Global Illumina Sequencing and the Development of EST-SSR Markers in Alfalfa

Zhipeng Liu
Lanzhou University, China

Yanrong Wang
Lanzhou University, China

The 22nd International Grassland Congress (Revitalising Grasslands to Sustain Our Communities) took place in Sydney, Australia from September 15 through September 19, 2013.

Proceedings Editors: David L. Michalk, Geoffrey D. Millar, Warwick B. Badgery, and Kim M. Broadfoot

Publisher: New South Wales Department of Primary Industry, Kite St., Orange New South Wales, Australia

Global Illumina sequencing and the development of EST-SSR markers in alfalfa

Zhipeng Liu and Yanrong Wang

State Key Laboratory of Grassland Agro-ecosystems, School of Pastoral Agricultural Science and Technology, Lanzhou University, Lanzhou 730020, People's Republic of China, caoye.lzu.edu.cn.

Contact email: lzp@lzu.edu.cn

Keywords: alfalfa, Illumina sequencing.

Introduction

RNA-Seq, a massively parallel sequencing method for transcriptome analysis, only analyzes transcribed portions of the genome. Recently, RNA-Seq has provided an opportunity to expand the identification of alfalfa (*Medicago sativa*) genes. Using Illumina sequencing, 124,025 unique sequences from MSGI 1.0 have been identified from the elongating stem and post-elongation stem internodes of two alfalfa genotypes (Yang *et al*. 2011). Using 454 sequencing, 54,216 unique sequences were obtained from the roots and shoots of two alfalfa genotypes (Han *et al*., 2011). In addition, Illumina sequencing of old and young stems of 27 alfalfa genotypes led to the identification of 25,183 contigs (Li *et al*. 2012). While these experiments have identified numerous transcripts, the transcripts were derived only from stems, roots, and shoots. Therefore, further transcriptome sequencing of a broader array of tissues permit the global identification of transcripts that would be useful in modern alfalfa breeding programs.

Method

Tissue Material

The alfalfa cultivar "Golden queen" was grown in a greenhouse. A total of 15 tissue types were collected, including germinated seeds (36 hours after seed germination), germinated seeds (48 hours after seed germination), cotyledons (from a 7-day-old seedling), unifoliate leaves (from a 20-day-old seedling), roots (from a 20-day-old seedling), compound leaves, young stems (less lignified), middle stems (moderately lignified), old stems (highly lignified), shoot apex, young inflorescences (diameter 0.4-0.5 cm), mature inflorescences (diameter 2 cm), young pods (16 days after pollination), and mature pods (24 days after pollination) and callus cells (Fig. 1).

Development of EST-SSR markers

The 40,433 unigenes of alfalfa obtained in the present study were subjected to SSRs detection using the Simple Sequence Repeat Identification Tool program (SSRIT).

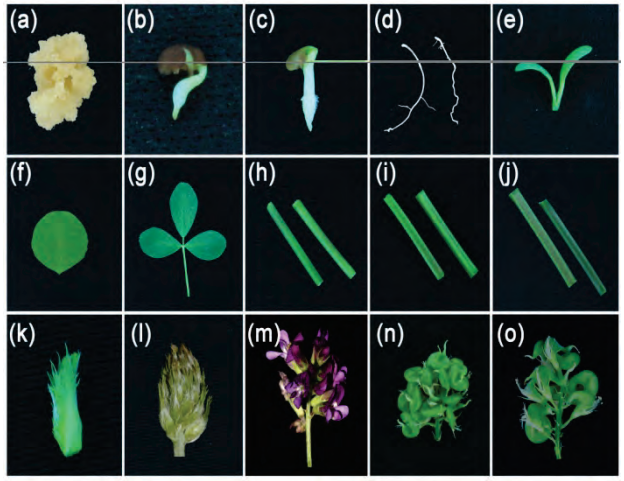

Figure 1. The tissues used in this study and all samples shown are from the alfalfa cultivar "Golden queen."

Discussion

Illumina sequencing and de novo assembly

All high-quality reads were assembled *de novo* using the Trinity program, which produced 40,433 unigenes that were obtained with an N50 length of 1,300 bp and a mean length of 803 bp. Altogether, 36,684 (90.73%) unigenes were successfully annotated in the Nr, Nt, Swiss-Prot, KEGG, COG, Ipr, and TrEMBL databases, suggesting that they have relatively well-conserved functions. To assess the extent of transcript coverage provided by the unigenes, we plotted the ratio of assembled unigene length to *M. truncatula* ortholog length. Among the 64,127 (Mt3.5.2) transcripts, 41,447 (64.63%) *M. truncatula* transcripts had homologous transcripts in the *M. sativa* genome. This finding suggests that most of the *M. truncatula* ortholog coding sequences could be covered by at least one individual unigene.

In addition, the available *M. truncatula* genome sequence was used as a scaffold to align the alfalfa unigene sequences. Under stringent conditions using Blat, including a threshold of 95% identity and 90% coverage, 27,853 (68.89%) unigenes were mapped to the

Table 1. Summary of the unigenes and their location on the Mt3.5.2 chromosomes.

Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr0	Total
3,128	3,069	3,841	4,047	4,348	1,491	3,363	2,969	1,597	27,853

Mt3.5.2 genome sequence assembly and their likely map positions inferred (Table 1).

SSR discovery

Using the SSRIT tool, a total of 1,649 potential EST-SSRs were identified from 1,494 unigenes. Of the 100 primer pairs, 82 were able to amplify PCR products from alfalfa genomic DNA, while 18 primer pairs failed to amplify PCR products. Of the 82 successful primer pairs, 37 PCR products were of the expected sizes, and 34 primer pairs generated PCR products that were larger than expected, suggesting that the amplified regions were likely to contain introns. The PCR products of the other 11 primer pairs were smaller than expected, suggesting a lack of specificity, assembly errors or deletions within the genomic sequences.

Conclusions

This work presents a *de novo* transcriptome sequencing analysis of mixed RNAs from 15 different tissues. A total of 5.64 Gb of data were generated and assembled into 40,433 unigenes. The 1,649 potential EST-SSRs predicted in this study provide a solid foundation for molecular marker development in alfalfa.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (31072072 and 31272492)

References

Han YH, Kang Y, Torres-Jerez I, Cheung F, Town CD, *et al*. (2011) Genome-wide SNP discovery in tetraploid alfalfa using 454 sequencing and high resolution melting analysis. *BMC Genomics* **12**, 350.

Li XH, Acharya A, Farmer AD, Crow JA, Bharti AK, *et al*. (2012) Prevalence of single nucleotide polymorphism among 27 diverse alfalfa genotypes as assessed by transcriptome sequencing. *BMC Genomics* **13**, 568.

Yang SS, Tu ZJ, Cheung F, Xu WW, Lamb JFS, *et al*. (2011) Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. *BMC Genomics* **12**, 199.