frontiers in
PLANT SCIENCE

ORIGINAL RESEARCH ARTICLE
published: 17 July 2014
doi: 10.3389/fpls.2014.00340

Sequencing and *de novo* assembly of a *Dahlia* hybrid cultivar transcriptome

Erik M. Lehnert[1,2]* and Virginia Walbot[3]

[1] *Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA*
[2] *Department of Medical Microbiology and Immunology, University of Wisconsin School of Medicine, Madison, WI, USA*
[3] *Department of Biology, Stanford University, Stanford, CA, USA*

Edited by:
Gane Ka-Shu Wong, University of Alberta, Canada

Reviewed by:
Xun Xu, BGI-Shenzhen, China
Jingfa Xiao, Chinese Academy of Sciences, China

****Correspondence:***
Erik M. Lehnert, Department of Medical Microbiology and Immunology, University of Wisconsin School of Medicine, 1550 Linden Dr., Madison, WI 53706, USA
e-mail: lehnert@wisc.edu

Dahlia variabilis, with an exceptionally high diversity of floral forms and colors, is a popular flower amongst both commercial growers and hobbyists. Recently, some genetic controls of pigment patterns have been elucidated. These studies have been limited, however, by the lack of comprehensive transcriptomic resources for this species. Here we report the sequencing, assembly, and annotation of the transcriptome of the developing leaves, stems, and floral buds of *D. variabilis*. This resulted in 35,638 contigs, most of which seem to contain the complete coding sequence, and of which 20,881 could be successfully annotated by similarity to UniProt. Furthermore, we conducted a preliminary investigation to identify contigs with expression patterns consistent with tissue-specificity. These results will accelerate research into the genetic controls of pigmentation and floral form of *D. variabilis*.

Keywords: *Dahlia*, next-generation sequencing, *de novo* transcriptome assembly, anthocyanin biosynthesis, floral homeotic factors

INTRODUCTION

Horticultural dahlias (*Dahlia variabilis*) are among the most diverse in floral form and colors of all popular garden flowers. Dahlias are also an important contributor to the more than US $100 Billion worldwide flower and potted plant market. Taxonomically dahlias are Compositae, and the center of species diversity is in Mexico and Central America. Likely through cross-breeding of two wild species followed by human selection since 1800 (Gatt et al., 1998), thousands of varieties have been generated by European horticulturists who prized novelty in overall flower size, color, petal number, and petal form.

Garden dahlias are proposed to be recent octoploid derivatives (2N = 64) (Gatt et al., 1998) from crosses between two natural species. Although this high ploidy may have been expected to result in redundancy of genetic factors, loci implicated in flower color have been elucidated and generally behave as diploid factors (Bate-Smith et al., 1955). More recently, genes that are highly expressed in flowers and encoding anthocyanin pathway enzymes have been cloned (Suzuki et al., 2002; Ohno et al., 2011b, 2013) and some enzymes characterized (Yamaguchi et al., 1999; Ogata et al., 2001). Furthermore, a partial reference transcriptome assembly was performed for a comparative genomics study of Compositae, but the majority of contigs did not contain the complete coding sequence (Hodgins et al., 2014).

Given the relatively low cost of deep transcriptome analysis using next-generation DNA sequencing, we wished to determine if reasonable descriptions of messenger RNA diversity could be obtained for leaves, stems, and floral buds of dahlia. The contigs assembled from the transcriptome data are sufficient to propose more than 20,000 likely protein coding genes in this species and can thus serve as a standard for allele comparison among dahlia varieties and for future studies of transposon-mediated variegation in anthocyanin pigmentation and the control of floral form. A cultivar containing anthocyanin in the stems and leaves and flavonoid precursors and anthocyanin in the petals was selected to assess recovery of known pigment factor transcripts as a test of the transcriptome completeness.

MATERIALS AND METHODS

PLANT GROWTH AND RNA ISOLATION

The cultivar "Rio Riata," originally obtained from Corralitos Gardens (http://www.cgdahlias.com/), was grown outdoors at Stanford University; this variety has 8 orange-red petals with yellow-tips, purple (anthocyanin pigment) stems, and green leaves tinged with purple (http://www.stanford.edu/group/dahlia_genetics/cultivars/rio_riata/rio_riata.htm). Total RNA was extracted from newly emerged stems, leaves, and buds (∼0.2–1 cm in diameter) using the RNAqueous-4PCR Kit (Ambion AM1914, Grand Island NY) following the manufacturer's instructions. The RNA-integrity number (RIN) of each sample was determined using an Agilent 2100 Bioanalyzer, Santa Clara CA), and only samples with a RIN ≥ 8 were used. Approximately 1 μg of total RNA was processed (including a poly-A^+-selection step) from the leaf, stem, and bud extracts using the TruSeq RNA Sample Prep Kit (Illumina FC-122-1001, Hayward CA) following the manufacturer's instructions to produce indexed libraries. The resulting libraries were pooled based on their indices (as described in the kit instructions). Clustering and sequencing were performed by the Stanford Center for Genomics and Personalized Medicine using an Illumina HiSeq 2000 sequencer to generate 101-bp paired-end reads. Accession numbers for reads from bud, leaf, and stem are: SRR1222985, SRR1226613, SRR1226614, respectively.

READ FILTERING, TRANSCRIPTOME ASSEMBLY, AND ANNOTATION

Transcriptome assembly was performed combining reads from all three tissues. Prior to assembly, the reads were processed as previously described (Lehnert et al., 2014) and outlined here: (1) the first 6 bp from the 5′ end of each read was discarded; (2) reads of <60 bp or containing ≥1 N were discarded; (3) low quality reads were discarded (if >10 of the first 35 bases had quality scores <30); and (4) reads were trimmed to the first position for which a sliding 4-bp window had an average quality-score of <20. The remaining read-pairs were then processed using FLASH (Magoc and Salzberg, 2011) to join reads whose ends overlapped by ≥10 bp with no mismatches. Finally, adapter sequences were removed using cutadapt (Martin, 2011) with default settings.

The processed reads were assembled using a 43-bp *k*-mer with the Velvet/Oases assembler (Velvet version 1.2.09 and Oases version 0.2.08) (Zerbino and Birney, 2008; Schulz et al., 2012) with the default settings. To choose a single representative contig from each "locus" (the Oases term for a connected component in the de Bruijn graph, which presumably consists of alternative transcripts, alleles, and extremely similar paralogs), the script "process_oases_transcripts.py" was used (Yang and Smith, 2013). This script designates as representative the contig with the highest geometric mean *k*-mer coverage that is also at least 30% as long as the longest contig within the locus. Only contigs longer than 300 bp were retained in the resulting set of representative contigs. Contigs were post-processed to remove terminal Ns, and such that internal runs of Ns were fewer than 14 bases in length.

To assign putative functional roles to the representative contigs, they were aligned to the SwissProt protein database and the NCBI Non-Redundant Protein Database (nr) using the blastx program from the standalone BLAST 2.2.25+software suite (Camacho et al., 2009) with an *E*-value cutoff of 1e-5. The Blast2GO software package (Conesa et al., 2005) was used with default settings to assign Gene Ontology (GO) terms and Enzyme codes to the predicted proteins based on their alignments to SwissProt. Blast2GO was also used to identify protein domains by InterProScan (Quevillon et al., 2005), whose associated GO terms were merged with those identified by alignment to SwissProt.

EXPRESSION ANALYSIS

The trimmed forward reads were aligned to the representative contigs using bwa (Li and Durbin, 2009). Aligned reads were counted using the samtools (Li et al., 2009; minimum mapping quality 30). The R package DESeq2 (Anders and Huber, 2010) was used to normalize read counts by library size.

RESULTS

SEQUENCING AND *DE NOVO* TRANSCRIPTOME ASSEMBLY

A set of 149,304,876 reads of 101-bp was generated in the initial sequencing run. After trimming and discarding low-quality reads, ∼73 million reads from buds, leaves, and stems (see **Table 1**) were assembled into 53,037 loci containing 122,053 contigs. After removing contigs that fell below the 300-bp length cutoff, choosing representative contigs for each locus, and removing unnecessary Ns (see Materials and Methods), 35,638 representative contigs remained. These ranged in size from 269 bp to 13,886 (see **Table 2**). These contigs have been deposited in the NCBI TSA (accession # GBDN01000001-GBDN01035638) and can also be found at (http://www.stanford.edu/group/dahlia_genetics/). 20,881 (58.5%) of contigs could be annotated by BLAST alignment to the UniProt protein database. Using these alignments and the results of InterProScan protein motif searches, 122,654 GO terms were assigned to 21,576 (60.5%) of the contigs (see **Table 3**).

COMPARISON TO KNOWN COMPONENTS OF ANTHOCYANIN BIOSYNTHESIS IN DAHLIAS AND IDENTIFICATION OF ADDITIONAL GENE COPIES

Previous research has identified and sequenced several genes of the dahlia anthocyanin biosynthetic pathway (Suzuki et al., 2002;

Table 1 | Summary of read metrics by tissue of origin.

	Total number of reads	Total nucleotides (Mb)	Average Length (bp)
Floral (paired)	12,792,456	1118	87
Floral (flash extended[a])	9,456,389	1264	133
Leaf (paired)	14,541,486	1258	
Leaf (flash extended)	13,742,631	1840	133
Stem (paired)	11,051,702	958	86
Stem (flash extended)	11,455,453	1499	130
Total	73,040,117	7938	–

[a] *Flash extended reads are paired reads that had 10-bp or more of perfect overlap at the ends of their sequence. These were joined to make longer reads, and are not included in the counts of paired reads (see Materials and Methods).*

Table 2 | Size distribution of the representative contigs.

	Representative contigs	All contigs
Number of loci	35,638	53,037
Number of contigs	35,638	122,053
Median contig size (bp)	906	943
Mean contig size (bp)	1166	1209
Minimum contig size (bp)	269	100
Maximum contig size (bp)	13,886	17,090
Total bases in assembly (Mb)	41.6	148

Table 3 | Summary of alignments to SwissProt and nr.

Number of contigs	Number (%) of contigs aligned to SwissProt[a]	Number (%[b]) of unique accessions	Number (%) of contigs aligned to nr[a]	Number (%[b]) of unique accessions
35,638	20,880 (58.6)	10,603 (50.8)	26,940 (75.6)	21,376 (79.3)

[a] *Alignments with E-value ≤ 10^{-5}.*

[b] *As % of all alignments.*

Ohno et al., 2011a). To investigate assembly quality, we compared these previously sequenced genes to the contigs in our assembly. As shown in **Table 4**, we identified all the genes previously known to be involved in the synthesis of the anthocyanidins beginning with chalcone synthase. Furthermore, we identified additional copies for several pathway genes by a best-reciprocal-blast approach, thereby expanding the dahlia gene list for the phenylpropanoid pathways. For example, we identified a total of six putative chalcone synthase genes, four more than were previously known. All contigs identified contained the complete coding sequence of the enzyme (as determined by alignment of sequence to the stop and start codon of reference protein), except for one chalcone synthase and for DvIVS, a basic helix-loop-helix transcription factor required for anthocyanin synthesis, in which the cDNA was split between two contigs.

INVESTIGATING TISSUE-SPECIFIC EXPRESSION

To achieve statistical power in assigning contigs to classes differentially expressed between tissue types, biological replicates would be required. Nonetheless, as a preliminary investigation that might guide future experimental design, we identified contigs with expression patterns consistent with tissue specificity. To do this, we used DESeq to perform a regularized log2 transformation on the read counts; this analysis generates normalized expression values after accounting for the differences in sequencing depth. We assigned a tentative tissue source to a contig by the following procedures: (1) We calculated the standard deviation of expression values across all tissues; (2) we calculated the standard deviation of expression values for each pair of tissues; (3) a contig's expression was classified as elevated or decreased in a specific tissue if the standard deviation of expression values of the pair of other tissues was 3-fold less than the standard deviation of the expression values of all three tissues and the standard deviation of the three tissues expression value was great than 10% the mean expression value. This selected for contigs that showed stable expression in two of three tissues, with a large difference in read counts in the third. **Table 4** presents this analysis for the anthocyanin biosynthetic pathway, and **Table 5** considers all classes of representative contigs.

This analysis resulted in tentative tissue-specific expression patterns for 8823 contigs (see **Table 5**). Given the distribution of pigment in Rio Riata, genes essential for anthocyanin synthesis would be expected in the transcriptomes of all three organs, and they were found. Interestingly, while genes for some of the enzymes exhibited similar expression in all three tissues, and all tissues had all the essential components of the pathway for pigment synthesis, some copies (alleles) exhibited patterns consistent with tissue-specific expression (see **Table 4**). We verified that these results were not affected by reads mapping to multiple contigs, as using only uniquely mapping reads gave nearly identical results. This implies that the apparent diploid nature of dahlia floral pigmentation genes may reflect not only the loss of function of redundant gene-copies, but also evolved tissue-specific expression patterns for some of the redundant genes, a process termed subfunctionalization. As a second test of the classification approach, we investigated the tissue-specificity of contigs annotated as floral homeotic proteins. Of the 22 contigs annotated as similar to floral homeotic proteins, 12 were elevated in buds as expected, while the remainder were classified as lacking tissue specificity (see **Table 6**).

Table 4 | Genes putatively required for anthocyanidin synthesis.

Putative enzyme	Locus # / transcript #	Tissue-specificity[a]	Previously identified?
CHS	27/7	All	CHS2[b]
CHS	49578/1	Elevated in stem	CHS1[b]
CHS	7505/1	Decreased in leaf	
CHS	3129/1	All	
CHS	18623/1	Elevated in bud	
CHS	9318/2[c]	Elevated in bud	
CHI	977/3	All	CHI[b]
CHI	2352/8	All	
CHI	13838/4	All	
DvIVS	7738/1; 36303/2[d]	All	DvIVS[c]
FLS	4250/4	Elevated in stem	
F3′H	824/8	All	
F3′H	16120/4	All	
F3′H	18455/5	All	
DFR	44537/3	All	DFR[b]
DFR	23328/5	Decreased in bud	
DFR	11673/3	All	
DFR	27896/1	Decreased in bud	
ANS	45613/2	All	ANS[b]
ANS	27069/1	All	
F3H	22534/2	All	F3H[b]
F3H	33298/1	All	
F3H	39959/2	Decreased in bud	
F3H	43247/1	All	

[a] Expressed in all three samples at levels that did not differ enough to be designated as consistent with tissue-specific expression, designated as All. [b] Ohno et al., 2011b. [c] Ohno et al., 2011a. [d] These loci lacked either conserved start or stop codons.

DISCUSSION

Despite the large genome size of *D. variabilis* (9.62 pg; Temsch et al., 2008) and the expectation of two to four paralogs per locus type in an octoploid (Gatt et al., 1998), a paired-end read strategy using next generation sequencing yielded sufficient data to assemble more than 20,000 contigs that could be annotated with GO terms. As flowering plants sequenced to date contain ∼25,000–40,000 genes (e.g., Arabidopsis Genome Initiative, 2000; Matsumoto et al., 2005; Amborella Genome Project, 2013) it is likely that the transcriptome data define approximately half

Table 5 | Tentative tissue specificity of contigs.

Expression pattern	Number in this class
Decreased in buds	1407
Elevated in buds	390
Decreased in leaves	2059
Elevated in leaves	1409
Decreased in stems	2690
Elevated in stems	868

Table 6 | Putative floral homeotic genes in Dahlia.

Locus# / transcript#	Match to floral homeotic gene	Best blast hit accession number	*e*-Value	Read counts in bud	Read counts in leaf	Read counts in stem	Tissue-specificity[a]
312/8	PMADS2	Q07474	4.47e-71	5697	1	69	Elevated in bud
3477/2	APETALA1	Q41276	9.28e-92	2034	0	22	Elevated in bud
16671/2	AGAMOUS	Q03489	8.38e-119	1192	3	17	Elevated in bud
324/3	AGAMOUS	Q40872	3.30e-128	1163	7	14	Elevated in bud
5269/3	DEFICIENS	P23706	3.68e-96	895	0	18	Elevated in bud
5867/6	DEFICIENS	P23706	1.82e-91	729	2	18	Elevated in bud
3943/2	AGAMOUS	Q03489	9.65e-133	645	41	11	Elevated in bud
643/1	AGAMOUS	Q40872	5.90e-127	637	3	13	Elevated in bud
8582/1	AGAMOUS	Q40872	4.94e-122	245	0	5	Elevated in bud
4783/1	AGAMOUS	Q01540	1.72e-06	157	0	0	Elevated in bud
9368/1	PMADS1	Q07472	1.28e-69	67	6	1	Elevated in bud
9276/1	APETALA2	P47927	1.35e-36	33	0	3	Elevated in bud
7009/1	DEFICIENS	P23706	2.93e-94	1675	27	227	All
1095/3	AGAMOUS	Q40872	4.44e-137	1118	0	54	All
15403/1	APETALA2	P47927	8.53e-111	693	1197	616	All
22399/7	APETALA2	P47927	5.50e-90	350	557	323	All
17401/4	APETALA2	P47927	1.45e-85	214	240	502	All
23847/1	APETALA2	P47927	3.84e-88	140	8	61	All
40009/3	APETALA2	P47927	7.44e-101	135	105	347	All
15550/2	APETALA2	P47927	2.8e-76	98	19	40	All
31911/1	APETALA1	D7KWY6	1.06e-09	30	14	6	All
38896/1	APETALA2	P47927	3.57e-106	16	3	183	All

[a] *Expressed in all three samples at levels that did not differ enough to be designated as consistent with tissue-specific expression, designated as All.*

of dahlia loci. Genes expressed in seeds, roots, and non-abundant specialized cell types were likely missed as well as genes expressed during biotic or abiotic stress. Given the scanty knowledge about *Dahlia* genes, this new transcriptome resource can serve as a guide in gene identification and cloning, as a standard for comparison among varieties and species, and as a database for identifying genes unique to *Dahlia* and paralogs and orthologs shared with other composites, flowering plants, and all plants. Furthermore, we identified contigs whose expression patterns were consistent with tissue-specificity; future researchers may test our assignments by qPCR or further sequencing. Our method was limited by a lack of biological replicates, which prevents any estimate of the variability of expression of genes within a tissue-type. For this reason, we attempted to be conservative in our calling of specificity and have likely both failed to identify some contigs whose expression is highly enriched in certain tissues, as well as potentially misassigned contigs whose expression is not tissue-specific but is highly variable. However, we believe this resource will be of use for researchers attempting to clone cDNAs or perform *in situ* hybridization to localize specific transcripts. This report indicates that reference transcriptomes for plant species with complex genomes are a feasible and relatively inexpensive method for generating a toolkit for molecular and genetic analysis.

AUTHOR CONTRIBUTIONS

Erik M. Lehnert performed all of the experiments; Virginia Walbot maintained the dahlias; both authors contributed to design of data collection and analysis and manuscript writing and editing.

ACKNOWLEDGMENTS

We thank Tim Culbertson and Blaine Marchant for constructing the Stanford Dahlia Project website (http://web.stanford.edu/group/dahliagenetics/dahliareferencetranscriptome.htm) and for starting the dahlia collection. John Fernandes incorporated the transcriptome data into the website. This effort was supported by the Savitsky Fund and by the outreach component of a National Science Foundation grant (PGRP 07-01880) to Virginia Walbot. Erik M. Lehnert was supported by the Gordon and Betty Moore Foundation (grant #2629).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpls.2014.00340/abstract

REFERENCES

Amborella Genome Project. (2013). The *Amborella* genome and the evolution of flowering plants. *Science* 342, 1241089. doi: 10.1126/science.1241089

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692

Bate-Smith, E. C., Swain, T., and Nördstrom, C. G. (1955). Chemistry and inheritance of flower colour in the Dahlia. *Nature* 176, 1016–1018. doi: 10.1038/1761016a0

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Conesa, A., Götz, S., García-Gómez, J., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Gatt, M., Ding, H., Hammet, K., and Murray, B. (1998). Polyploidy and evolution in wild and cultivated *Dahlia* species. *Ann. Bot.* 81, 647–656. doi: 10.1006/anbo.1998.0614

Hodgins, K. A., Lai, Z., Oliveira, L. O., Still, D. W., Scascitelli, M., Barker, M. S., et al. (2014). Genomics of Compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Mol. Ecol. Resour.* 14, 166–1177. doi: 10.1111/1755-0998.12163

Lehnert, E. M., Mouchka, M. E., Burriesci, M. S., Gallo, N. D., Schwarz, J. A., and Pringle, J. R. (2014). Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. *G3* 4, 277–295. doi: 10.1534/g3.113.009084

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 1–8. doi: 10.1093/bioinformatics/btr507

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Matsumoto, T., Wu, J. Z., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., et al. (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/Nature03895

Ogata, J., Sakamoto, T., Yamaguchi, M., Kawanobu, S., and Yoshitama, K. (2001). Isolation and characterization of anthocyanin 5-O-glucosyltransferase from flowers of *Dahlia variabilis*. *J. Plant Physiol.* 158, 709–714. doi: 10.1078/0176-1617-00370

Ohno, S., Deguchi, A., Hosokawa, M., Tatsuzawa, F., and Doi, M. (2013). A basic helix-loop-helix transcription factor DvIVS determines flower color intensity in cyanic dahlia cultivars. *Planta* 238, 331–343. doi: 10.1007/s00425-013-1897-x

Ohno, S., Hosokawa, M., Hoshino, A., Kitamura, Y., Morita, Y., Park, K.-I., et al. (2011a). A bHLH transcription factor, DvIVS, is involved in regulation of anthocyanin synthesis in dahlia (*Dahlia variabilis*). *J. Exp. Bot.* 62, 5105–5116. doi: 10.1093/jxb/err216

Ohno, S., Hosokawa, M., Kojima, M., Kitamura, Y., Hoshino, A., Tatsuzawa, F., et al. (2011b). Simultaneous post-transcriptional gene silencing of two different chalcone synthase genes resulting in pure white flowers in the octoploid dahlia. *Planta* 234, 945–958. doi: 10.1007/s00425-011-1456-2

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucl. Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442

Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi: 10.1093/bioinformatics/bts094

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Yamaguchi, M., et al. (2002). cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme a: anthocyanidin 3-o-glucoside-6"-o-malonyltransferase from dahlia flowers. *Plant Physiol.* 130, 2142–2151. doi: 10.1104/pp.010447

Temsch, E. M., Greilhuber, J., and Hammett, K. R. W. (2008). Genome size in Dahlia Cav. *(Asteraceae–Coreopsideae). Plant Syst. Evol.* 276, 157–166. doi: 10.1007/s00606-008-0077-0

Yamaguchi, M. A., Oshida, N., Nakayama, M., Koshioka, M., Yamaguchi, Y., and Ino, I. (1999). Anthocyanidin 3-glucoside malonyltransferase from *Dahlia variabilis*. *Phytochemistry* 52, 15–18. doi: 10.1016/S0031-9422(99)00099-0

Yang, Y., and Smith, S. A. (2013). Optimizing *de novo* assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics* 14:328. doi: 10.1186/1471-2164-14-328

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 May 2014; paper pending published: 06 June 2014; accepted: 25 June 2014; published online: 17 July 2014.

Citation: Lehnert EM and Walbot V (2014) Sequencing and de novo assembly of a Dahlia hybrid cultivar transcriptome. Front. Plant Sci. ***5****:340. doi: 10.3389/fpls.2014.00340*

This article was submitted to Plant Genetics and Genomics, a section of the journal Frontiers in Plant Science.