Aug 13th, 5:45 PM - 6:00 PM

RNA-seq combined with a bulked-segregant analysis identifies candidate genes for the waxy coating on blueberry fruit

Xinpeng Qi
USDA, Agricultural Research Service, Xinpeng.Qi@ars.usda.gov

Elizabeth L. Ogden
USDA, Agricultural Research Service, Elizabeth.Ogden@ars.usda.gov

Jose V. Die
University of Cordoba, Spain, jodiera@alumni.upv.es

Mark K. Ehlenfeldt
USDA, Agricultural Research Service, mark.ehlenfeldt@ars.usda.gov

James J. Polashock
USDA, Agricultural Research Service, james.polashock@ars.usda.gov

See next page for additional authors

RNA-seq Combined with Bulked-Segregant Analysis Identifies Candidate Genes for the Waxy Coating on Blueberry Fruit

Xinpeng Qi[1], Elizabeth L. Ogden[1], Jose V. Die[1], Mark K. Ehlenfeldt[2], James J. Polashock[2], and Lisa J. Rowland[1*]

[1]USDA-ARS, BARC-West, Genetic Improvement of Fruits and Vegetables Laboratory, Beltsville, MD 20705
[2] USDA-ARS, Blueberry and Cranberry Research Center, Genetic Improvement of Fruits and Vegetables Laboratory, Chatsworth, NJ 08019

Abstract
The most significant difference between blueberries with a light blue fruit color and black fruit color is the visible layer of an epicuticular waxy coating. This layer functions in disease defense and prevention of water loss. In this study, we constructed a northern-adapted rabbiteye hybrid breeding population, 'Nocturne' x T 300, which segregated for fruit color (light blue versus black). We screened this population and selected plants of each extreme phenotype, waxy-coated plants with light blue colored fruit versus non-waxy plants with black colored fruit, then isolated RNA from fruit tissue of each bulk, respectively. We sequenced the transcriptome of each bulk using RNA-seq, which resulted in a total of 167,093,354 reads for both libraries combined. We *de novo* assembled this data set into 171,678 contigs and used the assembled transcriptome as a reference for differential expression analysis using EdgeR. A total of 515 differentially expressed genes were identified with at least a four-fold difference in expression, and efforts were made to functionally annotate them using publicly available databases. From these, one excellent candidate 'waxy' gene has emerged, which we are investigating further.

Introduction

There are three main species of blueberries grown commercially, the tetraploid highbush blueberry (*Vaccinium corymbosum* and hybrids thereof), the hexaploid rabbiteye blueberry (*V. ashei*, syn. *V. virgatum*), and the wild tetraploid lowbush blueberry (*V. angustifolium*). A major focus in breeding of highbush and rabbiteye blueberry is developing cultivars with high fruit quality. Blueberry fruit quality includes attributes such as good flavor, large size, firmness, light blue color, and small scar. The color of blueberries is due to the presence of anthocyanins within the fruit skin and to a cuticular waxy layer on the outside of the fruit. This waxy coating or "bloom" gives a desirable light blue color, which is visually appealing and preferred for the fresh market (Nunes et al. 2004). The waxy coating also retards fruit desiccation and deterioration in storage (Albrigo et al. 1980).

In this study, we attempted to identify the gene(s) responsible for the protective waxy coating on blueberry fruit utilizing a unique northern-adapted rabbiteye hybrid breeding population. The population segregates visibly for the presence/absence of the waxy coating. We bulked fruit from progeny that have the waxy coating (light blue color) and those that do not (black color), extracted RNA from the two bulked groups, and performed RNA-seq. We then *de novo* assembled this data set and used the assembled transcriptome as a reference for differential expression analysis. Genes that were differentially expressed by at least four-fold between the two bulked groups were annotated, and the most likely candidate 'genes' for controlling the waxy coating were identified.

Materials and Methods

Plant material and RNA extractions

The northern rabbiteye breeding population used for this study resulted from a cross of 'Nocturne' (Ehlenfeldt et al. 2015) by T 300. The population segregated for the presence/absence of waxy coating on the fruit. 'Nocturne' is a complex mixed-species hexaploid composed primarily of *Vaccinium ashei* (syn. *V. virgatum*) and *V. constablaei*. 'Nocturne' is dark-fruited with little or no wax. T 300 is 100% *V. ashei* with light blue fruit. Photographs of typical plants from this population with and without the waxy coating are shown in Figure 1.

For RNA-seq libraries, RNA was extracted from bulked tissue samples from the population described above—10 waxy individuals and 9 non-waxy individuals were used for the bulks. To make each bulk, an equal amount of fruit tissue was used from each individual. RNA was extracted as previously described (Rowland et al. 2012), and the quality of the RNA was verified on a NanoDrop ND-1000 and run on a 1% agarose gel stained with ethidium bromide.

RNA-seq libraries, sequencing, trimming of reads and *de novo* assembly

RNA-seq libraries were prepared and sequenced at the David H. Murdock Research Institute in Kannapolis, NC, using the Illumina HiSeq2500 platform. Raw paired-end reads of 100 bp were trimmed to 85 bp by removing 10 nucleotides from the 5' ends and 5 nucleotides from the 3' ends. Reads containing more than 10 ambiguous nucleotides were discarded. Clean reads were assembled using Trinity (Grabherr et al. 2011) (version 2.1.1, --KMER_SIZE=25, --normalize_reads, --normalize_max_read_cov=60). Full-length gene hits were identified by BLASTN using a draft blueberry genome assembly and annotation as a reference (Dr. Robert Reid and Dr. Allan Brown, personal communication).

Identification of DEGs and functional annotation

Clean reads from waxy and non-waxy bulks were mapped back to the blueberry genome assembly using Bowtie (http://bowtie-bio.sourceforge.net/index.shtml). RNA-Seq Expectation Maximization (RSEM) method (Li and Dewey 2011) was applied to evaluate expression abundance of each bulked group. The EdgeR package (Robinson et al. 2010) was used to normalize expression profiles and identify differentially expressed genes (DEGs). NCBI Non-redundant protein sequence database (nr), SwissProt/Trembl (The UniProt Consortium 2017), PlantCyc (http://www.plantcyc.org/, 2016-09-19) and NCBI RefSeq (Pruitt et al. 2012) release for plants (refPlant) were downloaded and formatted into protein databases, and were used as references for BLASTP searches. A KEGG annotation was performed using the BlastKOALA (Kanehisa et al. 2016) web server (https://www.kegg.jp/blastkoala/).

Amplification and sequencing of candidate waxy cDNA(s)

Attempts were made to amplify a near full-length cDNA of the candidate waxy gene, which encodes acyl-[acyl-carrier-protein] hydrolase, from three waxy and three non-waxy plants that were included in the original bulks. Complementary DNAs were synthesized by priming with oligo-dT12–18 (Life Technologies, USA), using SuperScriptIII reverse transcriptase (Life Technologies) following the instructions of the provider. The cDNAs were diluted to a final volume of 50 µl. PCR primers were designed from the ends of the assembled transcript (Forward primer CATGCTTTCACGTTGCAGAT; Reverse primer CCGTCTCTCCTTGGATTTGA). PCR reaction volumes were 20 µL containing 1x Promega (Madison, WI) GoTaq Flexi Buffer, 3 mM $MgCl_2$, 0.2 mM each dNTP, 0.1 µM each of the forward and reverse primers, 0.5 units Promega GoTaq Flexi DNA polymerase and 1 µL cDNA. Amplification was carried out in a Bio-Rad (Hercules, CA) T100 thermal cycler with the following profile: an initial denaturation

of 95° for 5 minutes, then 40 cycles of denaturation (92°, 40s), annealing (60°, 30s), extension (72°, 60s), and a final extension step at 72° for 10 minutes.

Amplification products were purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA) and direct sequenced using the ABI Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol.

Results

Population construction and bulked RNA-seq sequencing

The cultivar Nocturne has black-colored fruit with little or no visible waxy coating (Ehlenfeldt et al. 2015). When crossed to the selection T 300, which has blue-colored fruit, a population resulted that clearly segregated for fruit color (Figure 1). We phenotyped the population for the presence/absence of the waxy coating on the fruit and bulked tissues from the two extreme types (having a heavy wax coating or no or little wax). We then constructed RNA-seq libraries of each bulk and generated a total of 167,093,354 paired-end reads (Table 1). Raw reads were quality trimmed and then used for *de novo* assembly.

Blueberry transcriptome *de novo* assembly

A total of 14.19 Gbp clean reads were used for the blueberry transcriptiome *de novo* assembly, which resulted in 171,678 'genes'. To eliminate assembly redundancy, we selected the longest isoform for each 'gene' as the representative transcript (Table 2). An average of 86.67% reads from the two libraries were successfully mapped back to this assembly, which confirmed that the quality of the assembly was good (Table 1).

Differentially expressed candidates between waxy and non-waxy bulks

A total of 515 genes with at least a four-fold ($\log_2$fold-change = 2) difference in expression level with false discovery rate lower than 0.05 between the waxy library and the non-waxy library were identified (Figure 2), including 318 genes which were expressed at a higher level in the non-waxy bulk and 197 genes which were expressed at a higher level in the waxy bulk group. We then used publicly available protein databases including SwissProt, Trembl, NCBI-nr, PlantCYC, refPlant, and KEGG to annotate the differentially expressed genes. This resulted in 137 genes being successfully annotated from hits in the databases. By analyzing the functional annotation of the DEGs, a total of 10 genes were found, which appeared to possibly be related to plant epicuticular wax pathways (Table 3), including one annotated as acyl-[acyl-carrier-protein] hydrolase (Liu and Post-Beittenmiller 1995, Bonaventure et al. 2003). Because of the well-described role of this gene in wax biosynthesis in *Arabidopsis*, discussed below, and its higher expression in the waxy bulk than the non-waxy bulk, we considered it an excellent candidate for further investigation as the 'waxy' gene in our population.

We made attempts to amplify a near full-length cDNA of the acyl-[acyl-carrier-protein] hydrolase from several waxy and non-waxy plants that comprised the bulks, using PCR primers designed from the ends of the assembled transcript and standard PCR conditions. Amplification was successful from only the waxy plants of the bulked group, presumably because of the low expression of the mRNA in the non-waxy plants. The amplification products from the waxy plants were sequenced; the translated sequence from one of the waxy plants was aligned to a conserved region of the acyl-[acyl-carrier-protein] hydrolase from *Arabidopsis* and soybean (Figure 3). Sequence similarity confirmed the blueberry amplification product as coding for an acyl-[acyl-carrier-protein] hydrolase.

Discussion and Conclusions

Blueberry, rich in anthocyanins and antioxidants, is one of the most economically important berry fruits, with worldwide production increasing each year. Blueberry genomic research is expanding as well, with more and more projects being published in recent years. Regarding transcriptome studies, several reports have been published focusing on fruit ripening (Die and Rowland 2014, Gupta et al. 2015, Li et al. 2016), cold acclimation (Rowland et al. 2012, Die and Rowland 2014), flowering (Walworth et al. 2016), and fruit antioxidant content (Li et al. 2012).

In this study, we describe a blueberry transcriptome assembly of a unique breeding population segregating for the waxy coating on the fruit. By comparing expression profiles between waxy-coated and non-waxy-coated fruit tissue bulks, we identified DEGs and performed functional annotation. The best candidates for the 'waxy' gene in this population were marked for further investigation, including one annotated as the FATB class of acyl-[acyl-carrier-protein] thioesterases/hydrolases. This class of proteins determines the type and amount of fatty acids that are exported from plastids. In maize, insertions in the FATB gene (GRMZM5G829544) have been shown to result in reduced palmitic acid (16:0) content in the seeds (Li et al. 2011, Zheng et al. 2014). In *Arabidopsis*, a T-DNA insertion in the FATB gene (AT1G08510) has been shown to result in lower levels of palmitate (16:0) and stearate (18:0) in various tissue types, resulting in a reduction in growth rate. Furthermore, the FATB gene has been implicated in supplying fatty acids for wax biosynthesis, as the T-DNA insertion/knock-out mutation in *Arabidopsis* resulted in a reduction of 20% and 50% of the total wax load in leaves and stems, respectively, of the mutant (Bonaventure et al. 2003).

This gene's differential expression has been confirmed by real-time qPCR (data not shown). We are currently attempting to amplify and sequence the entire gene from waxy and non-waxy plants included in the original bulked groups, to determine if there is a structural difference in the gene in the two types of plants. Alternatively, regulation of expression of the structural gene may be responsible for the differences in the waxy coating on fruit.

Figure 1. Blueberry fruit from two representative plants of the 'Nocturne' (NCT) x T 300 population, waxy (top panel) and non-waxy (bottom panel), respectively.

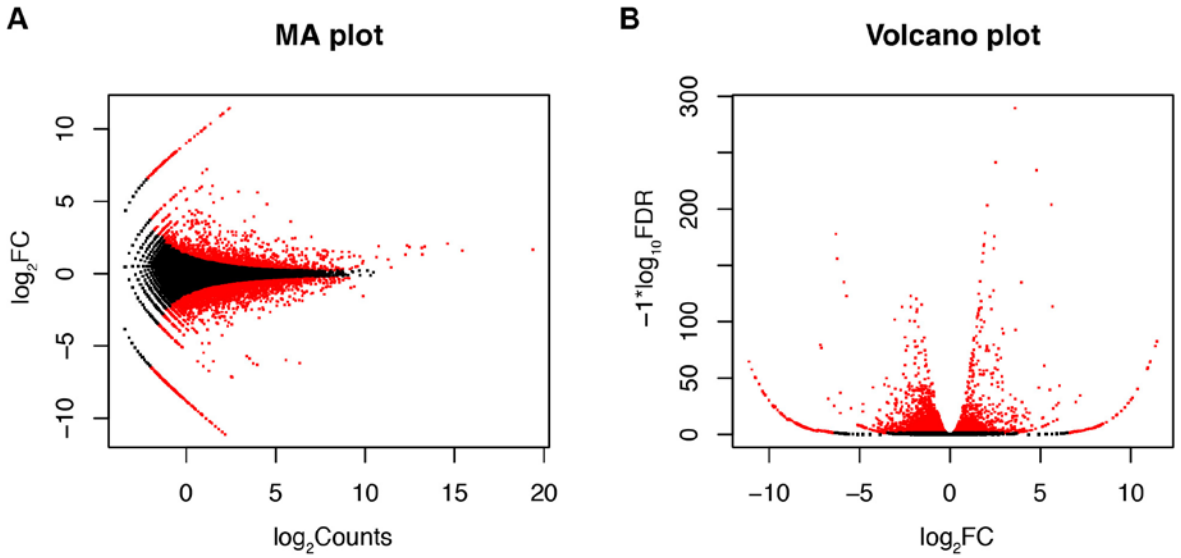

Figure 2. Expression profile of waxy versus non-waxy blueberry transcripts. A. MA plot (log_2

fold change versus $\log_2$ counts) of all assembled transcripts. B. Volcano plot of $\log_{10}$ false discovery rates versus $\log_2$ fold change. Differentially expressed genes with FDR <=0.05 were marked in red.

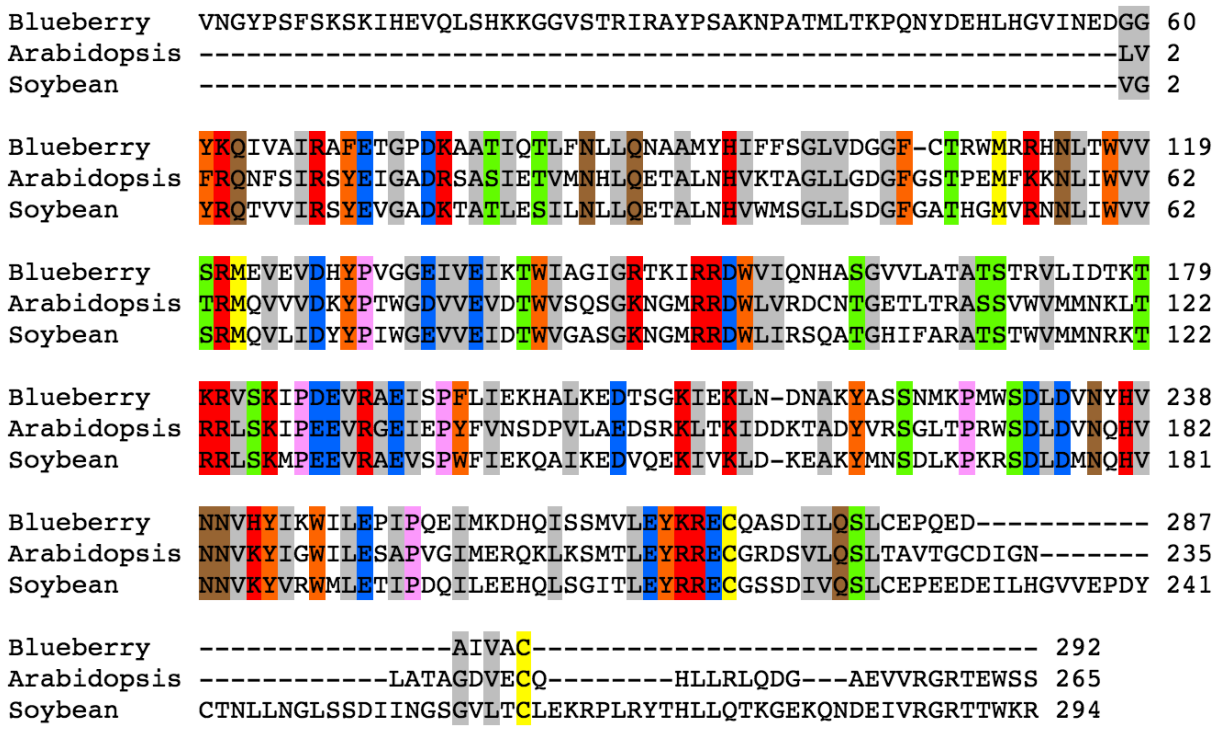

Figure 3. Alignment of deduced blueberry FATB protein sequence to *Arabidopsis* and soybean FATB protein sequences focusing on the conserved PF01643 domain of acyl-[acyl-carrier-protein] thioesterases. This domain gives the protein its catalytic function in terminating fatty acyl group extension by hydrolyzing the acyl group from the fatty acid.

Table 1. Summary of RNA-seq results from blueberry breeding population 'Nocturne' x T 300.

LibraryID	Raw Read Number	Clean Read Number	Total (bp)	Mapping Back Rate (%)
NCTxT300NonWaxy_R1	34,689,099	34,659,111	2,946,024,435	28,823,546 (83.16%)
NCTxT300NonWaxy_R2	34,689,099	34,659,111	2,946,024,435	28,728,565 (82.89%)
NCTxT300Waxy_R1	48,857,578	48,814,894	4,149,265,990	43,646,208 (89.41%)
NCTxT300Waxy_R2	48,857,578	48,814,894	4,149,265,990	43,503,450 (89.11%)

Table 2. Quality of *de novo* assembly of blueberry transcriptome

Total number of assembled 'genes'	171,678
Total length (bp)	84,121,379
N50 length (bp)	555
Mean length (bp)	490
Minimum contig length (bp)	201
Maximum contig length (bp)	27,467

Table 3. Blueberry DEGs related to plant wax biosynthesis pathways.

GeneID	Log_2 (Fold Change)[1]	False Dis-covery Rate (%)	Functional Annotation
VACC_DN29767_c0_g1	-7.67	0.279	Fatty acyl-CoA reductase
VACC_DN29767_c1_g1	-4.42	2.646	Fatty acyl-CoA reductase
VACC_DN32360_c0_g1	-3.36	2.055	Fatty acyl-CoA reductase
VACC_DN43292_c0_g1	2.91	0.007	Acyl-[acyl-carrier-protein] hydrolase
VACC_DN50504_c0_g1	-2.31	0.376	Allene oxide synthase, chloroplastic-like
VACC_DN51069_c2_g2	-2.56	1.559	Stearoyl-acyl carrier protein desaturase
VACC_DN52107_c0_g1	-2.05	2.055	Acetyl-coenzyme A synthetase
VACC_DN52309_c1_g1	2.89	0.002	Lipoxygenase
VACC_DN75555_c0_g1	-3.19	1.287	3-oxoacyl-[acyl-carrier-protein] reductase
VACC_DN92008_c0_g1	-3.92	0.072	Acyl-CoA delta (11) desaturase

[1]A negative number indicates expression is higher in the non-waxy bulk than the waxy bulk. A positive number indicates expression is higher in the waxy bulk than the non-waxy bulk.

References

Albrigo, L. G., Lyrene, P. M. and Freeman, B. 1980. Waxes and other characteristics of fruit and leaves of native *Vaccinium elliotti* Chapm. J Am Soc Hortic Sci. 105:230-235.

Bonaventure, G., Salas, J. J., Pollard, M. R. and Ohlrogge, J. B. 2003. Disruption of the FATB gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell. 15:1020-1033.

Die, J. V. and Rowland, L. J. 2014. Elucidating cold acclimation pathway in blueberry by transcriptome profiling. Environ Exp Bot. 106:87-98.

Ehlenfeldt, M. K., Martin, R. B. and Rowland, L. J. 2015. 'Nocturne' hybrid blueberry: a winter-hardy, mixed-species hexaploid with ornamental landscape interest and novel fruit quality. HortScience. 50:1825-1827.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech. 29:644-652.

Gupta, V., Estrada, A. D., Blakley, I., Reid, R., Patel, K., Meyer, M. D., Andersen, S. U., Brown, A. F., Lila, M. A. and Loraine, A. E. 2015. RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. Gigascience. 4:5.

Kanehisa, M., Sato, Y. and Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 428:726-731.

Li, B. and Dewey, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323.

Li, L., Li, H., Li, Q., Yang, X., Zheng, D., Warburton, M., Chai, Y., Zhang, P., Guo, Y., Yan, J. and Li, J. 2011. An 11-bp insertion in Zea mays *fatb* reduces the palmitic acid content of fatty acids in maize grain. PLoS ONE. 6:e24699.

Li, L., Zhang, H., Liu, Z., Cui, X., Zhang, T., Li, Y. and Zhang, L. 2016. Comparative transcriptome sequencing and de novo analysis of *Vaccinium corymbosum* during fruit and color development. BMC Plant Biol. 16:223.

Li, X., Sun, H., Pei, J., Dong, Y., Wang, F., Chen, H., Sun, Y., Wang, N., Li, H. and Li, Y. 2012. *De novo* sequencing and comparative analysis of the blueberry transcriptome to discover putative genes related to antioxidants. Gene. 511:54-61.

Liu, D. and Post-Beittenmiller, D. 1995. Discovery of an epidermal stearoyl-acyl carrier protein thioesterase. Its potential role in wax biosynthesis. J Biol Chem. 270:16962-16969.

Nunes, M. C. N., Emond, J.-P. and Brecht, J. K. 2004. Quality curves for highbush blueberries as a function of the storage temperature. Small Fruits Rev. 3:423-440.

Pruitt, K. D., Tatusova, T., Brown, G. R. and Maglott, D. R. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 40:D130-D135.

Robinson, M. D., Mccarthy, D. J. and Smyth, G. K. 2010. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139-140.

Rowland, L. J., Alkharouf, N., Darwish, O., Ogden, E. L., Polashock, J. J., Bassil, N. V. and Main, D. 2012. Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. BMC Plant Biol. 12:46.

The uniprot consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45:D158-D169.

Walworth, A. E., Chai, B. and Song, G. 2016. Transcript profile of flowering regulatory genes in VcFT-overexpressing blueberry plants. PLoS ONE. 11:e0156993.

Zheng, P., Babar, M. D. A., Parthasarathy, S., Gibson, R., Parliament, K., Flook, J., Patterson, T., Friedemann, P., Kumpatla, S. and Thompson, S. 2014. A truncated FatB resulting from a single nucleotide insertion is responsible for reducing saturated fatty acids in maize seed oil. Theor Appl Genet. 127:1537-1547.