Bangladesh J. Bot. **44**(5): 909-915, 2015 (December), Supplementary

DIFFERENTIALLY EXPRESSED GENES IN MALE AND FEMALE FLOWER BUDS OF HARDY KIWIFRUIT (*ACTINIDIA ARGUTA* (SIEB. *ET* ZUCC.) PLANCH. *EX* MIQ.)

XIAOYAN LI, HONGYAN QIN, ZHENXING WANG, QINGTIAN ZHANG, YINGXUE LIU, PEILEI XU, YING ZHAO, SHUTIAN FAN, YIMING YANG AND JUN AI*

Institute of Special Wild Economic Animal and Plant Science, China Academy of Agricultural Science, Changchun, Jilin, 130112, China

Key words: Actinidia arguta, Transcriptome, Illumina, Solexa, Sequencing, Sex determination

Abstract

The hardy kiwifruit (*Actinidia arguta* (Sieb. *et* Zucc.) Planch. *ex* Miq.) is a dioecious plant in which the sex type of seeding plants is difficult to determine morphologically before flowering. The available transcriptomic and genomic data are insufficient to understand the molecular mechanism of sex determination in *Actinidia arguta*. Therefore, the transcriptomes of male and female *Actinidia arguta* flower buds were determined. A total of 51,745 unigenes were obtained (N50 of 1,217 bp; average length of 694 bp). We performed Blast X alignments (E-value < $1.0e^{-5}$) against the NCBI NR, Swiss-Prot, GO, COG and KEGG data bases. 30,439 unigenes (58.8%) had matched to one or more of the data bases. About 1,625 differentially expressed genes (DEGs) between the two sex-type flower buds were identified, 27 of which might be involved in sex determination. This study furthered our understanding of the molecular mechanisms of plant sex determination and forms a resource for future functional genomics in *Actinidia arguta*.

Introduction

Actinidia arguta (Sieb. *et* Zucc.) Planch. *ex* Miq. is a member of Actinidiaceae, originating from northern China, Korea, Siberia and Japan, but is grown commercially in New Zealand, USA and many European countries (Ai 2014). *A. arguta* is a perennial, fast growing and deciduous twining vine, known as the hardy kiwifruit or baby kiwi in English. *A. arguta* fruit contains high levels of vitamin C, lutein and certain minerals, and extracts from the fruits, leaves and stem of *A. arguta* are used for the treatment of various inflammatory diseases and have the anti-cancer and antiallergic properties (Ravipati *et al*. 2012, Lee *et al*. 2014). Kiwifruit is known good for healthy and is preferred by consumers for its nutritional and medicinal value (Ferguson and Ferguson 2003). Increased demand has led its intensive cultivation, instead of collection from wild resources. Genetic improvement is required to increase its productivity and quality.

Indigenous varieties of *A. arguta* are dioecious, and the hermaphrodite type is extremely rare. To develop superior hardy kiwifruit, breeding projects have targeted both female and male cultivars. The success of a breeding program in dioecious plants depends upon early identification of the progeny's gender. Unfortunately, the gender of *A. arguta* seedlings cannot be determined morphologically until flowering, which usually occurs after 3 - 4 years in the field. This represents a serious problem for plant breeders, who are forced retain a large number of males for several years. The hardy kiwifruit genome is large and genomic information is lacking. As of August 2015, only 270 nucleotide sequences, 7,301 expressed sequence tags (ESTs) and 89 proteins from *A. arguta* had been deposited in the National Center for Biotechnology Information (NCBI) GenBank. The inherent limitations of EST sequencing and the small number of available ESTs suggested that present understanding of the *A. arguta* transcriptome is insufficient. Transcriptome sequencing has been proven to be an efficient method of gene discovery, especially using high-

*Author for correspondence. <aijun1005@163.com>.

throughput next generation sequencing technology. Studies have been conducted to detect novel candidate genes that may be linked to the trait of the sexual development, such as those in the Olive flounder (Fan *et al*. 2014), Seabuckthorn (Chawla *et al*. 2015) and Shrub willows (Liu *et al*. 2013).

In this study, we performed transcriptome sequencing for male and female flower buds of *A. arguta* using the Illumina HiSeq^{TM} 2500 technology to discover the differentially expressed genes (DEGs) that may be involved in sex determination of *A. arguta*. This study will provides a basis for identifying the sex determining genes and for reconstructing the regulatory network of sex determination for plants of the Actinidiaceae in future.

Materials and Methods

Male and female tree of *Actinidia arguta* were planted in the Institute of Special Wild Economic Animal and Plant Science (Zuojia, Jilin Province), China Academy of Agricultural Science. Flower buds were selected separately from the male and female trees in May, 2014. The expanded but unflushed flower buds 15 days were collected before the flowering and were kept in liquid nitrogen and stored at – 80°C for total RNA isolation.

Total RNA from male and female flower buds was isolated separately using the TRIzol® Reagent (Invitrogen). RNA quality was determined by the ratio of the 28S to 18S ribosomal RNA electropherogram peak using a Nanodrop 2000 spectrophotometer (Thermo Scientific) and a Bioanalyzer Chip RNA 7500 series II (Agilent). Two libraries were prepared using the mRNA-Seq 8 sample prep Kit (Illumina), according to the manufacturer's instructions, and sequencing was performed on the HiSeq2500 sequencing platform. The isolation of mRNA, fragment interruption, cDNA synthesis, addition of adapters, PCR amplification and RNA-Seq were performed by staff at Beijing BioMarker Technologies (Beijing, China).

Transcriptome *de novo* assembly was carried out using the Trinity software after the raw sequences were processed to remove low quality reads and adapters (Grabherr *et al*. 2011). Initially, Trinity was used with an optimized k-mer length of 25 for *de novo* assembly. Subsequently, the contigs were linked into transcripts according to the paired-end information of the sequences. The transcripts were clustered based on nucleotide sequence identity. The longest transcripts in the cluster units were regarded as unigenes to eliminate redundant sequences, and then the unigenes were combined to produce the final assembly used for annotation.

All unigene sequences were aligned by Blast X (cut off E-value < $1.0e^{-5}$) to the public non-redundant (NR), SwissProt, Kyoto encyclopedia of genes and genomes (KEGG) and clusters of orthologous groups (COG) protein data bases. The proteins with highest sequence similarity to the unigenes were used to assign functional annotations to the genes. Based on the NR annotation, present authors used the Blast2GO program (Conesa *et al*. 2005) to generate the gene ontology (GO) terms with an E-value $\leq 10^{-5}$, including molecular functions, biological processes and cellular components.

To screen the DEGs, a rigorous algorithm to identify differential transcript accumulation between two samples was developed based on the reads per kb per million reads (RPKM) method (Mortazavi *et al*. 2008). Equal expression between two samples can be calculated using the false discovery rate (FDR) method. The FDR was applied to determine the threshold of the p-value in multiple tests and analyses (Benjamini and Yekutieli 2001). An FDR < 0.001 and an absolute of $\log_2$ ratio value ≥ 1 were used as the threshold to judge the significance of the gene expression differences.

Results and Discussion

Illumina sequencing produced 19,244,670 clean paired-end reads with 3,887,029,780 nucleotides, and 20,103,690 clean paired-end reads with 4,060,504,304 nucleotides, from female and male flower buds, respectively. The raw reads are available in the Sequence Read Archive at the NCBI under accession number SRR2890157 and SRR2890158. Transcriptome *de novo* assembly resulted in 4,792,445 contigs and 108,325 scaffolds, with average lengths of 49 and 855 bp, respectively (Table 1). We then re-used the paired-end reads to fill the scaffold gaps to obtain unigenes with the few Ns (N was used to represent unknown sequences between two contigs and between the scaffolds that were built) that could not be extended on either end. Finally, we obtained 51,745 unigenes from both female and male flower buds, with a final unigene N50 (N50 represents median length of all unigenes) of 1,217 bp and a mean length of 695 bp, which included 11,946 unigenes (23.09%) with lengths greater than 1 kb (Table 1). These results showed that the throughput and sequencing quality was sufficient for subsequent analyses.

Table 1. Statistics of *A. arguta* expressed sequence tags (ESTs) generated by the Illumina sequencing platform.

Statistics	Contigs	Scaffolds	Unigenes
No. of sequences	4,792,445	108,325	51,745
Total length (bp)	235,292,998	92,669,299	35,937,366
N50 length (bp)	49	1,301	1,217
Mean length (bp)	49	855	695
No. of unigenes >1,000 (bp)	12,030	36,216	11,946

To annotate the transcriptome of *A. arguta,* we performed Blast X alignments (E-value < $1.0e^{-5}$) of the 51,745 unigenes against the NCBI NR, Swiss-Prot, GO, COG and KEGG data bases. We identified 30,439 unigenes (58.8% of all) that matched known proteins in the five public data bases, implying that the Illumina paired-end sequencing had produced a large proportion of the diverse genes expressed in *A. arguta* (Table 2). These unigenes were assigned gene or protein name descriptions and categorized with putative clusters of orthologous groups, gene ontology terms and metabolic pathways to provide functional information essential to understand the overall expression profiles of *A. arguta*. However, 42.2% of the unigenes did not show significant homology with any sequences in the NCBI public databases, probably because of the genome and EST information of *A. arguta* or closely related sequenced species is lacking.

Table 2. Functional annotation of *A. arguta* unigenes from matches in the public databases.

Annotated data base	No. of annotated sequences	Percentage of annotated sequences	≥ 300 nt	≥ 1000 nt
COG	8,618	16.7	7,748	5,131
GO	22,786	44.0	18,405	9,550
KEGG	6,640	12.8	5,396	3,071
Swiss-Prot	19,973	38.6	16,238	8,503
NR	30,311	58.6	24,067	11,746
All	30,439	58.8	24,129	11,749

Differences in tag frequencies in the female and male libraries were used to estimate differences in gene expression levels between the two libraries. 1,625 differentially expressed genes (DEGs) with at least a two-fold difference between the two libraries are shown in Fig.1. 768 (upregulated, highly expressed in male flower buds) and 857 (downregulated, highly expressed in female flower buds) represent transcripts with a more than two-fold higher or lower abundance, respectively, in the female compared with the male library.

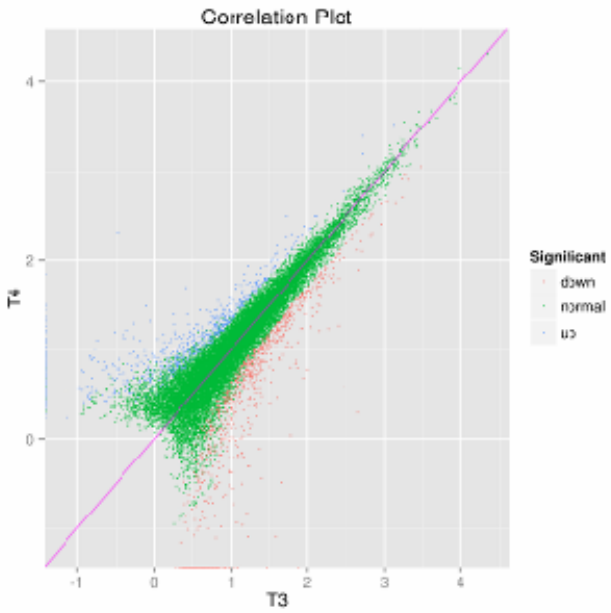

Fig. 1. Comparison of gene expression levels between the male and female libraries. To compare the gene expression levels between the two libraries, each library was normalized to 1 million tags. Red dots represent transcripts more prevalent in the female library (T3), blue dots show those present at a lower frequency in the male library (T4) and green dots indicate transcripts that did not change significantly. The parameters "false discovery rate (FDR) <0.001" and "$\log_2^{ratio} \geq 1$" were used as the threshold to judge the significance of gene expression differences.

Gene ontology (GO) functional analysis provides its functional classification annotation for DEGs as well as its functional enrichment analysis for DEGs. 22,786 unigenes (44.0% of all) were assigned at least one GO term, including 786 DEGs showing higher expression in male and female flower buds of *A. arguta*. In many cases, multiple terms were assigned to the same transcript; thus, 1,792 DEGs were assigned to one GO term in the cellular component category, 912 DEGs in the molecular function category and 2,118 DEGs in the biological process category (Fig. 2). Among biological processes category, cellular process (516 DEGs) and metabolic process (430 DEGs) were the most highly represented groups, indicating that the flower buds were undergoing rapid growth and extensive metabolic activities. We further identified GO terms in the biological process category that were over-represented in the lists of genes showing higher expression in male and female flower buds of *A. arguta*, respectively. It is worth noting that GO enrichment annotations revealed 29, 13 and 6 DEGs involved in pollen or pollen tube growth and development process, ovule or embryo development and stamen development process, these DEGs may have role in sex determination. It has been reported that a number of environment variables, such as light, temperature, water stress and disease, as well as exogenous treatment of

hormones or other growth-regulating substances, can directly influence plant sex expression. DEGs involved in other important biological processes such as stress response, signal transduction, and cell differentiation were also identified through GO terms. Plant hormones play an important role in the sexual organ development (Alvarez-Buylla *et al*. 2010). Many unigenes mapped to different hormone synthesis and metabolic pathways, including 6 DEGs in brassinosteroid biosynthetic process, 22 DEGs in auxin biosynthetic process, 11 DEGs in salicylic acid biosynthetic process, 7 DEGs in regulation of jasmonic acid biosynthetic process, 14 DEGs in regulation of ethylene biosynthetic process, 9 DEGs in regulation of gibberellin biosynthetic process and 5 DEGs in cytokinin metabolic process.

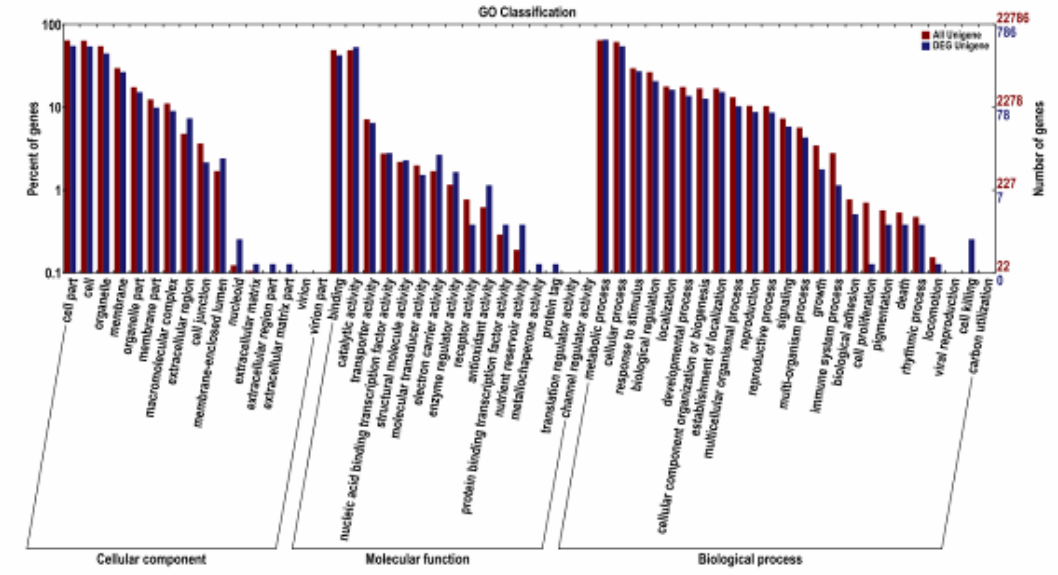

Fig. 2. Gene ontology (GO) classification of annotated unigenes. The results are summarized under the three main GO categories: biological process, cellular component and molecular function. The right y-axis indicates the number of genes in each category. The left y-axis indicates the percentage of a specific category of genes in the corresponding GO category.

Forty four putative genes that known to be involved in floral regulatory pathways and could be potential candidates for sex determination in *Arabidopsis* were identified following Chawla *et al*. 2015. Among them, 27 DEGs (16 highly expressed in the female and 11 highly expressed in the male) from *A. arguta* transcriptome resources were mapped to 15 known sex determination genes (Table 3), which may play an important role in sex determination. *WUSCHEL* (*WUS*) promotes central identity in both indeterminate shoot and determinate floral meristems, and plays an important role in maintaining their structural and functional integrity (Groß-Hardt *et al*. 2002). *MADS-box* genes (the ABCDE model genes) promote sexual organ differentiation and development (Quinet *et al*. 2014). *CONSTANS* (*CO*) and *CRYPTOCHROME* regulate flowering time in response to day length and blue ultraviolet (Samach *et al*. 2000). *LEAFY* (*LFY*) /*FLORICAULA* (*FLO*) promotes the expression of meristem identity *AP1*, and activates the floral organ identity genes, such as *AP3* and *AGM* (Reeves and Coupland 2000). *CRABS CLAW* and *SUPERMAN* (*SUP*) control cell proliferation in stamen, carpel primordia and ovules in flower development (Sakai *et al*. 2000). These potential candidate genes for sex determination in *A. arguta* were of great interest for further investigation.

Table 3. Potential candidate differentially expressed genes for sex determination in *A. arguta*.

No.	Gene name	Unigene ID	Annotation	High expression
1	WUSCHEL (WUS)	c59106.graph_c0	WUSCHEL-related homeobox 9-like	Female
		c51866.graph_c0	Protein WUSCHEL	Male
2	MADS-box	c55388.graph_c0	Agamous-like MADS-box protein AGL8	Male
		c55210.graph_c	MADS-box transcription factor 6	Male
		c45764.graph_c0	MADS-box transcription factor ANR1	Female
3	CLAVATA	c55980.graph_c1	PREDICTED: CLAVATA3/ESR (CLE)-related protein TDIF-like	Male
4	CONSTANS (CO)	c59864.graph_c0	Zinc finger protein CONSTANS-LIKE 2-like	Male
		c51139.graph_c0		Female
5	CRYPTOCHROME	c61648.graph_c1	Cryptochrome DASH	Male
6	LFY/FLO	c59727.graph_c0	LEAFY/FLORICAULA-like protein	Male
7	CRABS CLAW	c54633.graph_c0	Protein CRABS CLAW	Female
8	SUPERMAN (SUP)	c55116.graph_c0	Transcriptional regulator SUPERMAN	Female
		c54668.graph_c0	Transcriptional regulator SUPERMAN-like	,,
9	SOC1	c51821.graph_c0	MADS-box protein SOC1	Female
10	FLC	c65978.graph_c0	Flowering locus 1	Male
11	FPF1	c45720.graph_c0	Flowering-promoting factor 1-like protein	Female
		c46450.graph_c0		,,
		c27792.graph_c0		,,
12	DROOPING LEAF (DL)	c54633.graph_c0	PREDICTED: protein DROOPING LEAF-like	Female
13	Ethylene-response factor (ERF)	c56176.graph_c0	Ethylene response factor	Female
		c45598.graph_c0	Ethylene-responsive transcription factor ERF113	Male
		c58654.graph_c0	Ethylene-responsive transcription factor CRF2	Male
14	ERECTA (ER)	c64256.graph_c0	LRR receptor-like serine/threonine-protein kinase ERL	Male
15	Oleosin 16 (OLE16)	c56116.graph_c0	PREDICTED: oleosin 16 kDa	Female
		c45590.graph_c0	Oleosin 16 kDa	,,
		c39356.graph_c0	Oleosin 16 kDa	,,
		c68849.graph_c0	OleI	,,

Here, the present authors surveyed the transcriptome of *A. arguta* by using Illumina sequencing technology, and produced 51,745 assembled unigenes and identified 1,625 genes that were differentially expressed in the male and female flower buds of *A. arguta*. Notably, they detected 27 DEGs involved in floral regulatory pathways that could be potential candidates for sex determination in *A. arguta*. Their findings also provided novel insights into the molecular

mechanisms of *A. arguta* sex determination, as well as a rich list of candidate genes for further functional analysis. The transcriptome dataset is a substantial contribution to the existing genomic resources for *A. arguta* and will serve as an important public information platform to accelerate research into gene expression, genomics and functional genomics in *A. arguta*.

Acknowledgments

This work was supported by grants from the project of Jilin Provincial Science & Technology Department (Grant No. 20130102046JC and Grant No. 20140204034NY).

References

Ai J 2014. Cultivation and processing technologies of *Actinidia arguta* Beijing: China Agriculture Press. (in Chinese)

Alvarez-Buylla ER, Benítez M and Corvera-Poiré A 2010. Flower Development. The Arabidopsis Book **8**: e0127.

Benjamini Y and Yekutieli D 2001. The control of the false discovery rate in multiple testing under dependency. Ann Stat. **29**: 1165-1188.

Chawla A, Stobdan T, Srivastava RB, Jaiswal V, Chauhan RS and Kant A 2015. Sex-biased temporal gene expression in male and female floral buds of seabuckthorn (*Hippophae rhamnoides*). PLoS One **10(**4): e0124890.

Conesa A, Götz S, García-Gómez JM, Terol J and Talón M 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**: 3674-3676.

Fan Z, You F, Wang L, Weng S, Wu Z, Hu J, Zou Y, Tan X and Zhang P 2014. Gonadal transcriptome analysis of male and female olive flounder (*Paralichthys olivaceus*). Bio. Med. Res. Int. **2014**: 291067.

Ferguson AR and Ferguson LR 2003. Are kiwifruits really good for you? Acta Hortic. **610**: 132-138.

Grabherr MG, Haas BJ, Yassour M, Levin JZ and Thompson DA 2011. Full length transcriptome assembly from RNA Seq data without a reference genome. Nat. Biotechnol. **29** (7): 644-652.

Groß-Hardt R, Lenhard M and Laux T 2002. *WUSCHEL* signaling functions in interregional communication during *Arabidopsis* ovule development. Genes Dev. **16**(9):1129-1138.

Lee J, Sowndhararajan K, Kim M, Kim J, Kim D, Kim S, Kim GY, Kim S and Jhoo JW 2014. Antioxidant, inhibition of α-glucosidase and suppression of nitric oxide production in LPS-induced murine macrophages by different fractions of Actinidia arguta stem. Saudi J. Biol. Sci. **21**: 532-538.

Liu J, Yin T, Ye N, Chen Y, Yin T, Liu M and Hassani D 2013. Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). PloS one **8**(4): e60181.

Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods **5**: 621-628.

Niharika Sharma1, Chol-Hee Jung, Prem L. Bhalla1, Mohan B. Singh 2014. RNA sequencing analysis of the gametophyte transcriptome from the liverwort, *Marchantia polymorpha*. PloS one **9**(5): e97497.

Quinet M, Bataille G, Dobrev PI, Capel C and Lozano R 2014. Transcriptional and hormonal regulation of petal and stamen development by *STAMENLESS,* the tomato (*Solanum lycopersicum* L.) orthologue to the B-class *APETALA3* gene. J. Expt. Bot. **65**(9): 2243-2256.

Ravipati AS, Zhang L, Koyyalamudi SR, Jeong SC, Reddy N, Bartlett J, Smith PT, Shanmugam K, Munch G, Wu MJ, Satyanarayanan M and Vysetti B 2012. Antioxidant and anti-inflammatory activities of selected Chinese medicinal plants and their relation with antioxidant content. BMC Complement. Altern. Med. 1**2**, 173: 1-14.

Reeves PH and Coupland G 2000. Response of plant development to environment: control of flowering by day length and temperature. Curr. Opin. Plant Biol. **3**(1): 37-42.

Sakai H, Krizek BA, Jacobsen SE and Meyerowitz EM 2000. Regulation of *SUP* expression identifies multiple regulators involved in *Arabidopsis* floral meristem development. Plant Cell. **12**(9):1607-1618.

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z and Yanofsky MF 2000. Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis.* Science **288**(5471):1613-1616.

(Manuscript received on 11 September, 2015; revised on 2 November, 2015)