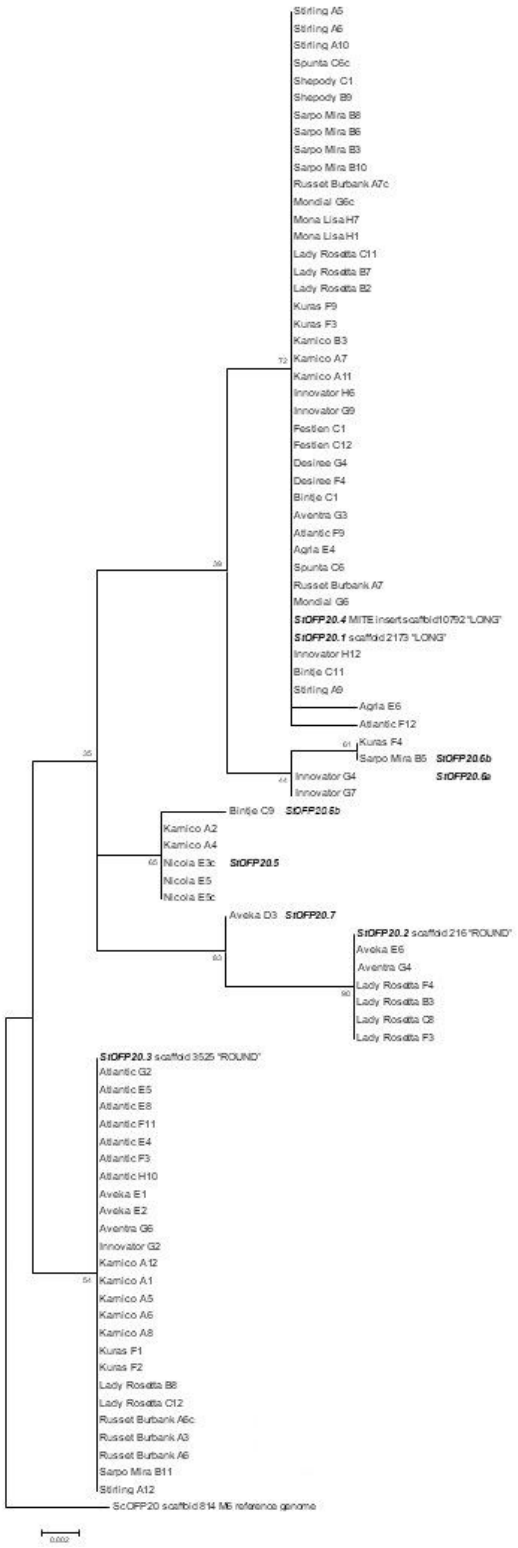

Figure 2

Maximum Likelihood tree of *StOFP20* PCR amplicons along with four *de novo* assembled sequences. The percentage of trees, in which the associated taxa clustered together, is shown next to the branches. From 19 tetraploid varieties 11 unique haplotypes were identified. Within the main clusters the representative sequences are named, ranging from *StOFP20.1* to *StOFP20.7* (Data S2). Sequences represented only once remain unnamed. Haplotypes *StOFP20.1* and *StOFP20.4* represent “long” alleles, whereas

StOFP20.3 represents a “round” alleles. The functionality of other haplotypes with low allele frequencies remains unclear.

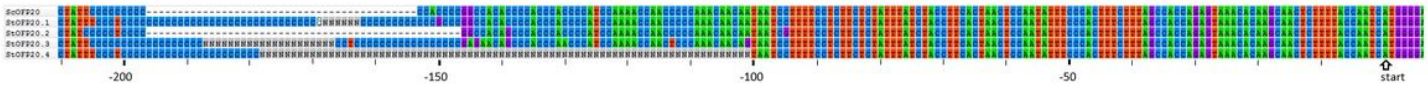

Figure 3

The promoter region of *StOFP20* (coordinates: -210; +3) contains the mononucleotide motif $(C)_n$ which can assume lengths exceeding 50 bp of C’s. The last three letters of this sequence alignment (ATG) is the start codon.

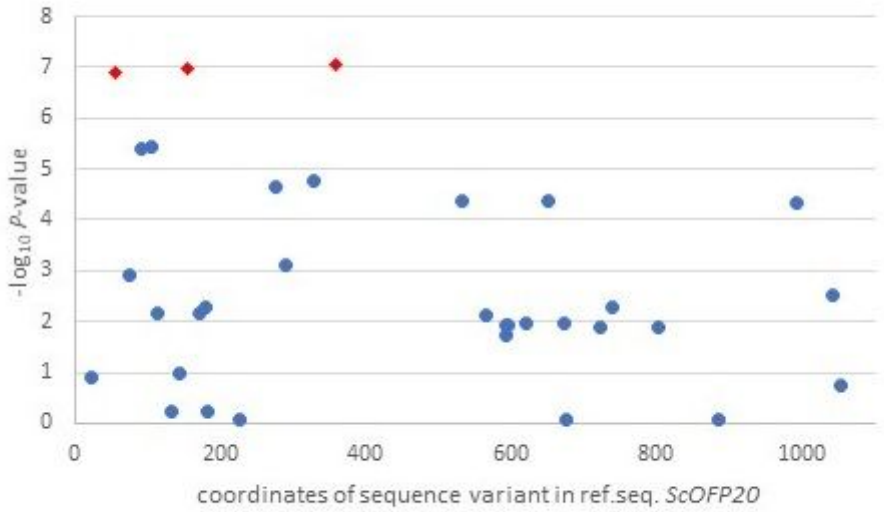

Figure 4

Manhattan plot of the significance of the association between StOFP20 sequence variants and tuber shape. Three variants (red diamonds) SNP_57_G/A, SNP_156_A/G and SNP_359_G/A are sufficient to contrast the “round” *StOFP20.2* and *StOFP20.3* haplotypes from the “long” *StOFP20.1* and *StOFP20.4* haplotypes. Other SNPs are less haplotype specific or indicate rare haplotypes

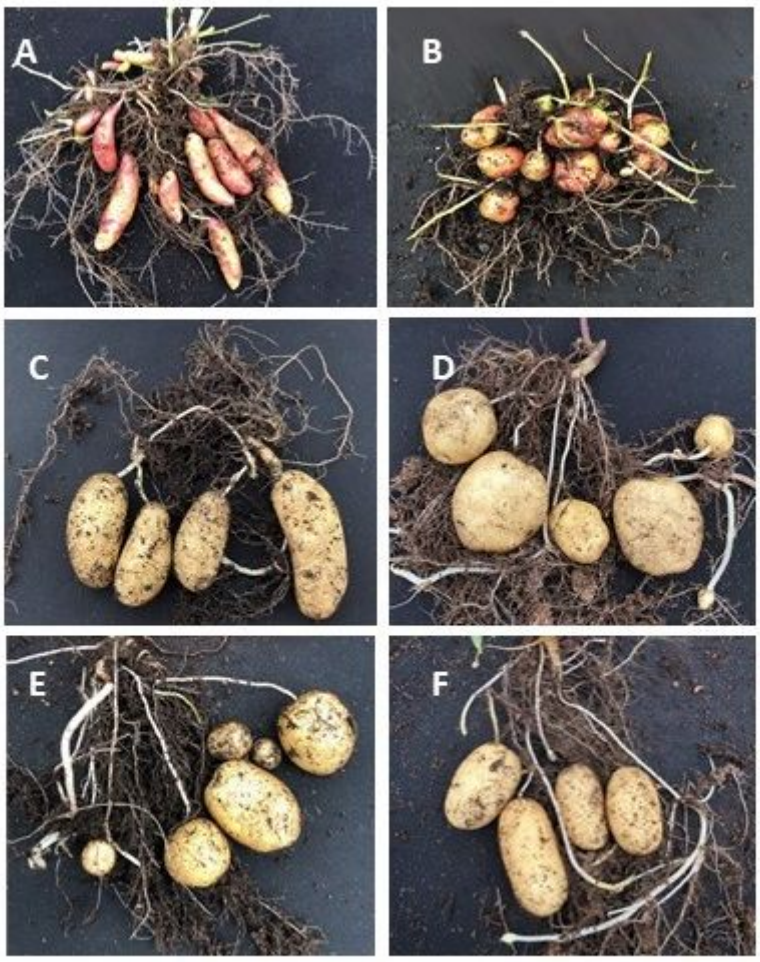

Figure 5

StOFP20 regulation of potato tuber shape. Wild type potato clone DM showing long tuber shape (A), DM transformed with *StOFP20.2* expressed from a CaMV 35S promoter showing round tubers (B). Wild type potato cultivar Spunta (C) with long tubers, Spunta transformed with *StOFP20.2* expressed from a CaMV 35S promoter showing round tuber shape (D). Wild type cultivar Atlantic with round tubers (E), Atlantic transformed with an RNAi construct for *StOFP20* showing elongated tuber shape. Images were taken from individual plants showing the strongest phenotypic effect.

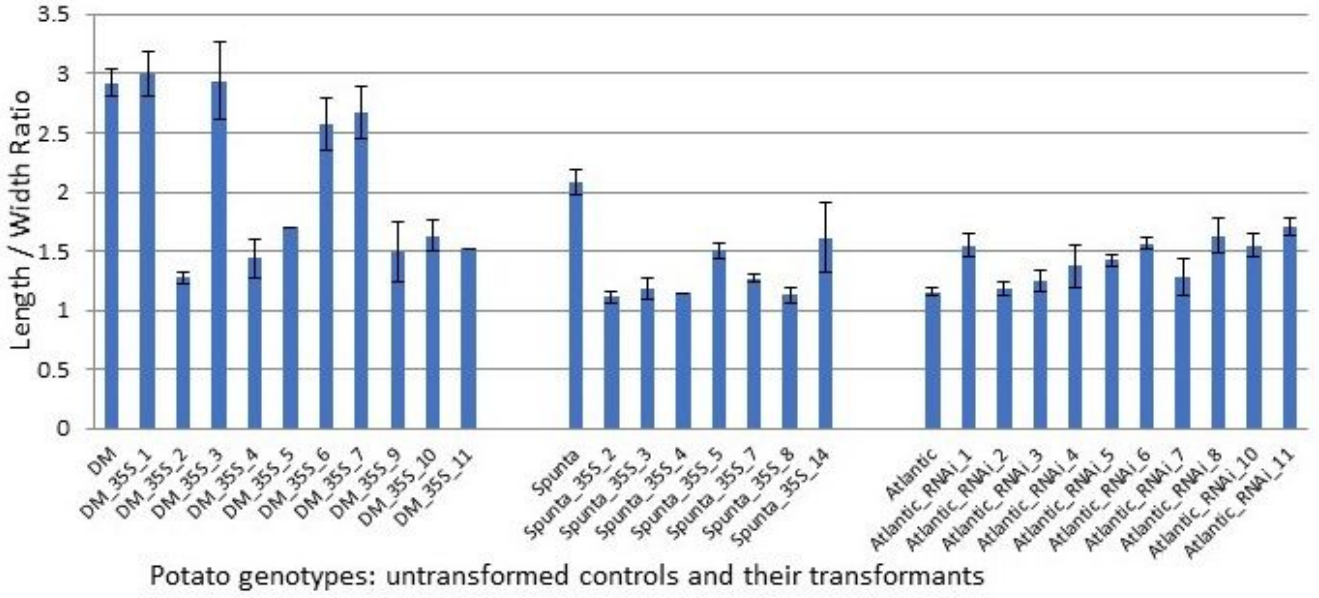

Figure 6

Significant changes in tuber shape measured as length/width (L/W) ratio were induced by transformation of Spunta and DM with *StOFP20* expressed from a CaMV 35S promoter, or by transformation of Atlantic with an RNAi construct.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarydataS1ScOFP20andthe42kbindel.docx
- SupplementarydataS2StOFP20haplotypesinfasta.docx
- SupplementarydataS3StOFP20haplotypesproteinalign.docx
- SupplementarydataS4Varietypanel.docx
- SupplementarydataS5SNPcalling136varieties.xlsx