Leveraging Transcriptome Data for Enhanced Gene Expression Analysis in Apple

Heidi Hargarten[1]
Physiology and Pathology of Tree Fruits Research Unit, 1104 North Western Avenue, U.S. Department of Agriculture, Agricultural Research Service, Wenatchee, WA 98801

Sumyya Waliullah[1]
Tree Fruit Research and Extension Center, 1100 North Western Avenue, Washington State University, Wenatchee, WA 98801

Lee Kalcsits
Tree Fruit Research and Extension Center, 1100 North Western Avenue, Washington State University, Wenatchee, WA 98801

Loren A. Honaas[2]
Physiology and Pathology of Tree Fruits Research Unit, 1104 North Western Avenue, U.S. Department of Agriculture, Agricultural Research Service, Wenatchee, WA 98801

ABSTRACT. Complex changes in gene expression occur during postharvest storage of apple (*Malus ×domestica*) and often precede or accompany changes in ripening and disorder development. Targeted gene expression analysis fundamentally relies on previous knowledge of the targeted gene. Minimally, a substantial fragment of the gene sequence must be known with high accuracy so that primers and probes, which bind to their targets in a complimentary fashion, are highly specific. Here, we describe a workflow that leverages publicly available transcriptome data to discover apple cultivar–specific gene sequences to guide primer design for quantitative real-time polymerase chain reaction (qPCR). We find that problematic polymorphisms occur frequently in 'Granny Smith' and 'Honeycrisp' apple when candidate primer binding sites were selected using the 'Golden Delicious' genome. We attempted to validate qPCR-based gene expression measurements with RNA sequencing (RNA-Seq) analysis of the same RNA samples. However, we found that agreement between the two technologies was highly variable and positively correlated with the similarity between cultivar-specific genes and RNA-Seq reference genes. Thus, we offer insight that 1) improves the accuracy and efficiency of qPCR primer design in cultivars that lack sufficient sequence resources and 2) better guides the essential step of validation of RNA-Seq data with a subset of genes of interest examined via qPCR.

Physiological disorders of apple that develop during storage contribute to significant postharvest losses (up to 30% cullage), causing significant economic losses for the apple industry (Bramlage and Watkins 1994; Doerflinger et al., 2015; Rosenberger et al., 2001). Our understanding of the molecular mechanisms of postharvest tree fruit disorders is continuously evolving (Johnson and Zhu 2015; Leisso et al., 2015; Lum et al., 2016; Sevillano et al., 2009). Disorder incidence and severity is often cultivar dependent, with some cultivars being predisposed to specific physiological defects over the course of long-term storage (Larrigaudière et al., 2016). Examining changes in gene expression during the postharvest period will 1) shed light on the molecular mechanisms underlying the physiology of tree fruit disorders; 2) be useful in classifying disorder susceptibility; 3) act as a guide for new storage strategies; and 4) improve risk assessment and management before, during, and after storage (Duan et al., 2017; Leisso et al., 2016; Nham et al., 2015).

Global-scale gene expression analysis afforded by second-generation sequencing of mRNA allows the activity of all genes to be monitored simultaneously. This powerful approach has been used extensively to deepen our understanding of numerous plant processes (Thudi et al., 2012; Voelckel et al., 2017; Wang et al., 2009). However, a primary limitation to gene activity measurements for both targeted (e.g., qPCR) and untargeted (e.g., RNA-Seq) techniques is accurate previous knowledge of gene sequences. As more genomes become available for important crop species, recently so for rosaceous specialty crops such as *Malus* (Daccord et al., 2017; Velasco et al., 2010), *Prunus* (Ahmad et al., 2011), *Pyrus* (Chagne et al., 2014; Wu et al., 2013), and *Rubus* (VanBuren et al., 2016; Ward et al., 2013), this wave of gene discovery can be leveraged for enhanced gene expression analysis. However, genetic diversity among cultivars in these specialty crops can cause a loss of fidelity in gene expression measurements due to cryptic polymorphisms or

mismatches in primers/probes or digital gene expression reference sequences (Clark et al., 2007; Takahashi et al., 2009). Development of detailed and accurate knowledge of cultivar-specific gene sequences that recover this loss in fidelity is an essential step to obtain highly accurate estimates of gene expression in non-sequenced cultivars of interest. This is especially relevant in duplicated and highly heterozygous plant genomes like apple (Daccord et al., 2017; Velasco et al., 2010), where differentiation of highly similar but unique genes (and alleles) is challenging and requires highly accurate sequence information.

qPCR is the reference standard for gene expression analysis, as the detection of transcripts is direct—oligo nucleotide primers and/or probes physically bind to a cDNA sequence of interest (Arya et al., 2005; O'Driscoll 2011; Wong and Medrano 2005; Wu et al., 2014). The highly accurate and highly sensitive nature of this interaction is a key feature of the technology, making it robust for the differentiation of highly similar sequences, even down to single-nucleotide polymorphisms (SNPs) (Gehring and Geider 2012; Jatayev et al., 2017). High specificity makes this technique susceptible to cryptic polymorphisms when, for instance, the genome of one cultivar is used for primer or probe design for another genetically distinct cultivar. The reliance on accurate previous knowledge of gene sequences has long been acknowledged as a hurdle for gene expression analysis (Freeman et al., 1999).

To create cultivar-specific and robust qPCR primers for targeted gene expression analysis, we assembled transcriptomes de novo from publicly available data for two apple cultivars (Granny Smith and Honeycrisp). We selected candidate genes using two different strategies; a forward approach in which we selected gene targets from published results (de Freitas et al., 2010, 2011) that implicated genes in a disorder of interest, and a reverse approach in which we selected genes with expression that was correlated with disorder incidence in a previous publication (Gapper et al., 2017). Using these candidates, we developed an efficient workflow (Fig. 1) that includes de novo assembly, contig selection, transcript validation, informed primer design, and qPCR assay validation. For validated transcripts, our qPCR assay success rate exceeded 95%. In addition to efficient primer design, we identified cultivar-specific polymorphisms affording a deeper understanding of how genes of interest in specific cultivars may contribute to the development of physiological disorders. We also developed strategies to enhance the necessary qPCR validation of RNA-Seq data to provide more meaningful interrogation of changes in global gene expression during postharvest storage of apple.

Materials and Methods

Transcriptome assembly using public data. Transcriptome data were retrieved from the Sequencing Read Archive (SRA) at the National Center for Biotechnology Information [NCBI (National Institutes of Health, 2018)] in compressed SRA archive format using the prefetch command. Archives were validated with the vdb-validate command, and archives were extracted with the fastq-dump command using the SRA Toolkit v2.8.2-1 (National Institutes of Health, 2018). The 'Granny Smith' apple fruit peel transcriptome was reported in Gapper et al. (2017) (SRA experiment SRP100589). The 'Honeycrisp' apple fruit peel transcriptome was reported in Leisso et al. (2016) (SRA experiment SRP081273).

Raw fastq data from the aforenamed NCBI SRA archives were quality trimmed and adapter filtered (referencing TruSeqv3 adapters; Illumina, San Diego, CA) using CLC Genomics Workbench 9.5.3 (QIAGEN, Hilden, Germany) with default parameters (e.g., trim bases less than Q20) except that all ambiguous bases were trimmed. The CLC Genomics Workbench 9.5.3 Toolkit de novo sequencing > de novo assembly function was used for de novo assembly with default parameters, except minimum contig length was decreased to 200 bp to allow inclusion of as many transcripts (and fragments thereof) as possible in our BLASTn searches. See Supplemental Table 1 for assembly statistics.

Identification of cultivar-specific gene sequences using BLAST. Gene sequences from two versions of the apple genome appear in this manuscript. Apple V1 refers to the genome of the commercially produced 'Golden Delicious' apple [Velasco et al., 2010; specifically: *Malus ×domestica* Whole Genome v1.0.p - Assembly and Annotation at the Genome Database for Rosaceae (GDR) (Jung et al., 2014)]. Apple V2 refers to the genome of a 'Golden Delicious' double haploid from INRA breeding efforts in the 1960s [Daccord et al., 2017; specifically: *Malus ×domestica* GDDH13 v1.1 - Assembly and Annotation at GDR (Jung et al., 2014)]. The steps described to follow were performed with both Apple V1 and Apple V2 (beginning with the BLASTn search of de novo assembly databases). Because the annotations of these two genomes have not been reconciled, we used Apple V1 transcripts to identify best matches to Apple V2 transcripts.

Initially, 28 candidate apple genes for 'Granny Smith' and 14 for 'Honeycrisp' were selected for this analysis (see Supplemental Files 1 and 2, final lists in Tables 1 and 2). To create our 'Granny Smith' candidate list, we used Apple V1 gene IDs from Gapper et al. (2017) to retrieve coding sequences (CDS) from GDR. To create our 'Honeycrisp' candidate list, clone sequences from de Freitas et al. (2010, 2011) were retrieved from GenBank (Benson et al., 2012) using the published NCBI clone accession numbers, or using best hits to the *M. ×domestica* expressed sequence tags (ESTs; the referenced University of California at Davis URL to retrieve ESTs was not found—see Supplemental Table 2). Apple V1 was then queried with BLASTn (via GDR, default parameters, excluding hits <95% identity) to retrieve best matching 'Golden Delicious' CDS sequences for each 'Honeycrisp' candidate. For those that did not result in good matches to Apple V1 genes, the NCBI annotation of *M. ×domestica* was similarly searched using BLASTn (see Supplemental File 3 and Supplemental Table 2).

BLASTn (built into CLC Genomics Workbench 9.5.3) was used to build nucleotide BLAST databases from the de novo transcriptome assemblies for each cultivar of interest. To retrieve cultivar-specific gene sequences, we used the known 'Golden Delicious' apple gene sequences to search each respective nucleotide BLASTn database. BLASTn parameters were default (including setting the e-value threshold to 1^{-10} and enabling the "low complexity filter"), except that the maximum number of hits was reduced from 250 to five. For each gene, we selected de novo contigs from the multiple BLASTn hits that had the best combination of high percent identity (80% to 100%), greatest subject coverage, and highest bit score. Known apple genes that did not produce sufficiently long, high-identity hits were excluded from further analysis. When a BLASTn search produced multiple good hits (e.g., two ≈600 bp, >95%
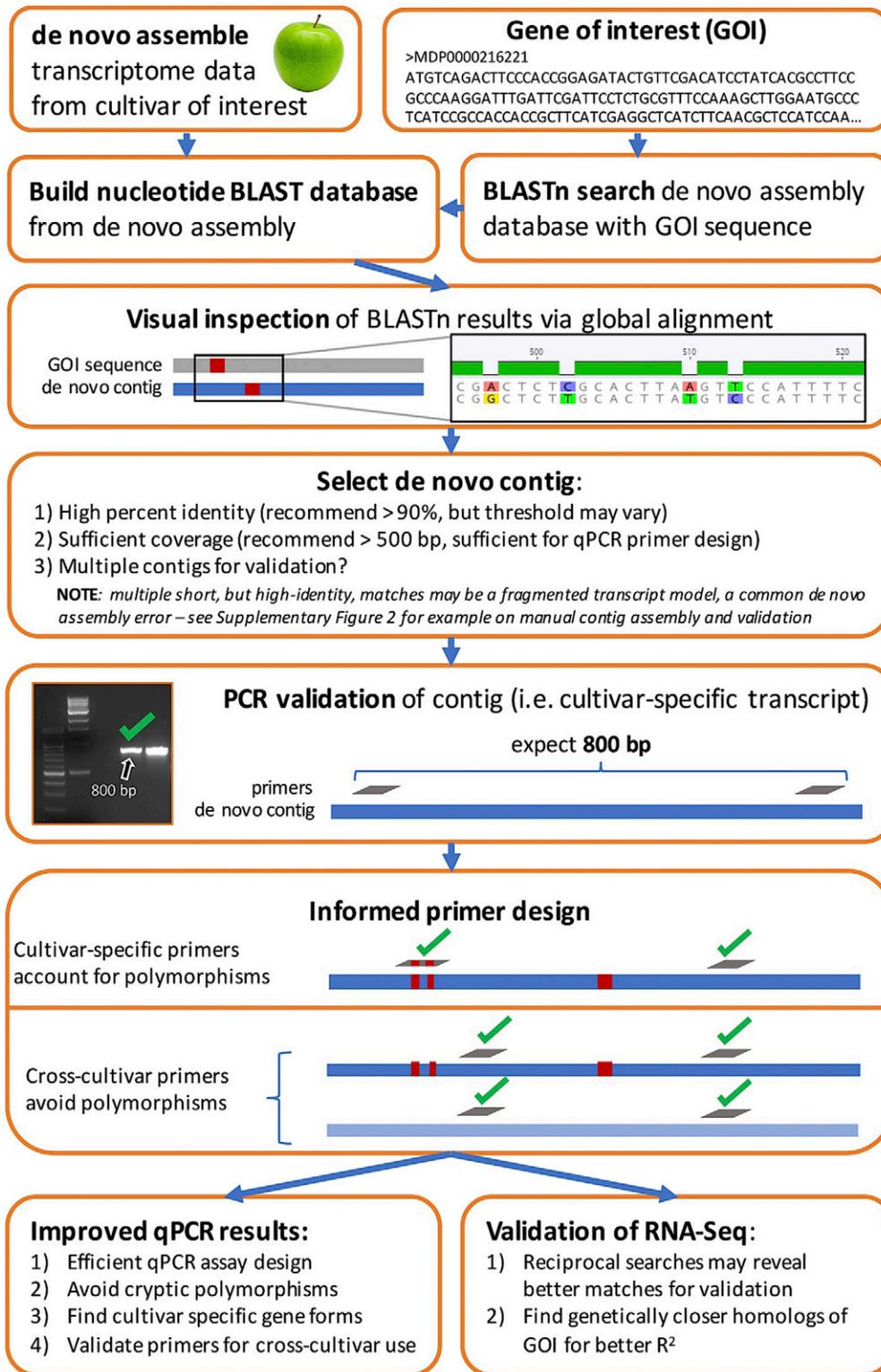
Fig. 1. Workflow for quantitative real-time polymerase chain reaction (qPCR) primer design. This workflow uses the 'Golden Delicious' reference genome and cultivar-specific de novo assemblies for enhanced qPCR primer design, with applications for gene expression and RNA Sequencing (RNA-Seq) – qPCR cross validation. PCR = polymerase chain reaction.

identify alignments to distinct portions of the query sequence), these hits were taken into subsequent steps in the analysis and either manually assembled for validation [e.g., HC_02 (see Supplemental Fig. 1)], excluded based on failure of PCR validation, or the higher identity match was chosen. The selected contigs and Apple V1 and V2 transcripts were imported into Geneious v10.1.2 (Kearse et al., 2012) for further analysis and primer design.

GLOBAL ALIGNMENTS OF REFERENCE GENES AND DE NOVO CONTIGS. Alignments of each qualifying BLASTn hit to the

Table 1. Summary of 'Granny Smith' candidate genes.[z]

Apple V1 ID	ID	qPCR primers	Gene description
MDP0000145813	GS_01	Fwd CGCAAAAAGAAGTGTTGGCC	UDP-glycosyltransferase superfamily protein
		Rev GCGTGCAAATCATAGGGACC	
MDP0000150382	GS_02	Fwd ATGGGAGATGACAAGGATGG	UDP-glucosyltransferase 74F2
		Rev TGTTATCCGAAGTGCCACC	
MDP0000165325	GS_03	Fwd CCTCGTACAACAACCGATACG	Unknown function
		Rev TTGCTTGTGACCCATCTGC	
MDP0000203897	GS_04	Fwd CTAGCCAACACCATCTTCATCG	UDP-glucosyltransferase 74F2
		Rev CAAATCCGCCTTCATCGTAGC	
MDP0000215525	GS_05	Fwd ACGGTCGAGATGGCAAAGC	UDP-glucosyl transferase 71B6
		Rev TCAGTGTTGGTGAATGGTTGGG	
MDP0000275383	GS_06	Fwd GAGCACCACATCAACAAGC	Aldolase-type TIM barrel family protein
		Rev CCTTTCGTCCGGATCATCG	
MDP0000292164	GS_07	Fwd AGTAGGATGTCTTAGCGTGC	ASP1 - aspartate aminotransferase 1
		Rev CAGGTGGGTTACTATACATGGG	
MDP0000551974	GS_08	Fwd TGTCACATATCCGGCTCAAGG	Indole-3-acetate beta-D-glucosyltransferase
		Rev GATGATGGGCTGAGAGACTGG	
MDP0000564193	GS_09	Fwd TACGAGGAATGGGCTGTCG	Senescence-associated gene 21
		Rev GATCAGAGGACGGAGGAACC	
MDP0000617956	GS_10	Fwd AGGTGTTTGGAGTTGGTCATGG	Indole-3-acetate beta-D-glucosyltransferase
		Rev CTTCACTCACAGCCTCTCTTGC	
MDP0000688645	GS_11	Fwd TTGTGCTCTATGACACCAAGGC	Aluminum-induced protein with YGL and LRDR motifs
		Rev CACCACCGAACCATCTACTGC	
MDP0000748916	GS_12	Fwd ACACTGGTTTGCAGCTTTGG	RmlC-like cupins superfamily protein
		Rev TCTGCCCTAGTTATGTCCTTGC	
MDP0000755567	GS_13	Fwd TACGAGGAATGGGCTGTCG	Senescence-associated gene 21
		Rev GATCAGAGGACGGAGGAACC	
MDP0000858559	GS_14	Fwd AGGAGGGGAGAACAAGAGC	Seed maturation protein
		Rev CTCCACCTTGTCTTCGAGC	
MDP0000940411	GS_15	Fwd CGACTTTGTTGGACAGAATGGC	AOX1B - alternative oxidase 1B
		Rev ACTGCTGCCACTGTTTCTAGC	
MDP0000274900	GSR_01[y]	Fwd CAAGGACTGAAAGCGGGAAC	CKL2 - casein kinase 1-like protein 2
		Rev GGATACCCACGGCATAATGC	
MDP0000173025	GSR_02[x]	Fwd CCATATCCAGGCTTGCCTAA	LTL1 - lipid transport superfamily protein
		Rev ACGAGGGTAACCTCACATGC	
MDP0000223691	GSR_03[w]	Fwd CATATTTGGCAGCAGAGCAA	Histone 1
		Rev CTCGTTAGCCAACTGCATCA	

[z]All 'Granny Smith' gene candidates were selected from Gapper et al. (2017). Annotations were obtained from Phytozome v12 (Joint Genome Institute, 2018). qPCR primers were designed in Geneious using the Primer3 plugin. Reference genes were selected from the indicated sources (descriptions were transferred from sources).
[y]Perini et al. (2014).
[x]Bowen et al. (2014).
[w]Storch et al. (2015).

query sequence was done using the Align/Assemble > Pairwise/Multiple Align function in Geneious (see Supplemental File 4). Multiple alignment algorithms were used to produce alignments that were then curated by hand. Generally, the "Geneious Alignment" and "MUSCLE Alignment" options performed well, but we note that the quality of the alignment depended heavily on sequence characteristics and therefore an exploration of aligner settings is prudent. Instances of ambiguities, SNPs (e.g., bp mismatches with regard to primer binding), splice variants, and insertions/deletions were recorded for each candidate. For all contigs, open reading frames were identified and translated for subsequent protein alignments to verify the CDS. Protein alignments were examined to find changes in the encoded proteins. For reference genes that lacked nucleotide ambiguity codes (common in the 454-based Apple V1), EMBOSS 6.5.7 (Rice et al., 2000) was used to predict secondary protein structure to scan for differences via a Geneious plugin.

Tissue handling and quality control. 'Granny Smith' peel tissue was removed from fruit at or shortly after harvest using a vegetable peeler. Three evenly spaced peels cut from the stem end to calyx end (sections of ≈20 × 3 cm) of each apple fruit were immediately flash frozen on liquid nitrogen and stored at –80 °C. Four samples in biological triplicate (a total of 12 observations—each replicate was a pool of tissues from six fruit) represent a short time course sampling scheme in the first 2 weeks of 1 °C air storage. 'Honeycrisp' peel and cortical tissue were collected from individual apples after about 6 months of storage at 1 °C in air using a 4-mm biopsy punch to an approximate depth of 6 to 8 mm, where the peel was immediately separated from the cortex using a razor blade and both tissues immediately frozen separately in liquid nitrogen,

336

J. Amer. Soc. Hort. Sci. 143(5):333–346. 2018.

Table 2. Summary of 'Honeycrisp' candidate genes.[z]

Apple V1 ID	ID	qPCR primers	Gene description
MDP0000141033	HC_01[y]	Fwd TTTGTCGATGGTTGGACC	Pectin methylesterase 44
		Rev GGCATCTGTAGAACACCG	
NA	HC_02[y]	NA	NA
NA	HC_03[y]	Fwd GATCACGTACATCCACATGG	NA
		Rev CACCTGGGATGTAATGACG	
MDP0000141033	HC_04[y]	NA	Pectin methylesterase 44
MDP0000326734	HC_05[y]	Fwd GGAAGCTCAACACAATCACC	ADPG2,PGAZAT- polygalacturonase
		Rev CTCCACTGGAAGAACAAGC	
MDP0000416548	HC_06[y]	Fwd GTTGAATGGGTAGAAGGACC	BGAL12 - beta-galactosidase 12
		Rev CACGCTCTGTCCATTTATCC	
MDP0000718765	HC_07[y]	Fwd TTGGGGCATAGTGATTCG	Pectin lyase-like superfamily protein
		Rev AGCGTACATCTCCCAATGG	
MDP0000289952	HC_08[y]	Fwd TTCCAACAACCTCCTCTCC	CAX5 - cation exchanger 5
		Rev ACAGCCATACAAGCTCTC	
MDP0000165349	HC_09[x]	Fwd GCTGTTATGTTCGCCATG	CAX5 - cation exchanger 5
		Rev GAAAAGTGTCGCTGTCTC	
MDP0000320927	HC_10[y]	Fwd CTTTTAGTGCTCGGATGTGC	CAX2 - cation exchanger 2
		Rev CGTAAAATGAAGCACAGCCG	
MDP0000230664	HC_11[y]	Fwd GGTGATATAGAGGTTTCTGG	CIF1 - autoinhibited Ca(2+)-ATPase 10
		Rev TTTCTCAGAGTTGAATGGG	
MDP0000121822	HC_12[y]	Fwd CTGTTAATGTTGCCGCTC	ACA9 - autoinhibited Ca(2+)-ATPase 9
		Rev GTTGTCTGTAGGTGGTTCG	
MDP0000248012	HC_13[y]	Fwd GGCCAACACTTCTAACATGC	VHA-A - vacuolar ATP synthase subunit A
		Rev GAAATCTCACGCAATGCC	
MDP0000022745	HC_14[y]	Fwd CTGTACTCATGACCGTTGG	AVP1 - Inorganic H pyrophosphatase
		Rev CAATGATAAGGCCAGCCC	
MDP0000213603	HCR_01[w,t]	Fwd CTCTTCTTGCTTTCACCC	EF1-alpha
		Rev GTACCCAACCTTCTTCAG	
MDP0000095375	HCR_02[v,t]	Fwd GAATCAGCAGCAGAGATG	CKB4 - casein kinase II beta subunit 4
		Rev GATATCCGACTGGCCAAC	
MDP0000223691	HCR_03[u,t]	Fwd CAGGTAGGTTGAGAAGGC	Histone-1
		Rev TGATCTGTATTCTGCTTCGG	

[z]'Honeycrisp' candidates were selected from the literature. Annotations were obtained from Phytozome v12 (Joint Genome Institute, 2018). qPCR primers were designed in Geneious using the Primer3 plugin. Reference genes were selected from the indicated sources (descriptions were transferred from sources).
[y]de Freitas et al. (2010).
[x]de Freitas et al. (2011).
[w]Perini et al. (2014).
[v]Bowen et al. (2014).
[u]Storch et al. (2015).
[t]Primers designed based on contigs.
qPCR = quantitative real-time polymerase chain reaction

then stored at –80 °C. Four samples, in biological triplicate (a total of 12 observations—where each replicate was a pool of tissues from 10 fruits) represent peel and cortical tissues each from fruit with and without bitter pit lesions. RNA was extracted using a CTAB/chloroform protocol modified specifically for pome fruit tissue (Honaas and Kahn, 2017). Extracted RNA was analyzed for quantity and purity using a spectrophotometer (NanoDrop ND-1000; Thermo Fisher Scientific, Waltham, MA) and for quantity and integrity on an automated electrophoresis system (Bioanalzyer 2100 G2938C; Agilent Technologies, Santa Clara, CA) with the Agilent-RNA Pico Kit (catalog no. 5067-1513). Only RNA that met the following standards were used for downstream analysis: $A_{260}/A_{280} \approx 2.0$, RNA integrity number of ≥ 8.0.

CANDIDATE GENE TRANSCRIPT VALIDATION. Putative cultivar-specific transcripts (i.e., de novo contigs) were verified by PCR where the amplicon size was maximized to capture as much of the putative transcript as possible. PCR primers were designed in Geneious using Primer3 v2.3.4 (Untergasser et al., 2012). For each candidate gene transcript, five to 10 primer pairs were designed using the parameters found in Table 3. Primer pairs were selected based on closest melt temperature (T_m), lowest hairpin and dimer T_m, longest sequence length, and percent GC content closest to 50%. Preferred primer pairs for each putative transcript were then BLASTn searched against all de novo contigs to ensure primer binding specificity. In instances in which primer pairs matched multiple contigs, new primer pairs were selected and checked in the same manner until each primer pair had a unique target in the respective de novo transcriptome assembly. Primers were synthesized by Integrated DNA Technologies (IDT, Coralville, IA), dissolved in qPCR-grade water (catalog no. W4502; Sigma-Aldrich, St. Louis, MO) to produce 100 μM solutions, and stored at –20 °C. Recommendations by IDT were used for calculating final T_m.

Table 3. PCR assay information: A summary of assay parameters used in this study for PCR primer and qPCR primer design.[z]

Parameter	Standard PCR	qPCR
Product size, bps	>80% contig length	75–150 (optimal 100)
Primer length, bps	18–28	18–22 (optimal 20)
T_m, °C	57–63 (optimal 60)	58–62 (optimal 60)
%GC	45–60 (optimal 50)	50–60 (optimal 55)
GC clamp	Preferred	Preferred
Max poly-X	5	5
Max dimer T_m, °C	40	0
Max T_m difference, °C	2	1
Assay conditions		
Denature (initial)	95 °C, 1 min (5 min)	95 °C, 10 s (30 s)
Anneal/elongate	Varies with assay	60 °C/30 s
Cycles	35–40	40
Melt curve	NA	65–95 °C at 0.5 °C increments

[z]All primers were designed in Geneious using the Primer3 plugin.
PCR = polymerase chain reaction; qPCR = quantitative real-time polymerase chain reaction; T_m = melting temperature; G = guanine; C = cytosine; %GC = percent GC content; GC clamp = presence of a GC end clamp at the end of designed primer; Max = maximum; NA = not applicable.

qPCR assay design and execution. Apple reference genes identified specifically as suitable reference genes for gene expression studies of tree fruits were selected from the literature (Tables 1 and 2; Supplemental Table 2). The following reference genes from Apple V1 MDP0000274900 (Perini et al., 2014), MDP0000173025 (Bowen et al., 2014), and MDP0000223691 (Storch et al., 2015) were used for 'Granny Smith' (Table 1) and MDP0000223691 (Storch et al., 2015), MDP0000095375 (Bowen et al., 2014), and MDP0000213603 (Perini et al., 2014) for 'Honeycrisp' (Table 2). A BLASTn search was used to find best matching contigs within respective cultivar-specific de novo assemblies. Published primers (Tables 1 and 2) were checked for perfect binding site matches or new primers were designed based on the de novo cultivar-specific sequences as for the candidate genes. All qPCR primer design followed the parameters in Table 3. Multiple primer pairs (n = 5) were scanned for binding sites that overlapped with polymorphisms in each alignment and tallied. The best qPCR primer pair for each transcript (no overlap with polymorphisms and closest to optimum parameters) was used to query the respective de novo assembly with BLASTn to ensure primer specificity. Primer sequences were synthesized by IDT and prepared as described above (see the section "Candidate gene transcript validation").

Conventional standard curve analysis to determine primer efficiency was performed for a subset of candidates (n = 4). Starting with 1 µg of total RNA, cDNA was synthesized (including DNase treatment) and a template dilution series of 10^{-1} (2 ng) to 10^{-4} (2 pg) was prepared by serial dilution. The standard curve was prepared based on the Ct value of the dilution series run and primer efficiency was calculated as described in Ginzinger (2002) (Supplemental Table 3). Reaction efficiency based on raw amplification data also were estimated using the R 3.2.1 (R Core Team, 2017) package "qpcR" (Ritz and Spiess, 2008). The "Cy0" method was used in the efficiency calculation function (Guescini et al., 2008). The reaction efficiency estimates were generally concordant (Supplemental Table 3); therefore, the "Cy0" method was used to estimate efficiencies for all remaining candidate genes (Table 4).

Total RNA (1 µg) was converted to cDNA using qScript (catalog no. 95048-025; QuantaBio, Beverly, MA) for 'Granny Smith' and iScript (catalog no. 1708840; Bio-Rad Laboratories, Hercules, CA) for 'Honeycrisp'. Template amounts were generally between 5 and 10 ng but were variable depending on amplicon yield. For 'Granny Smith', PCRs were run on a thermocycler (T100, catalog no. 1861096; Bio-Rad Laboratories) using KAPA2G Fast ReadyMix (catalog no. KK5021; Roche, Basel, Switzerland) and the manufacturer's suggested reaction protocol. For 'Honeycrisp', PCRs were run on a thermocycler (Veriti 96-Well, catalog no. 4375786; Applied Biosystems, Foster City, CA) (using the gradient feature) using EconoTaq PLUS GREEN 2X Master Mix (catalog no. 30033-1; Lucigen, Madison, WI) with the manufacturer's suggested protocol. Annealing temperatures predicted by Primer3 and IDT ± 1.0 °C defined the gradient range for all reactions. All PCRs were checked on 1.5% Tris-acetate-EDTA (TAE) or Tris-borate-EDTA (TBE) agarose gel. In situations in which at least one reaction yielded a single, expected-size product, the reaction was considered a successful validation of the transcript. Reactions that failed to generate a product, generated the wrong-sized product, or resulted in multiple products were re-run with the same reagents using touch-down protocol (touch-down PCR) where the initial annealing steps (for 10 to 15 cycles, less 0.5 °C each cycle) were ≈10 °C higher than the final annealing temperature. After touch-down PCR, any reaction that still failed to produce a clean product resulted in removal of the gene from the study.

In most cases, overlapping contig alignments were either easy to manually assemble (100% identical alignment overlap >30 bp required) or clearly different genes. In cases in which overlapping contig alignments were clearly different transcripts, the best was chosen for primer design and downstream analysis. In one instance (Supplemental Fig. 1) in which multiple short contigs aligned to query sequences with a substantial gap, the intervening sequences were determined by amplicon sequencing. The QiaQuick Gel Extraction kit (catalog no. 28704; QIAGEN) was used to extract gel-purified amplicons from PCR validation reaction (described previously) and was sequenced by Retrogen (San Diego, CA).

For both 'Granny Smith' and 'Honeycrisp' samples, total RNA was concentrated in a Concentrator plus/Vacufuge® (AG 5305; Eppendorf, Hamburg, Germany) to a minimum concentration of 125 ng·µL⁻¹. One microgram of total RNA from each sample was used to synthesize single-strand cDNA using the iScript gDNA Clear cDNA Synthesis Kit (kit includes a DNase treatment, catalog no. 1725034; Bio-Rad Laboratories). All nucleic acid samples were stored at –20 °C. For 'Granny Smith' genes, all qPCRs were run on a CFX384 Touch (catalog no. 1855485; Bio-Rad Laboratories) and for 'Honeycrisp' on a CFX96 Real-Time PCR Detection System (catalog no. 1855195; Bio-Rad Laboratories) using SsoAdvanced Universal

Table 4. Summary statistics for the best overall match between the 'Golden Delicious' reference sequences (Apple V1 and Apple V2) and the de novo 'Granny Smith' and 'Honeycrisp' contig.[z]

Gene ID	qPCR efficiency (%)	Apple V1 best match vs. de novo contig					Apple V2 best match vs. de novo contig				
		Identity (%)	Protein identity (%)	Polymorphic primer (Y/N)	qPCR vs. RNA-Seq R^2	Identity difference with next best hit (%)	Identity (%)	Protein identity (%)	Polymorphic primer (Y/N)	qPCR vs. RNA-Seq R^2	Identity difference with next best hit (%)
GS_01	98.0	94.0	72.0	Yes	0.6734	NA	100	100	No	0.6995	NA
GS_02	92.3	99.3	97.3	No	0.9198	NA	99.8	99.4	No	0.9150	NA
GS_03	94.8	99.5	97.8	No	0.8774	NA	NA	NA	NA	NA	NA
GS_04	97.6	93.8	92.2	Yes	0.9730	NA	100	100	No	0.9853	NA
GS_05	96.4	96.2	90.9	Yes	0.8947	NA	97.4	94.4	Yes	0.8559	1.3
GS_06	93.2	99.8	99.7	Yes	0.3527	5.8	91.4	95.3	Yes	0.3367	41.3
GS_07	94.6	99.9	99.5	No	0.8713	NA	100	100	No	0.8880	NA
GS_08	96.2	99.0	97.1	Yes	0.5441	2.3	99.0	97.1	Yes	0.5197	5.2
GS_09	109.7	100	100	No	0.7900	0.5	97.5	98.5	Yes	0.8120	NA
GS_10	106.5	100	100	Yes	0.9401	7.2	100	100	No	0.9316	10.3
GS_11	96.2	98.4	97.7	Yes	0.7996	NA	98.1	98.1	Yes	0.8627	NA
GS_12	96.3	97.2	94.5	Yes	0.8614	NA	99.5	98.4	Yes	0.8657	NA
GS_13	108.5	99.0	97.1	No	0.9570	NA	99.0	100	Yes	0.9651	NA
GS_14	100.3	98.3	97.4	Yes	0.7818	NA	89.0	89.7	Yes	0.3299	NA
GS_15	92.3	98.7	77.5	No	0.8279	NA	99.6	99.4	No	0.9344	NA
HC_01	108.0	NA	NA	NA	NA	NA	99.4	99.6	No	0.9050	NA
HC_02	98.1	NA	NA	NA	NA	NA	100	100	No	NA	NA
HC_03	97.8	NA	NA	NA	NA	NA	100	100	No	0.1120	NA
HC_04	108.0	NA	NA	NA	NA	NA	99.4	99.6	No	NA	NA
HC_05	98.6	97.1	96.4	No	0.3948	NA	97.7	96.3	No	0.3772	NA
HC_06	113.4	85.8	81.1	Yes	0.6706	4.0	98.5	97.4	Yes	0.7852	4.7
HC_07	102.1	99.8	100.0	Yes	0.2618	6.4	99.6	100	No	0.2638	6.4
HC_08	97.3	97.6	95.8	Yes	0.3198	NA	79.4	96.3	Yes	0.0521	1.1
HC_09	94.2	92.6	98.6	Yes	0.0908	NA	99.1	98.6	No	0.1030	6.4
HC_10	113.5	94.5	94.4	Yes	0.6824	3.6	99.8	100	No	0.6962	3.7
HC_11	106.6	98.7	98.3	Yes	0.4460	NA	98.7	98.3	Yes	0.3633	NA
HC_12	103.3	88.1	82.6	No	0.0601	NA	100	100	No	0.3345	NA
HC_13	105.3	97.7	99.7	Yes	0.4830	14.9	99.5	100	Yes	0.4645	15.4
HC_14	102.9	97.5	97.1	Yes	0.3106	NA	98.5	99.3	Yes	0.3031	1.2

[z]qPCR efficiency (%): qPCR efficiency, as calculated by the Cy0 method in the R Package qpcR; Identity (%) = Pairwise identity between nucleotide sequences of 'Golden Delicious' reference gene and cultivar-specific de novo contig; Protein identity (%) = Pairwise identity between amino acid sequences of 'Golden Delicious' reference gene and cultivar-specific de novo contig; Polymorphic Primer Y/N = whether any predicted primer sequences fall on alignment polymorphisms; qPCR vs. RNA-Seq R^2 = Pearson's R^2 correlation between qPCR relative expression and RNA-Seq reads per kilobase of transcript per million mapped reads expression levels for each biological sample; Identity difference with next best hit (%) = where improvements were found, column indicate the change in pairwise nucleotide sequence identity when a better sequence match was found upon reciprocal BLASTn search. Overall best matches were determined by two BLASTn searches for Apple V1: 1) Apple V1 into the de novo contig assembly, then 2) the de novo contig back into Apple V1, and two BLASTn searches for Apple V2: 1) Apple V1 into Apple V2, and then 2) the de novo contigs identified from Apple V1 Blast 1 into Apple V2.
qPCR = quantitative real-time polymerase chain reaction; RNA-Seq = RNA sequencing; NA = not applicable.

SYBR® Green Supermix (catalog no. 1725270; Bio-Rad Laboratories). The reaction volume was 10 μL, the template mass per reaction was 10 pg cDNA, and primer concentrations were 300 nM ('Granny Smith') and 500 nM ('Honeycrisp'). The recommended thermal cycling protocol for SsoAdvanced SYBR Green was used: activation/DNA denaturation at 95 °C for 30 s, denaturation at 95 °C for 10 s, and annealing/extension at 60 °C for 30 s for 40 cycles. A melt curve analysis was included: 65 to 95 °C at 0.5-°C increments, 5 s per step. Samples were run in Bio-Rad plastics (catalog no. HSP-3801) and sealed with optical adhesive seals (catalog no. MSB-1001; Bio-Rad Laboratories). All assays included reverse transcription–negative controls to check for genomic DNA contamination and no template controls to check for other contamination. Each reaction was run in technical triplicate. Reverse transcription–negative controls for each sample were run on a single plate, as were reactions for all three reference genes. Template amounts were optimized such that the crossing point (in the linear phase) was at or before cycle 35 of 40. Only one assay, GS_02, produced a bad (bimodal) melt curve, and even though the correlation with our RNA-Seq data was high, this assay requires further optimization. We followed the "sample maximization" scheme described by Hellemans et al. (2007), where all samples are run on each plate, with each plate containing a subset of gene tests obviating the need for interrun calibrations.

TARGETED VALIDATION WITH RNA-SEQ OF THE 29 CANDIDATE GENES. As part of ongoing and separate research efforts, we are generating transcriptome data for apple fruit. To summarize, total RNA (described previously) was provided to the Penn State Genomics Core Facility in University Park for library preparation. Libraries were constructed with 600 ng of total RNA using TruSeq Stranded mRNA Library Prep kit (catalog no. RS-122-2103; Illumina) according to the manufacturer's instructions. Libraries were sequenced on a 150-bp single-end protocol to a target volume of ≥20 million reads per biological replicate on HiSEq. 2500 in Rapid Mode (Illumina). Raw reads are available at the NCBI's SRA (SRA accession SRP150622).

Illumina read data (150 bp single-end mRNA) were analyzed with FastQC (Babraham Bioinformatics, 2018) to survey overall data quality and identify adapter sequences. This was done iteratively after data trimming and filtering with Trimmomatic v0.36 (Bolger et al., 2014) and the following command shows parameters that were sufficient to trim and filter raw data to remove adapter sequences and low quality data: - *java -jar trimmomatic.jar phred33 Sample.fastq Sample.trimmed.fastq ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:50*. Clean reads were mapped to the predicted 'Golden Delicious' apple v1.0 gene annotations obtained from Phytozome v12 (Joint Genome Institute, 2018 - Mdomestica_196_v1.0.transcript.fa, mean mapping rate 59.5% ± 1.7%) and to the double haploid 'Golden Delicious' apple version 1.1 (Bucher Laboratory, 2018; GDDH13_1–1_mrna.fasta, - mean mapping rate 84.2% ± 0.8%). Expression abundance was estimated for both using the RSEM v1.3.0 (Li and Dewey, 2011) pipeline with the inbuilt Bowtie2 (Langmead and Salzberg, 2012) read aligner option. Example commands show parameters: - *rsem-prepare-reference--bowtie2--num-threads 50 Mdomestica_196_v1.0.transcript.fa Mdomestica & - rsem-calculate-expression --phred33-quals--num-threads 50 --bowtie2 Sample.trimmed.fastq Mdomestica Sample.*

DATA ANALYSIS. The "sample maximization" experimental set-up for multiplate qPCR studies is a prerequisite for creation of a "Gene Study" using the CFX Maestro Software (Bio-Rad Laboratories) in which plate data are combined to evaluate relative expression across experimental groups for each candidate gene. The CFX Maestro Software was used to analyze all qPCR data (crossing points in linear range) using the Pfaffl method (Pfaffl 2001) with three reference genes and accounting for primer efficiencies for each qPCR primer pair considered. For graph preparation and statistical analysis, R 3.2.1 (R Core Team, 2017) was used.

Relative normalized expression values (calculated using the CFX Maestro Software) from qPCR experiments were correlated with normalized digital count data (reads per kilobase of transcript per million mapped reads) for each triplicated apple fruit biological sample (n = 4 for both 'Granny Smith' and 'Honeycrisp'). Linear regression analysis of each observation (four biological samples in triplicate) was performed for each candidate gene ($\alpha = 0.05$). If $R^2 \leq 0.80$, the cultivar-specific transcript was used to query the annotated genes in Apple V1 and Apple V2 with BLASTn (default parameters) to search for other high-identity gene matches with better agreement to our qPCR gene expression estimates. These secondary matches were analyzed as described previously.

Results

DE NOVO ASSEMBLY OF PUBLICLY AVAILABLE TRANSCRIPTOME DATA. The first step in our workflow to design efficient qPCR assays was to de novo assemble publicly available fruit transcriptome data for 'Granny Smith' [Gapper et al., 2017 (SRA experiment SRP100589)] and 'Honeycrisp' [Leisso et al., 2016 (SRA experiment SRP081273)]. The 'Granny Smith' assembly of 355 million reads resulted in ≈68,000 contigs with >9000 contigs over 750 bp in length, and the 'Honeycrisp' assembly of 254 million reads resulted in ≈50,000 contigs with >4500 contigs over 750 bp in length (Supplemental Table 1). As a quality check, reads were mapped back to the cultivar-specific assembly (see Honaas et al., 2016); >80% of reads mapped back, indicating an assembly that was representative of the input data and therefore high quality (Supplemental Table 1). The CLC Genomics Workbench de novo transcriptome assemblies were used without extensive postprocessing to maximize recovery of cultivar-specific transcripts (or sufficiently large fragments thereof) because no contigs would be removed during any cleanup or refinement steps.

GENE OF INTEREST AND CONTIG SELECTION. Candidate gene mRNA transcripts for the cultivar 'Golden Delicious' were obtained from GDR by simple sequence retrieval in the case of 'Granny Smith' candidates. For 'Honeycrisp,' apple ESTs related to sequences in de Freitas et al. (2010, 2011) were used to search 'Golden Delicious' nucleotide database to retrieve candidates (see Supplemental File 2 for BLASTn results). These sequences were then used to query cultivar-specific transcriptome assemblies with BLASTn (see Supplemental Files 3 and 4 for BLASTn results). From the multiple hits, cultivar-specific contigs (putative transcripts or fragments thereof) were selected, resulting in an initial candidate list of 25 'Granny Smith' genes and 14 'Honeycrisp' genes.

TRANSCRIPT VALIDATION. Contigs that resulted from de novo assembly are essentially a hypothesis about oligonucleotides present in the sample from which a library was made. In the

case of RNA-Seq, these oligonucleotides are typically enriched for mRNA [selected by hybridization of polyadenylated mature mRNA transcripts to oligo(dT) probes]; thus, the majority of de novo assembled contigs are hypothetical mRNA transcripts. Therefore, to quickly test these hypotheses, conventional PCR primers were developed for all candidates and tested for expected amplicon size (see Supplemental File 5 for primer sequences). A majority of 'Granny Smith' candidates (15 of 25) readily yielded clean amplicons of the expected size (Supplemental Fig. 2A). All 'Honeycrisp' candidates (n = 14) passed this validation step with clean and specific PCR tests (Supplemental Fig. 2B). This indicates that the de novo contigs were accurate hypotheses of cultivar-specific transcripts.

INFORMED PRIMER DESIGN: CULTIVAR-SPECIFIC TRANSCRIPTS ARE POLYMORPHIC. We next examined alignments to 'Golden Delicious' sequences of the 15 validated 'Granny Smith' candidates and 14 validated 'Honeycrisp' candidates. Compared with Apple V1, the cultivar-specific transcripts (or fragments thereof) produced global alignments that ranged from 88.1% to 100% nucleotide identity and for Apple V2 ranged from 79.4% to 100% nucleotide identity (Table 4; Supplemental File 4). Referencing Apple V1, the overall average of nucleotide identity was 96.9% ± 3.7% ('Granny Smith' 98.2% ± 2.0%, 'Honeycrisp' 94.9% ± 4.7%). For Apple V2, the average identity was slightly greater at 97.8% ± 4.4% ('Granny Smith' 97.9% ± 3.4%, 'Honeycrisp' 97.8% ± 5.3%), likely due to resolution of ambiguous base calls. Approximately 28% of the candidate genes contained insertions or deletions compared with the Apple V1 and Apple V2 transcripts, including a deletion that produced an open reading frame for a divergent and extended C-terminal poly-peptide for GS_15. A majority of candidates (82.8% for Apple V1 and 85.7% for Apple V2) had SNPs (Table 4, Supplemental File 4). SNPs were observed in potential primer binding sites in 72.0% and 46.6% of Apple V1 and Apple V2 candidates, respectively (examples in Fig. 2, Table 4, and Supplemental Fig. 3). Referencing Apple V1, 65.4%, and Apple V2, 75.0%, of candidate genes had SNPs that altered predicted protein sequences (Table 4; Supplemental File 4). The protein identity ranged from 71% to 100% between contigs and Apple V1 transcripts and between 89.7% to 100% for Apple V2 (Table 4; Supplemental File 4). For those candidates lacking ambiguity codes, the amino acid sequence (thus allowing secondary structure prediction) changes altered the prediction of secondary protein structure 90% of the time (Supplemental File 4).

In one 'Honeycrisp' example (HC_02), two short contigs aligned at either end of a single 'Golden Delicious' gene obtained from GenBank with high sequence similarity (>99%), yet a substantial gap remained in the middle of the reference transcript (Supplemental Fig. 1). This putative transcript was validated with conventional PCR, the resulting amplicon was sequenced, and it was aligned with 99.0% identity to the NCBI Expressed Sequence Tag cDNA clone and 100% identity with the GenBank reference apple genome transcript model (Supplemental File 3). This case demonstrates that even partial information can be successfully leveraged to develop primers and probes for cultivar-specific genes of interest. Also, contigs were manually assembled for HC_05 and HC_14 based on overlapping alignments that were validated by PCR (Supplemental File 4).

PRIMER BINDING SITES FREQUENTLY CONTAIN POLYMORPHISMS. Approximately 60% of the putative qPCR primers spanned polymorphisms between the reference 'Golden Delicious' sequence and the cultivar-specific de novo contig (Supplemental File 4). This is likely an overestimate of polymorphic primer sites, given the use of one primer pair per gene. However, because of the nuanced nature of selecting primers and the frequent practice of evaluating multiple primer pairs, we reported the frequency of potentially problematic sites. Ambiguity codes in reference gene models from Apple V1 masked superior primer binding sites in ≈10% of the candidate transcripts (Supplemental File 4).

AVOIDING PITFALLS IN CROSS-TECHNOLOGY VALIDATION OF GENE EXPRESSION DATA. Transcriptome data were used to validate the qPCR results for the biological samples used in this study. Since cross-platform validation is a critical step for RNA-Seq experiments, a similar analysis was used for this cross-cultivar mapping experiment where we mapped Illumina transcriptome data from 'Granny Smith' and 'Honeycrisp' to the 'Golden Delicious' genome. A correlation analysis of relative expression and normalized count data, reads per kilobase of transcript per million mapped reads, mapping against Apple V1 and Apple V2 showed that in nearly one-half of the candidates, the estimates agreed ($R^2 > 0.8$) (Table 4; Supplemental Fig. 4A and B). For most of the remaining candidates in this study, lower, but still significant, positive correlations between RNA-Seq and qPCR were observed (Table 4; Supplemental Fig. 4A and B).

We hypothesized that the correlation between RNA-Seq and qPCR could be influenced by sequence similarity between qPCR targets and RNA-Seq reference sequences. Furthermore, because draft genome annotations and de novo transcriptome assemblies were incomplete, reciprocal searching for better matches to candidates revealed matches that showed improved agreement between qPCR and RNA-Seq. Apple V1 and Apple V2 transcripts were searched with BLASTn using cultivar-specific contigs. Generally, where BLASTn hits showed a higher sequence identity, the correlation between qPCR and RNA-Seq improved (Supplemental Fig. 4A and B). The improvement in agreement ranged from dramatic (Fig. 3A) to nominal (Fig. 3B). When we filtered out alignments of cultivar-specific genes that covered less than 75% of the reference sequences, we observed a positive correlation ($R^2 = 0.60$) between agreement of qPCR vs. RNA-Seq estimates of gene expression and alignment identity (Supplemental Fig. 5). Although concordant with our other analyses, this relationship may have cultivar-specific characteristics, could be nonlinear, and will require additional data to fully resolve.

Discussion

OVERVIEW. For our explorations of apple fruit cross-cultivar gene expression analysis, we ultimately selected 15 'Granny Smith' and 14 'Honeycrisp' candidate genes for analysis using qPCR. For 'Granny Smith' apple genes, a reverse approach was employed in which a subset of genes was hand selected from thousands that were identified ab initio (i.e., expression that was significantly correlated with superficial scald incidence) from a transcriptome analysis of 'Granny Smith' apple fruit (Gapper et al., 2017). For 'Honeycrisp' apple genes, a forward approach, beginning with a literature search, led to a short list of genes implicated in bitter pit (de Freitas et al., 2010, 2011).

The initial 'Granny Smith' candidates were reduced from 28 to 15 by our validation criteria. The excluded genes either lacked
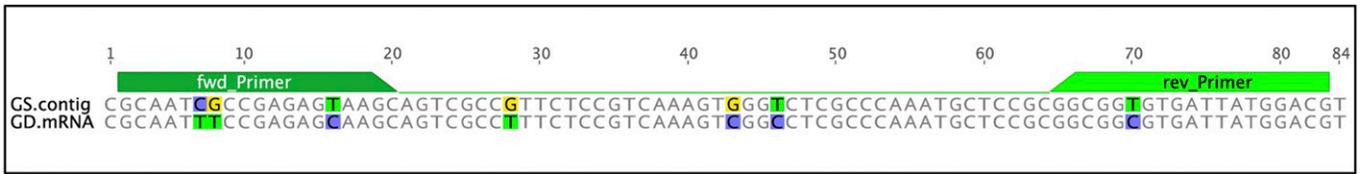
Fig. 2. Cultivar-specific genes are polymorphic compared with 'Golden Delicious' reference genes. An alignment showing single-nucleotide polymorphisms (colored letters) between a 'Granny Smith' de novo transcriptome assembly contig (GS.contig, GS_06) and the 'Golden Delicious' predicted mRNA (GD.mRNA, MDP0000275383). Polymorphisms, especially those occurring in primer binding sites, are potentially problematic for gene expression measurements.
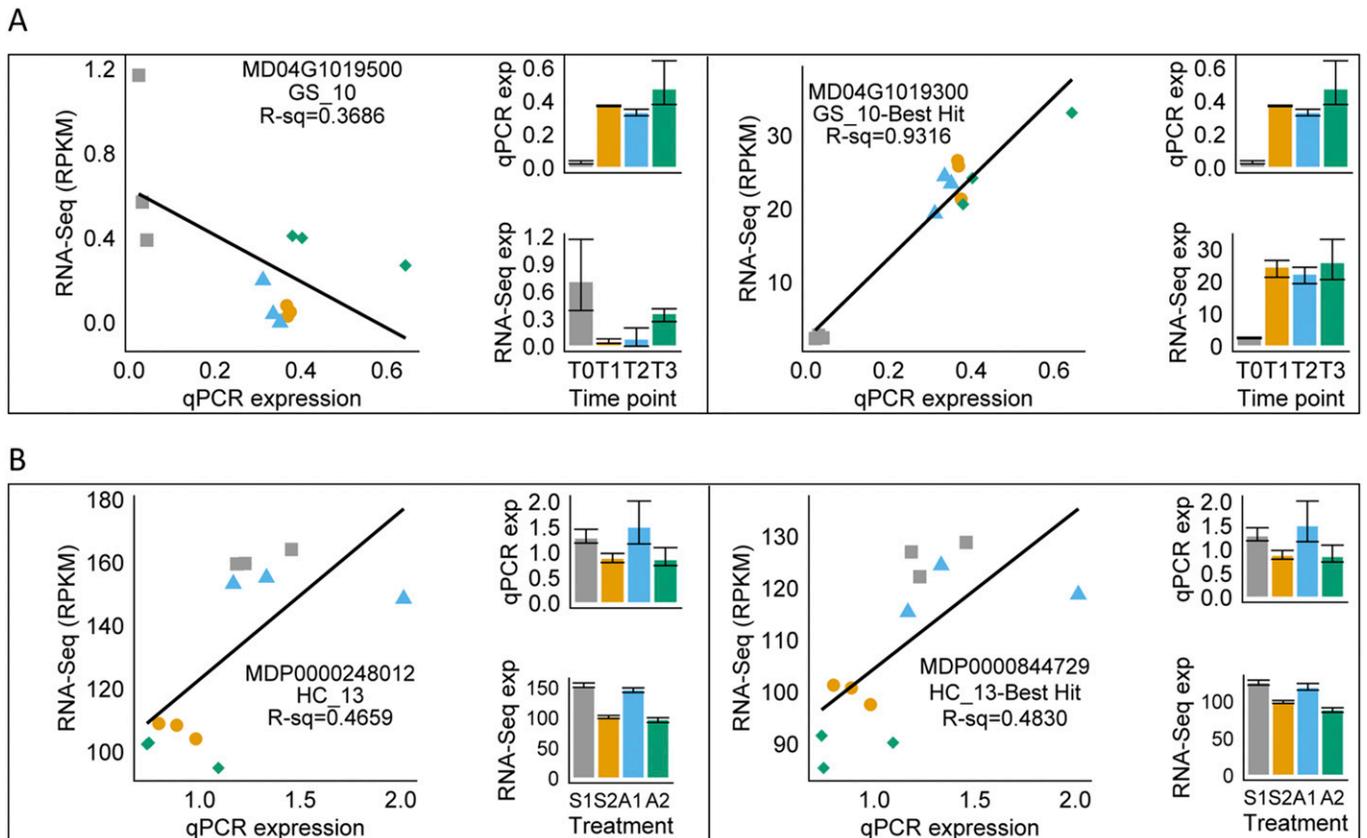

Fig. 3. Agreement between quantitative real-time polymerase chain reaction (qPCR) and RNA sequencing (RNA-Seq) is variable but can be improved in some cases (for all regressions of all candidate genes, see Supplemental Fig. 4). The change in agreement (Pearson's R^2 correlation between qPCR relative expression and RNA-seq normalized expression (reads per kilobase of transcript, per million mapped reads) between the two technologies ranged from dramatic (**A**) to minimal (**B**) when higher identity matches were found by additional and reciprocal searches of reference sequences for better matches. In (**A**), the identity increased by 10.3% (in the BestHit alignment) between matches for GS_10, resulting in a much better R^2. In (**B**), the identity increased by 14.9% (in the BestHit alignment) between matches for HC_13 with virtually no change in R^2. Biological samples in 3A represent a time–course experiment of 'Granny smith' apple fruit: T0 = at harvest; T1 = 1 week of storage in 1 °C air; T2 = 2 weeks of storage in 1 °C air; T3 = 1 week of storage in 1 °C air and then 1 week of storage at ambient temperature in air. Biological samples in 3B represent 'Honeycrisp' tissue taken after 6 months of 1 °C in air: S1 = symptomatic cortical tissue; S2 = symptomatic peel tissue; A1 = asymptomatic cortical tissue; A2 = asymptomatic peel tissue.

BLASTn search matches in the 'Granny Smith' de novo transcriptome assembly (n = 3) or failed to produce clean PCR products at the conventional PCR validation step (n = 10). All 'Honeycrisp' gene candidates passed the validation process, likely due to previous molecular validation (de Freitas et al., 2010, 2011). Four 'Honeycrisp' candidates (identified as four distinct clones in de Freitas et al., 2010) did not produce sufficiently good alignments with predicted 'Golden Delicious' genes in Apple V1. In addition, alignments to the NCBI *M. ×domestica* gene predictions of Apple V2 genes for those four 'Honeycrisp' candidates showed only two unique matches. The clones are unique, yet the two pairs are highly similar and may be alleles; the number of 'Honeycrisp' genes

remains unclear. The final 29 candidate genes (Tables 1, 2, and 4; Supplemental Files 1 and 5) were subject to our full analysis that resulted in successful qPCR assay development for 28 candidate genes, a success rate exceeding 95%. The sole assay, which requires further optimization, produced a bimodal melt curve (GS_02), perhaps due to cryptic alleles present in the qPCR amplicon.

Second-generation sequencing data can be leveraged for gene discovery. Gene discovery in apple was stimulated with the release of the apple genome (Daccord et al., 2017; Velasco et al., 2010). However, substantial genetic diversity among apple cultivars likely presents a hurdle to gene expression analysis via loss of signal fidelity due to cryptic polymorphisms.

342

J. Amer. Soc. Hort. Sci. 143(5):333–346. 2018.

Indeed, genomic surveys of apple during the development of a genotyping array showed more than 15 million variants detected across 63 apple cultivars. However, only 3.2% are represented in the final array (Bianco et al., 2016). A majority of gene candidates in this study (selected for their potential involvement in economically relevant disorders of apple fruit) are sufficiently divergent in coding sequences to impede targeted gene expression analysis (Lefever et al., 2013; Wu et al., 2009). We resolved these differences by gene discovery in cultivars of interest using publicly available transcriptome data. Transcriptome assembly is a more efficient approach to explore gene space (i.e., discover gene sequences) compared with genome assembly (De Wit et al., 2012), and in our case easily afforded by publicly available data (Gapper et al., 2017; Leisso et al., 2016). Using a proven de novo transcriptome assembly tool, CLC Genomics Workbench, accurate cultivar-specific gene models were developed. Thus, a workflow for enhanced targeted gene expression analysis was built around the new reference transcriptomes (Fig. 1).

Upfront informatics avoids pitfalls of cryptic polymorphisms. SNPs have been shown to have dramatic effects on qPCR experiments, with four mismatches able to completely block amplification (Lefever et al., 2013). Fewer than four mismatches can be tolerated, although less so when mismatches occur nearer to the 3′ end of the primer. Just one SNP at a primer binding site can cause clear and significant effects on gene expression measurements (Lefever et al., 2013). To avoid these potential pitfalls, transcriptomes were assembled de novo to discover cultivar-specific versions of apple genes of interest. De novo transcriptomes were previously shown to be highly accurate with regard to base-call errors (Honaas et al., 2016). Therefore, we reasoned that de novo assembled 'Granny Smith' and 'Honeycrisp' transcripts would be excellent targets for high accuracy qPCR assay design. Here, we report that previously unknown transcript polymorphisms in 'Granny Smith' and 'Honeycrisp' compared with 'Golden Delicious' apple, occur in candidate primer binding sites. In this study, qPCR assays showed high specificity and high efficiency often with good agreement to RNA-Seq estimates of the same genes. Potential assay failures/problems due to cryptic polymorphisms were avoided by leveraging freely available public data.

Cultivar-specific gene discovery reveals altered protein sequences. Polymorphisms in cultivar-specific genes were discovered that result in alterations of encoded proteins. SNPs that are silent and do not specify a different protein sequence are still equally problematic for primer design. Furthermore, protein sequence changes have the potential to alter biological activity of the protein. In addition to learning about nucleotide sequences to inform primer design, the biological significance of candidate genes can be enhanced by learning about the alterations in encoded proteins. In approximately two-thirds of cases, the cultivar-specific genetic differences resulted in altered protein-coding sequences. Although is it purely speculative to assume some biologically relevant change in these cases, it does provide additional information that will guide future experiments and interpretations thereof. Importantly, this information would remain cryptic without the gene discovery afforded by de novo assembly.

RNA-Seq validation. It is common practice to validate RNA-Seq data with qPCR data. Several studies report comparisons between RNA-Seq and qPCR that found highly significant, positive correlations between the two methods in a diverse set of experiments (Asmann et al., 2009; Griffith et al., 2010; Gusberti et al., 2013; Wu et al., 2014; Xu et al., 2017). This cross-platform check is a critical step because two concordant estimates of gene expression from fundamentally different technologies is robust validation. Here, we validated our gene expression estimates because we had available RNA-Seq data to examine correlations in cross-cultivar RNA-Seq experiments for our candidates.

Initially, several candidates showed poor correlations between cultivar-specific qPCR assays and cross-cultivar RNA-Seq. Because the qPCR assays were highly specific (with clean melt curves and no off-target primer sites detected in the de novo assembly) and highly efficient (mean efficiency 100.8 ± 6.3), we suspected that the poor correlations were due to issues with the RNA-Seq analysis rather than the qPCR assay, likely due to high genetic diversity among apple cultivars (Bianco et al., 2016). By searching for reciprocal best hits, better gene matches were often found in the 'Golden Delicious' genome (Table 4; Supplemental Fig. 4A and B). Although we frequently found better matches, the higher identity matches did not always result in proportional increases in correlation coefficients (for example, see Fig. 3A vs. Fig. 3B). The presence of suboptimal matches is likely due to an incomplete annotation of the 'Golden Delicious' genome, genome mis-assembly (e.g., missing genes), transcriptome mis-assembly, or a combination thereof. Each of these could result in inferior matches when searching the de novo transcriptome assembly with predicted 'Golden Delicious' genes and vice versa. That closely related genes were often found in our analysis illustrates a key issue of large, complex gene families (that contain many highly similar members) in duplicated plant genomes (Jiao et al., 2011); this problem is confounded by high heterozygosity in apple (Daccord et al., 2017).

It is unclear why, in general, agreement between RNA-Seq and qPCR for the 'Granny Smith' candidates was better than for the 'Honeycrisp' candidates. It has been shown that even in isogenic comparisons, individual gene characteristics (e.g., gene length and expression level) can influence agreement between the technologies (Everaert et al., 2017). Thus, the difference in our study may arise from an issue of sampling effort and may require a much larger or even global-scale analysis to fully resolve. The different gene selection methods for each cultivar may also help explain the difference. Our 'Granny Smith' candidates were selected from a list of genes included in an RNA-Seq analysis, thus weeding out problematic loci—a step to which the 'Honeycrisp' candidates were not subjected. Parsing out which differences arise from genetic distinctness vs. other gene characteristics and how these interact with each technology will require additional work.

From a practical standpoint, our workflow provides an easy step to improve the agreement between RNA-seq and qPCR for a set of validation genes. Typically, validation genes are either chosen at random or represent a small set of genes of interest. qPCR is used to verify that, generally, the estimates of gene expression are concordant between the two technologies. This ranges from excellent correlations of $R^2 > 0.9$ for genes of interest (Honaas et al., 2016; Zermiani et al., 2015) to non-mathematical visualization meant to show concordance (Busatto et al., 2018). Here, we show that disagreement may

J. Amer. Soc. Hort. Sci. 143(5):333–346. 2018.

343

simply result from misidentification of the corresponding homologous gene, but we note that it is also variable from gene to gene.

Beginning a study with a gene discovery step like that described here will promote efficient and accurate qPCR assay design for genes of interest. For example, instead of choosing just 20 genes randomly for validation, we suggest selecting greater than 20, then screening them for the highest identity matches in a de novo transcriptome for the cultivar of interest. Primers could then be designed for the best (i.e., highest identity) RNA-Seq validation candidates, thus providing data with fewer artifacts resulting from cross-cultivar polymorphisms. If genes have already been chosen and qPCRs have already been run, reciprocal searches between de novo transcriptomes and reference genomes may yield matches in the genome with improved agreement to the existing qPCR data.

CONSIDERATIONS FOR DATA USAGE AND SOFTWARE SELECTION. In the present study, we used free data and closed-source software to develop our workflow. However, if free and suitable transcriptome data for a cultivar of interest is not available, there are affordable options to generate sufficient data for robust gene discovery. We previously reported that a data set of roughly 4 Gbp ($\approx$25 million 76 × 76 bp reads) was sufficient to reconstruct a majority of detected transcripts (Honaas et al., 2016) from a single tissue sample. The cost to generate such a transcriptome data set is easily on the order of the cost to develop qPCR primers for just a handful of candidates and should therefore be considered when no public data are available.

The workflow presented here does not rely on functionality unique to the closed-source software used in this study (CLC Genomics Work Bench and Geneious). The data retrieval and validation application SRA Toolkit from NCBI is open-source. Trinity (Grabherr et al., 2011) is an excellent open-source de novo assembler, and Honaas et al. (2016) have previously shown it to be among the top de novo transcriptome assemblers along with CLC Genomics Workbench and SOAPdenovo-Trans (also open-source; Beijing Genomics Institute, 2013). BLASTn, Primer3 (Untergasser et al., 2012), R, and the package we used for determining primer efficiency, "qpcR" (Ritz and Spiess 2008), are all open-source and freely available. An open source alternative for sequence alignments is MEGA (Kumar et al., 2016).

Conclusion

As genomes and transcriptomes become available for more specialty crops, opportunities to use these resources emerge. In this study, as a test group we focused on genes potentially involved in physiological disorders of apple fruit: superficial scald in 'Granny Smith' and bitter pit in 'Honeycrisp'. To develop predictive and diagnostic tests for these disorders, knowledge of molecular mechanisms underlying superficial scald and bitter pit development are essential. This knowledge relies on high-fidelity measurements of gene activity as well as high-confidence RNA-Seq validation tests. We aimed to enhance gene expression analysis, and to that end, we have shown that transcriptome data can be leveraged before assay development to avoid pitfalls, enhance efficiency, and learn about genes of interest in genetically distinct apple cultivars, as well as improve RNA-Seq validation with qPCR.

Literature Cited

Ahmad, R., D.E. Parfitt, J. Fass, E. Ogundiwin, A. Dhingra, T.M. Gradziel, D. Lin, N.A. Joshi, P.J. Martinez-Garcia, and C.H. Crisosto. 2011. Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. BMC Genomics 12(1):569.

Arya, M., I.S. Shergill, M. Williamson, L. Gommersall, N. Arya, and H.R.H. Patel. 2005. Basic principles of real-time quantitative PCR. Expert Rev. Mol. Diagn. 5(2):209–219.

Asmann, Y., E. Klee, E.A. Thompson, E. Perez, S. Middha, A. Oberg, T. Therneau, D. Smith, G. Poland, E. Wieben, and J.P. Kocher. 2009. 3′ tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. BMC Genomics 10(1):531.

Babraham Bioinformatics. 2018. FastQC. 20 Apr. 2017. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Beijing Genomics Institute. 2013. SOAP (Short Oligonucleotide Analysis Package). 11 Jan. 2018. <http://soap.genomics.org.cn/SOAPdenovo-Trans.html#faq2>.

Benson, D.A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and E.W. Sayers. 2012. GenBank. Nucleic Acids Res. 41(D1):D36–D42.

Bianco, L., A. Cestaro, G. Linsmith, H. Muranty, C. Denance, A. Theron, C. Poncet, D. Micheletti, E. Kerschbamer, E.A. Di Pierro, S. Larger, M. Pindo, E. Van de Weg, A. Davassi, F. Laurens, R. Velasco, C.E. Durel, and M. Troggio. 2016. Development and validation of the Axiom® Apple480K SNP genotyping array. Plant J. 86(1):62–74.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Bowen, J., H.S. Ireland, R. Crowhurst, Z. Luo, A.E. Watson, T. Foster, N. Gapper, J.J. Giovanonni, J.P. Mattheis, C. Watkins, D. Rudell, J.W. Johnston, and R.J. Schaffer. 2014. Selection of low-variance expressed *Malus × domestica* (apple) genes for use as quantitative PCR reference genes (housekeepers). Tree Genet. Genomes 10(3):751–759.

Bramlage, W.J. and C.B. Watkins. 1994. Influences of preharvest temperature and harvest maturity on susceptibility of New Zealand and North American apples to superficial scald. N. Z. J. Crop Hort. Sci. 22(1):69–79.

Bucher Laboratory. 2018. The apple genome and epigenome: A project led by the Bucher lab. 15 Dec. 2017. <https://iris.angers.inra.fr/gddh13/the-apple-genome-downloads.html>.

Busatto, N., B. Farneti, M. Commisso, M. Bianconi, B. Iadarola, E. Zago, B. Ruperti, F. Spinelli, A. Zanella, R. Velasco, A. Ferrarini, G. Chitarrini, U. Vrhovsek, M. Delledonne, F. Guzzo, G. Costa, and F. Costa. 2018. Apple fruit superficial scald resistance mediated by ethylene inhibition is associated with diverse metabolic processes. Plant J. 93(2):270–285.

Chagne, D., R.N. Crowhurst, M. Pindo, A. Thrimawithana, C. Deng, H. Ireland, M. Fiers, H. Dzierzon, A. Cestaro, P. Fontana, L. Bianco, A. Lu, R. Storey, M. Knabel, M. Saeed, S. Montanari, Y.K. Kim, D. Nicolini, S. Larger, E. Stefani, A.C. Allan, J. Bowen, I. Harvey, J. Johnston, M. Malnoy, M. Troggio, L. Perchepied, G. Sawyer, C. Wiedow, K. Won, R. Viola, R.P. Hellens, L. Brewer, V.G. Bus, R.J. Schaffer, S.E. Gardiner, and R. Velasco. 2014. The draft genome sequence of european pear (*Pyrus communis* L. 'Bartlett'). PLoS One 9:e92644.

Clark, R.M., G. Schweikert, C. Toomajian, S. Ossowski, G. Zeller, P. Shinn, N. Warthmann, T.T. Hu, G. Fu, D.A. Hinds, H. Chen, K.A. Frazer, D.H. Huson, B. Scholkopf, M. Nordborg, G. Ratsch, J.R. Ecker, and D. Weigel. 2007. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science 317(5836):338–342.

Daccord, N., J.-M. Celton, G. Linsmith, C. Becker, N. Choisne, E. Schijlen, H. van de Geest, L. Bianco, D. Micheletti, R. Velasco, E. Pierro, J. Gouzy, J.D.G. Rees, P. Guérif, H. Muranty, C.-E. Durel, F. Laurens, Y. Lespinasse, S. Gaillard, S. Aubourg, H. Quesneville, D. Weigel, E. van de Weg, M. Troggio, and E. Bucher. 2017. High-quality

de novo assembly of the apple genome and methylome dynamics of early fruit development. Nat. Genet. 49(7):1099–1106.

de Freitas, S.T., C.V.T. do Amarante, J.M. Labavitch, and E.J. Mitcham. 2010. Cellular approach to understand bitter pit development in apple fruit. Postharvest Biol. Technol. 57(1):6–13.

de Freitas, S.T., M. Padda, Q. Wu, S. Park, and E.J. Mitcham. 2011. Dynamic alternations in cellular and molecular components during blossom-end rot development in tomatoes expressing sCAX1, a constitutively active Ca2+/H+ antiporter from Arabidopsis. Plant Physiol. 156(2):844–855.

De Wit, P., M.H. Pespeni, J.T. Ladner, D.J. Barshis, F. Seneca, H. Jaris, N.O. Therkildsen, M. Morikawa, and S.R. Palumbi. 2012. The simple fool's guide to population genomics via RNA-Seq: An introduction to high-throughput sequencing data analysis. Mol. Ecol. Resour. 12(6):1058–1067.

Doerflinger, F.C., B.J. Rickard, J.F. Nock, and C.B. Watkins. 2015. An economic analysis of harvest timing to manage the physiological storage disorder firm flesh browning in 'Empire' apples. Postharvest Biol. Technol. 107:1–8.

Duan, N., Y. Bai, H. Sun, N. Wang, Y. Ma, M. Li, X. Wang, C. Jiao, N. Legall, L. Mao, S. Wan, K. Wang, T. He, S. Feng, Z. Zhang, Z. Mao, X. Shen, X. Chen, Y. Jiang, S. Wu, C. Yin, S. Ge, L. Yang, S. Jiang, H. Xu, J. Liu, D. Wang, C. Qu, Y. Wang, W. Zuo, L. Xiang, C. Liu, D. Zhang, Y. Gao, Y. Xu, K. Xu, T. Chao, G. Fazio, H. Shu, G.Y. Zhong, L. Cheng, Z. Fei, and X. Chen. 2017. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. Nat. Commun. 8(1):249.

Everaert, C., M. Luypaert, J.L.V. Maag, Q.X. Cheng, M.E. Dinger, J. Hellemans, and P. Mestdagh. 2017. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Scientific Rpt. 7(1):1559.

Freeman, W.M., S.J. Walker, and K.E. Vrana. 1999. Quantitative RT-PCR: Pitfalls and potential. Biotechniques 26(1):112–125.

Gapper, N.E., M.L. Hertog, J. Lee, D.A. Buchanan, R.S. Leisso, Z. Fei, G. Qu, J.J. Giovannoni, J.W. Johnston, R.J. Schaffer, B.M. Nicolaï, J.P. Mattheis, C.B. Watkins, and D.R. Rudell. 2017. Delayed response to cold stress is characterized by successive metabolic shifts culminating in apple fruit peel necrosis. BMC Plant Biol. 17(1):77.

Gehring, I. and K. Geider. 2012. Identification of Erwinia species isolated from apples and pears by differential PCR. J. Microbiol. Methods 89(1):57–62.

Ginzinger, D.G. 2002. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. Exp. Hematol. 30(6):503–512.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29(7):644–652.

Griffith, M., O.L. Griffith, J. Mwenifumbo, R. Goya, A.S. Morrissy, R.D. Morin, R. Corbett, M.J. Tang, Y.C. Hou, T.J. Pugh, G. Robertson, S. Chittaranjan, A. Ally, J.K. Asano, S.Y. Chan, H.I. Li, H. McDonald, K. Teague, Y. Zhao, T. Zeng, A. Delaney, M. Hirst, G.B. Morin, S.J. Jones, I.T. Tai, and M.A. Marra. 2010. Alternative expression analysis by RNA sequencing. Nat. Methods 7(10):843–847.

Guescini, M., D. Sisti, M.B. Rocchi, L. Stocchi, and V. Stocchi. 2008. A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. BMC Bioinformatics 9(1):326.

Gusberti, M., C. Gessler, and G.A. Broggini. 2013. RNA-Seq analysis reveals candidate genes for ontogenic resistance in Malus-Venturia pathosystem. PLoS One 8:e78457.

Hellemans, J., G. Mortier, A. De Paepe, F. Speleman, and J. Vandesompele. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 8(2):R19.

Honaas, L.A., E.K. Wafula, N.J. Wickett, J.P. Der, Y. Zhang, P.P. Edger, N.S. Altman, J.C. Pires, J.H. Leebens-Mack, and C.W. dePamphilis. 2016. Selecting superior de novo transcriptome assemblies: Lessons learned by leveraging the best plant genome. PLoS One 11:e0146062.

Honaas, L.A. and E. Kahn. 2017. A practical examination of RNA isolation methods for european pear (Pyrus communis). BMC Res. Notes 10(1):237.

Jatayev, S., A. Kurishbayev, L. Zotova, G. Khasanova, D. Serikbay, A. Zhubatkanov, M. Botayeva, A. Zhumalin, A. Turbekova, K. Soole, P. Langridge, and Y. Shavrukov. 2017. Advantages of Amplifluor-like SNP markers over KASP in plant genotyping. BMC Plant Biol. 17(2):254.

Jiao, Y.N., N.J. Wickett, S. Ayyampalayam, A.S. Chanderbali, L. Landherr, P.E. Ralph, L.P. Tomsho, Y. Hu, H.Y. Liang, P.S. Soltis, D.E. Soltis, S.W. Clifton, S.E. Schlarbaum, S.C. Schuster, H. Ma, J. Leebens-Mack, and C.W. dePamphilis. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473(7345):97–100.

Johnson, F.T. and Y. Zhu. 2015. Transcriptome changes in apple peel tissues during CO2 injury symptom development under controlled atmosphere storage regimens. Hort. Res. 2:15061.

Joint Genome Institute. 2018. Phytozome v12, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute. 8 Sept. 2017. <https://phytozome.jgi.doe.gov/pz/portal.html>.

Jung, S., S.P. Ficklin, T. Lee, C.H. Cheng, A. Blenda, P. Zheng, J. Yu, A. Bombarely, I. Cho, S. Ru, K. Evans, C. Peace, A.G. Abbott, L.A. Mueller, M.A. Olmstead, and D. Main. 2014. The Genome Database for Rosaceae (GDR): Year 10 update. Nucleic Acids Res. 42(D1):D1237–D1244.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33(7):1870–1874.

Langmead, B. and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9(4):357–359.

Larrigaudière, C., A.P. Candan, J. Giné-Bordonaba, M. Civello, and G. Calvo. 2016. Unravelling the physiological basis of superficial scald in pears based on cultivar differences. Scientia Hort. 213:340–345.

Lefever, S., F. Pattyn, J. Hellemans, and J. Vandesompele. 2013. Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. Clin. Chem. 59(10):1470–1480.

Leisso, R.S., D.A. Buchanan, J. Lee, J.P. Mattheis, C. Sater, I. Hanrahan, C.B. Watkins, N. Gapper, J.W. Johnston, R.J. Schaffer, M.L. Hertog, B.M. Nicolai, and D.R. Rudell. 2015. Chilling-related cell damage of apple (Malus x domestica Borkh.) fruit cortical tissue impacts antioxidant, lipid and phenolic metabolism. Physiol. Plant. 153(2):204–220.

Leisso, R.S., N.E. Gapper, J.P. Mattheis, N.L. Sullivan, C.B. Watkins, J.J. Giovannoni, R.J. Schaffer, J.W. Johnston, I. Hanrahan, M.L. Hertog, B.M. Nicolai, and D.R. Rudell. 2016. Gene expression and metabolism preceding soft scald, a chilling injury of 'Honeycrisp' apple fruit. BMC Genomics 17(1):798.

Li, B. and C.N. Dewey. 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12(1):323.

Lum, G.B., B.J. Shelp, J.R. DeEll, and G.G. Bozzo. 2016. Oxidative metabolism is associated with physiological disorders in fruits stored under multiple environmental stresses. Plant Sci. 245:143–152.

National Institutes of Health. 2018. The National Center for Biotechnology Information. 20 Apr. 2017. <https://www.ncbi.nlm.nih.gov/>.

J. Amer. Soc. Hort. Sci. 143(5):333–346. 2018.

345

Nham, N.T., S.T. de Freitas, A.J. Macnish, K.M. Carr, T. Kietikul, A.J. Guilatco, C.Z. Jiang, F. Zakharov, and E.J. Mitcham. 2015. A transcriptome approach towards understanding the development of ripening capacity in 'Bartlett' pears (*Pyrus communis* L.). BMC Genomics 16(1):762.

O'Driscoll, L. 2011. Gene expression profiling. Humana Press, New York, NY.

Perini, P., G. Pasquali, M. Margis-Pinheiro, P.R.D. de Oliviera, and L. Fernando Revers. 2014. Reference genes for transcriptional analysis of flowering and fruit ripening stages in apple (*Malus × domestica* Borkh.). Mol. Breed. 34(3):829–842.

Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29(9):2002–2007.

R Core Team. 2017. R: A language and environment for statistical computing. 31 Oct. 2016. <https://www.R-project.org/>.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European molecular biology open source software suite. Trends Genet. 16(6):276–277.

Ritz, C. and A. Spiess. 2008. qpcR: An R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. Bioinformatics 24(13):1549–1551.

Rosenberger, D., J. Schupp, C.B. Watkins, K. Iungerman, S. Hoying, D. Straub, and L. Cheng. 2001. Honeycrisp: Promising profit maker or just another problem child? New York Fruit Qrtly. 9(3):9–13.

Sevillano, L., M.T. Sanchez-Ballesta, F. Romojaro, and F.B. Flores. 2009. Physiological, hormonal and molecular mechanisms regulating chilling injury in horticultural species. Postharvest technologies applied to reduce its impact. J. Sci. Food Agr. 89(4):555–573.

Storch, T.T., C. Pegoraro, T. Finatto, V. Quecini, C.V. Rombaldi, and C.L. Girardi. 2015. Identification of a novel reference gene for apple transcriptional profiling under postharvest conditions. PLoS One 10: e0120599.

Takahashi, Y., K.M. Teshima, S. Yokoi, H. Innan, and K. Shimamoto. 2009. Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. Proc. Natl. Acad. Sci. USA 106(11):4555–4560.

Thudi, M., Y. Li, S.A. Jackson, G.D. May, and R.K. Varshney. 2012. Current state-of-art of sequencing technologies for plant genomics research. Brief. Funct. Genomics 11(1):3–11.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res. 40(15):1–12.

VanBuren, R., D. Bryant, J.M. Bushakra, K.J. Vining, P.P. Edger, E.R. Rowley, H.D. Priest, T.P. Michael, E. Lyons, S.A. Filichkin, M. Dossett, C.E. Finn, N.V. Bassil, and T.C. Mockler. 2016. The genome of black raspberry (*Rubus occidentalis*). Plant J. 87(6):535–547.

Velasco, R., A. Zharkikh, J. Affourtit, A. Dhingra, A. Cestaro, A. Kalyanaraman, P. Fontana, S.K. Bhatnagar, M. Troggio, D. Pruss, S. Salvi, M. Pindo, P. Baldi, S. Castelletti, M. Cavaiuolo, G. Coppola, F. Costa, V. Cova, A. Ri, V. Goremykin, M. Komjanc, S. Longhi, P. Magnago, G. Malacarne, M. Malnoy, D. Micheletti, M. Moretto, M. Perazzolli, A. Si-Ammour, S. Vezzulli, E. Zini, G. Eldredge, L.M. Fitzgerald, N. Gutin, J. Lanchbury, T. Macalma, J.T. Mitchell, J. Reid, B. Wardell, C. Kodira, Z. Chen, B. Desany, F. Niazi, M. Palmer, T. Koepke, D. Jiwan, S. Schaeffer, V. Krishnan, C. Wu, V.T. Chu, S.T. King, J. Vick, Q. Tao, A. Mraz, A. Stormo, K. Stormo, R. Bogden, D. Ederle, A. Stella, A. Vecchietti, M.M. Kater, S. Masiero, P. Lasserre, Y. Lespinasse, A.C. Allan, V. Bus, D. Chagné, R.N. Crowhurst, A.P. Gleave, E. Lavezzo, J.A. Fawcett, S. Proost, P. Rouzé, L. Sterck, S. Toppo, B. Lazzari, R.P. Hellens, C.-E. Durel, A. Gutin, R.E. Bumgarner, S.E. Gardiner, M. Skolnick, M. Egholm, Y. de Peer, F. Salamini, and R. Viola. 2010. The genome of the domesticated apple (*Malus × domestica* Borkh.). Nat. Genet. 42(10): 833–839.

Voelckel, C., N. Gruenheit, and P. Lockhart. 2017. Evolutionary transcriptomics and proteomics: Insight into plant adaptation. Trends Plant Sci. 22(6):462–471.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. 10(1):57–63.

Ward, J., B. Jasbir, F. Fernaìndez-Fernaìndez, P. Moore, J.D. Swanson, R. Viola, R. Velasco, N. Bassil, C.A. Weber, and D.J. Sargent. 2013. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics 14(1):2.

Wong, M.L. and J.F. Medrano. 2005. Real-time PCR for mRNA quantitation. Biotechniques 39(1):75–85.

Wu, A.R., N.F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M.E. Rothenberg, F.M. Mburu, G.L. Mantalas, S. Sim, M.F. Clarke, and S. R. Quake. 2014. Quantitative assessment of single-cell RNA-sequencing methods. Nat. Methods 11(1):41–46.

Wu, J., Z. Wang, Z. Shi, S. Zhang, R. Ming, S. Zhu, M.A. Khan, S. Tao, S.S. Korban, H. Wang, N.J. Chen, T. Nishio, X. Xu, L. Cong, K. Qi, X. Huang, Y. Wang, X. Zhao, J. Wu, C. Deng, C. Gou, W. Zhou, H. Yin, G. Qin, Y. Sha, Y. Tao, H. Chen, Y. Yang, Y. Song, D. Zhan, J. Wang, L. Li, M. Dai, C. Gu, Y. Wang, D. Shi, X. Wang, H. Zhang, L. Zeng, D. Zheng, C. Wang, M. Chen, G. Wang, L. Xie, V. Sovero, S. Sha, W. Huang, S. Zhang, M. Zhang, J. Sun, L. Xu, Y. Li, X. Liu, Q. Li, J. Shen, J. Wang, R.E. Paull, J.L. Bennetzen, J. Wang, and S. Zhang. 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Res. 23(2):396–408.

Wu, J.-H., P.-Y. Hong, and W.-T. Liu. 2009. Quantitative effects of position and type of single mismatch on single base primer extension. J. Microbiol. Methods 77(3):267–275.

Xu, X., M. Chen, J. Ji, Q. Xu, X. Qi, and X. Chen. 2017. Comparative RNA-seq based transcriptome profiling of waterlogging response in cucumber hypocotyls reveals novel insights into the de novo adventitious root primordia initiation. BMC Plant Biol. 17(1):129.

Zermiani, M., E. Zonin, A. Nonis, M. Begheldo, L. Ceccato, A. Vezzaro, B. Baldan, A. Trentin, A. Masi, M. Pegoraro, L. Fadanelli, W. Teale, K. Palme, L. Quintieri, and B. Ruperti. 2015. Ethylene negatively regulates transcript abundance of ROP-GAP rheostat-encoding genes and affects apoplastic reactive oxygen species homeostasis in epicarps of cold stored apple fruits. J. Expt. Bot. 66(22):7255–7270.

346

J. Amer. Soc. Hort. Sci. 143(5):333–346. 2018.