ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2014) **12**(S1); S130–S133
doi:10.1017/S1479262114000410

De novo transcriptome assembly of *Vitis flexuosa* grapevines inoculated with *Elsinoe ampelina*

Soon Young Ahn[1], Seon Ae Kim[1], Sung Hwan Jo[2] and Hae Keun Yun[1]*

[1]*Department of Horticulture and Life Science, Yeungnam University, Gyeongsan 712-749, Republic of Korea and* [2]*SEEDERS Inc., Daejeon 302-735, Republic of Korea*

Abstract

In this study, the transcriptome of *Vitis flexuosa* leaves inoculated with *Elsinoe ampelina* was analysed to identify useful genes and elucidate their function and differential expression patterns through assembly and annotation gene ontology of data from sequencing short reads on the Illumina platform. We assembled ~121 million high-quality trimmed reads using Velvet and Oases with optimal parameters into a non-redundant set of 70,899 transcripts (≥200 bp in length). The transcripts exhibited an average length of 1138 bp and a N50 length of 1695 bp, with the largest contig length being 9623 bp. Functional categorization revealed the conservation of genes involved in various molecular functions, including protein binding (21.1%) and oxidoreductase activity (11.7%), in *V. flexuosa*. The *V. flexuosa* transcript set generated in this study will serve as a resource for gene discovery and development of functional molecular markers.

Keywords: DEGs; short reads; transcripts; *Vitis*

Introduction

Grapevines (*Vitis* spp.) are the most economically important fruit crops worldwide (FAOSTAT, 2011); they are exposed to a number of pathogens, including *Elsinoe ampelina* (Pearson and Goheen, 1998). The grapevine genome is highly polymorphic with a moderate genome size of about 475 Mb (Lodhi *et al.*, 1995; Lijavetzky *et al.*, 2007).

Transcriptome studies have been conducted to investigate the function and expression pattern of a number of genes that accumulated in various tissues of model species (Fung *et al.*, 2008; Polesani *et al.*, 2010). In a molecular breeding programme of grapevines, transcriptome data could serve as useful genomic resources of valuable information for future studies. Because transcript sequences generally lack the repetitive sequences that make genome assembly complicated (Venturini *et al.*, 2013), *de novo* transcriptome assembly has also been employed as a supplement to provide useful information regarding the whole-genome sequence of grapevines.

In this study, we conducted *de novo* assembly of the transcriptome of *Vitis flexuosa* leaves inoculated with *E. ampelina*, a pathogen of grapevine anthracnose, to obtain useful gene resources and detailed genomic information for grape breeding.

Materials and methods

V. flexuosa VISKO001, a type of wild grape from Korea, was maintained in a grapevine germplasm collection field of Yeungnam University, Gyeongsan, and cultivated in an experimental glasshouse at 25°C under a 16 h photoperiod. For pathogen inoculation, when plants reached the 15- and 20-leaf stage, fully expanded leaves were sprayed with an *E. ampelina* spore suspension (10^5 spores/ml). Infected leaves were collected 24 h after inoculation, immediately frozen in liquid nitrogen

* Corresponding author. E-mail: haekeun@ynu.ac.kr

and then stored at −80°C. Controls were harvested from water-treated leaves under the same conditions. RNA was extracted using the modified pine tree method (Chang *et al.*, 1993). The quality of total RNA was determined based on the absorbance at 230, 260 and 280 nm, which was measured using a NanoDrop spectrophotometer (ND-1000; NanoDrop Technologies Inc., Wilmington, Delaware, USA). After carrying out a quality check and trimming (SolexaQA package) the raw data, reference genes were mapped using Bowtie (version 0.12.7). A flowchart of the transcriptome analysis is shown in Fig. 1. Libraries were prepared using the TruSeq SBS v5 protocol, and more than 121 sequencing paired-end reads were generated using Illumina HiSequation 2000 (Illumina, San Diego, California, USA). The sequences were deposited in the National Agricultural Biotechnology Information Center (NABIC, Rural Development Administration, Korea) database (NN-0682 and NN-0683). *De novo* assembly of trimmed reads was carried out using Velvet (version 1.2.07) and Oases (version 0.2.08), with *k*-mer sizes of 61, 65, 67 and 69, a minimum contig length of 200 and an insert length of 300. We then reassembled with *k*-mer sizes of 69 using the transcripts of $k = 67$ and $k = 69$. The longest transcripts from $k = 69$ in the same locus were selected to assemble them based on putative amino acids obtained from nucleotide sequences. Gene expression was measured using a short read number, and differentially expressed genes (DEGs) were identified, clustered and annotated. DEGs were identified using the DESeq Bioconductor package (version 1.6.0) (Anders and Huber, 2010) with a false discovery rate (FDR) of the Benjamini–Hochberg multiple tests of 1% ($P < 0.01$) based on read counts and a minimum fold change of 2 in at least one pairwise comparison. Pearson's correlation hierarchical clustering was used to analyse gene expression patterns. Expressed genes of *V. flexuosa* inoculated with the pathogen against *Vitis vinifera*/*Arabidopsis* were annotated using the BLASTN and BLASTX search, respectively. Annotation information was connected to gene ontology (GO) and the biological pathway (Kyoto Encyclopedia of Genes and Genomes).

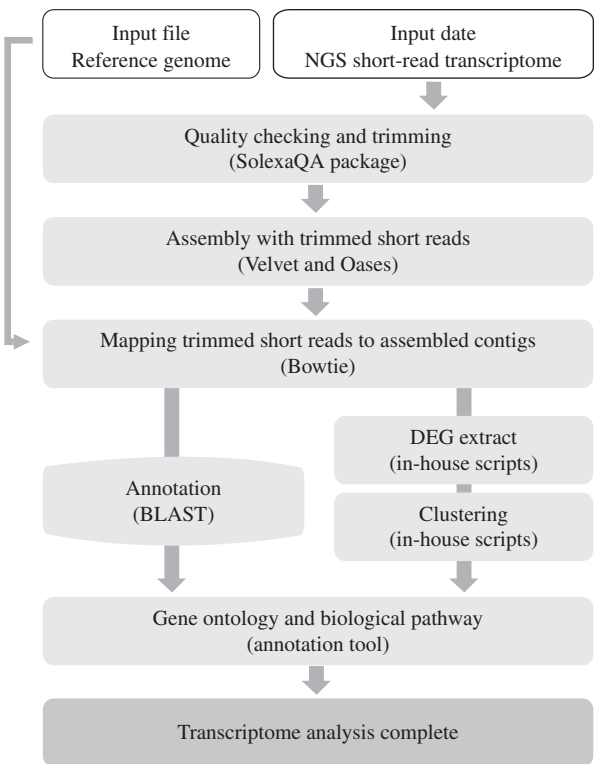

Fig. 1. Flowchart of transcriptome analysis by the RNA-Seq method.

Results and discussion

We obtained the assembly results with a total number of 70,899 contigs (≥200 bp in length) with a N50 length of 1695 bp, a maximum contig length of 9623 bp and an average contig length of 1138 bp from the *E. ampelina*-inoculated leaves of *V. flexuosa*. We assessed the effects of sequence quality, various assembly parameters and assembly programs on the final assembly output.

The expression profiles of DEGs from the *E. ampelina*-inoculated grapevines were determined based on their expression modulation. We identified 10,064 loci that were expressed in at least one of the samples using the combination of a FDR ≤ 0.01% and a log2 ratio ≥2. Among the DEGs, 4675 genes were up-regulated and 5389 genes were down-regulated in the leaves 24 h after *E. ampelina* inoculation compared with the control. BLASTN analysis of *V. flexuosa* transcripts revealed 8399 genes to be homologous to *V. vinifera* (*E* value ≤1e-80), and BLASTX analysis of *V. flexuosa* transcripts revealed 7973 genes to be homologous to *Arabidopsis* (*E* value ≤1e-06).

The annotated function of DEGs from the *E. ampelina*-inoculated *V. flexuosa* grapevines was assigned by mapping unigenes onto the GO structure using the TGICL (TIGR Gene Indices Clustering Tools). Unigenes with the assigned putative functions were classified into three ontologies, molecular function, biological process and cellular component, by controlled GO term.

DEGs that were up-regulated in response to *E. ampelina* infection were divided into 9, 15 and 17 main functional categories of molecular function, biological process and cellular component with different degrees at level 2, respectively (Fig. 2). In the GO terms of molecular function, protein binding (21.1%) was the largest category induced by pathogen inoculation,

https://doi.org/10.1017/S1479262114000410 Published online by Cambridge University Press

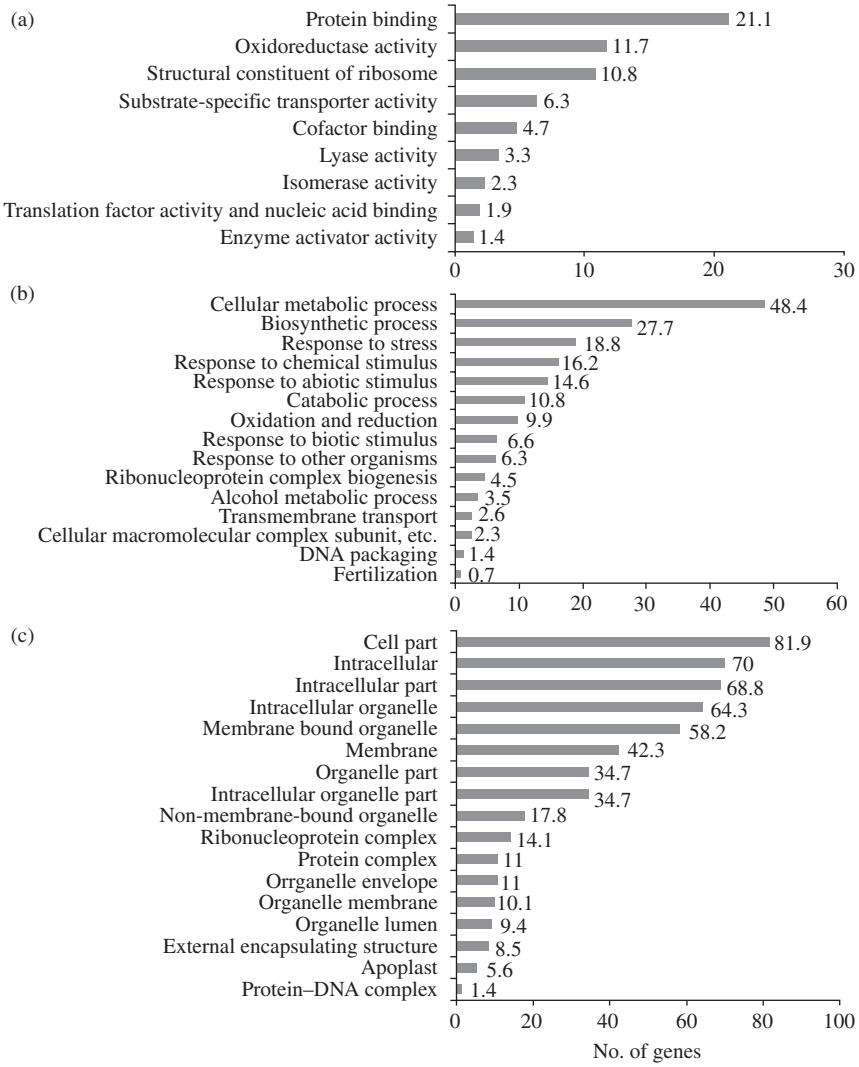

Fig. 2. Functional categories of highly up-regulated differentially expressed genes with putative functions assigned through gene ontology annotation at level 2 from *Elsinoe ampelina*-inoculated *Vitis flexuosa* grapevines. (a) Molecular function, (b) biological process and (c) cellular component. Assignments were based on the data available in TIGR Gene Indices Clustering Tools.

followed by oxidoreductase activity (11.7%) and structural constituent of ribosome (10.8%) (Fig. 2(a)). The largest proportions of biological process were annotated as cellular metabolic process (48.4%) and biosynthetic process (27.7%), meaning comprehensive changes in grapevine gene expression. Response to stress (18.8%), response to biotic stimulus (6.6%) and response to other organisms (6.3%) were other major assignments under the biological process categories, which included responses to stress, suggesting that a significant proportion of DEGs were produced in leaves infected with *E. ampelina* in grapevines (Fig. 2(b)). Cell and intracellular parts were assigned into the largest categories in cellular component (Fig. 2(c)).

DEGs that were down-regulated in response to *E. ampelina* infection were assigned to two main molecular function categories, ten main biological process categories and five main cellular component categories at level 2 (data not shown). Among the biological process categories, biosynthetic process (29.4%) and response to stress (27.1%) comprised the two most abundant clusters. When compared with the up-regulated DEGs, the down-regulated genes exhibited simple clusters in the biological process and cellular component categories by controlled GO term.

De novo assembly is a powerful tool for the reconstruction of transcriptomes of non-model species lacking reference (Venturini *et al.*, 2013) and for the selection of valuable genes and genetic information in grape breeding programmes.

In this study, we demonstrated a strategy for *de novo* assembly of a transcriptome using short reads for

https://doi.org/10.1017/S1479262114000410 Published online by Cambridge University Press

V. flexuosa, which is resistant to anthracnose disease. *De novo* assembly of transcriptomes using short read sequences has received a great deal of attention (Gibbons *et al*., 2009) and been applied to analyses of a variety of organisms (Jackson *et al*., 2009; Garg *et al*., 2011). We produced more than 121 million sequence reads for *V. flexuosa* and revealed a non-redundant set of 70,899 transcripts representing about 80 Mb of a unique transcriptome sequence. Detailed analysis of the transcript dataset has led to the analysis of important features such as genes conserved across grapevines and *Arabidopsis* and assignment of functional categories.

The *V. flexuosa* transcript set generated in this study will serve as a resource for gene discovery and development of functional molecular markers related to disease resistance response in grapevine breeding programmes.

Acknowledgements

This study was supported by a grant from the Next-Generation BioGreen 21 program (no. PJ008213), Rural Development Administration, Republic of Korea.

References

Anders S and Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology* 11: R106.

Chang S, Puryear J and Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* 11: 113–116.

FAOSTAT (2011) Food and agricultural commodities production. Available at http://faostat.fao.org

Fung RW, Gonzalo M, Fekete C, Kovacs LG, He Y, Marsh E, McIntyre LM, Schachtman DP and Qiu W (2008) Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. *Plant Physiology* 146: 236–249.

Garg R, Patel RK, Tyagi AK and Jain M (2011) *De Novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research* 18: 53–63.

Gibbons JG, Janson EM, Hittinger CT, Johnston M, Abbot P and Rokas A (2009) Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Molecular Biology and Evolution* 26: 2731–2744.

Jackson BG, Schnable PS and Aluru S (2009) Parallel short sequence assembly of transcriptomes. *BMC Bioinformatics* 10(Suppl. 1): S14.

Lijavetzky D, Cabezas JA, Ibáñez A, Rodríguez V and Martínez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genome* 8: 424.

Lodhi MA, Daly MJ, Ye GN, Weeden NF and Reisch BI (1995) A molecular marker based linkage map of *Vitis*. *Genome* 38: 786–794.

Pearson RC and Goheen AC (1998) *Compendium of Grape Disease*. Saint Paul, Minnesota: APS Press.

Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, Zadra C, Lovato A, Pezzotti M, Delledonne M and Polverari A (2010) General and species-specific transcriptional responses to downy mildew infection in a susceptible (*Vitis vinifera*) and a resistant (*V. riparia*) grapevine species. *BMC Genomics* 11: 117.

Venturini L, Ferrarini A, Zenoni S, Tornielli GB, Fasoli M, Santo SD, Minio A, Buson G, Tononi P, Zago ED, Zamperin G, Bellin D, Pezzotti M and Delledonne M (2013) *De novo* transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity. *BMC Genomics* 14: 41.