Establishment of a Tea Custom Array Based on EST Data Obtained with Next-generation Sequencing

Yasuo Niwa[1], Shoko Watanabe[1], Takeshi Ogawa[1], Hideki Nagasaki[2], Yasukazu Nakamura[2], Hideki Nishimura[3], Takashi Hirayama[3], Eiji Kobayashi[4], Yoriyuki Nakamura[1], and Hirokazu Kobayashi[1]

1: Univ. Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka, 422-8526, Japan
2: Natl. Inst. Genetics, 1111 Yata, Mishima, Shizuoka, 411-0801, Japan
3: IPSR, Okayama Univ., 2-20-1, Chuo, Kurashiki, Okayama, 710-0046, Japan
4: Shizuoka Tea Res. Ctr., 1706-11 Kurasawa, Kikugawa, Shizuoka, 439-0002, Japan
Correspondent email: niwa@u-shizuoka-ken.ac.jp

Summary

Light is one of the key factors controlling the development and differentiation of plants. To elucidate the mechanisms of etiolation and the increase in amino acid content in tea leaves due to complete shading, comprehensive gene expression analysis is essential. So far, however, the sequence information of tea genes, especially from the Japanese major variety Yabukita, has been very limited. Therefore, in this study we carried out an expressed sequence tag (EST) analysis of Yabukita using next-generation sequencing. The results provided EST data for more than 50,000 contigs. A custom array was created by designing probes consisting of a 60-mer oligonucleotide for each gene. Highly reproducible data was obtained by microarray analysis.

Introduction

Tea is the most widely consumed beverage in the world. Since tea is not a model plant and it has a very large genome size of about 4 Gb, the sequence information of tea genes, especially from the Japanese major variety Yabukita, has been very limited. Therefore, in this study we carried out an expressed sequence tag (EST) analysis of Yabukita using next-generation sequencing.

Materials and methods

Total RNA was extracted by RNeasy Mini Kit (QIAGEN) from five different stages of the tea leaf (*Camellia sinensis* cv. Yabukita) according to manufacturer's instructions. After DNase (Wako) treatment, cDNA was synthesized using a Transcriptor First Strand cDNA Synthesis Kit (Roche). To normalize the amplified cDNA, a TRIMMER DIRECT kit (Evrogen) was used. The cDNA library was sequenced from 5' end using a Genome Sequencer FLX+ System (Roche Diagnostics). Image processing was performed by 454 BaseCaller version 2.6, and MIRA version 3.4.0 (Chevreux, *et al.* (2004)) was used for assembly. The assembled contigs and singlets were compared to the NT and NR databases using blastn and blastx, with a significant threshold set at an E-value smaller than 10.0. The custom micro array for Yabukita was designed by eArray system (Agilent). The custom

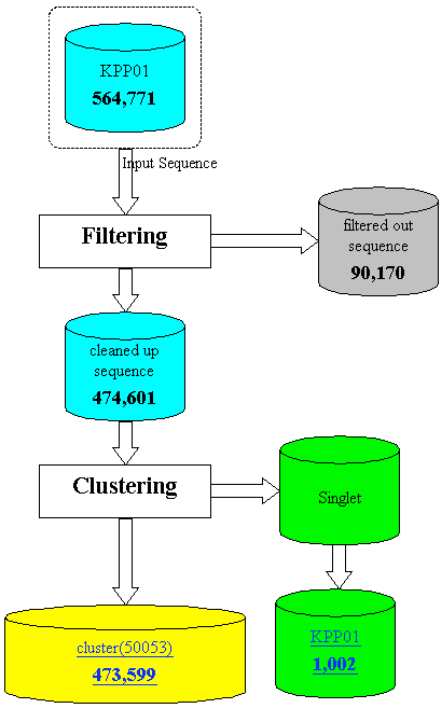

Figure 1. Overview of sequencing and assembly. 5,0053 contigs were assembled from 473,599 reads. 1,002 reads remained as singlets.

array was printed in 8 x 60K format.

Results and discussion

A total of 564,771 reads were obtained from the Genome Sequencer FLX+ System. An overview of the sequencing and assembly is outlined in Figure 1. After removal of adaptor sequences and invalid reads, 474,601 cleaned up sequences remained. The clean sequences were assembled using the MIRA program, producing 50,053 contigs with an average contig length of 626 bp. This length is three times longer than the contigs obtained by Shi *et al.* (2011). This is because the average read length of our data was six times longer than theirs. 1,002 reads remained as singlets with an average length of 241 bp.

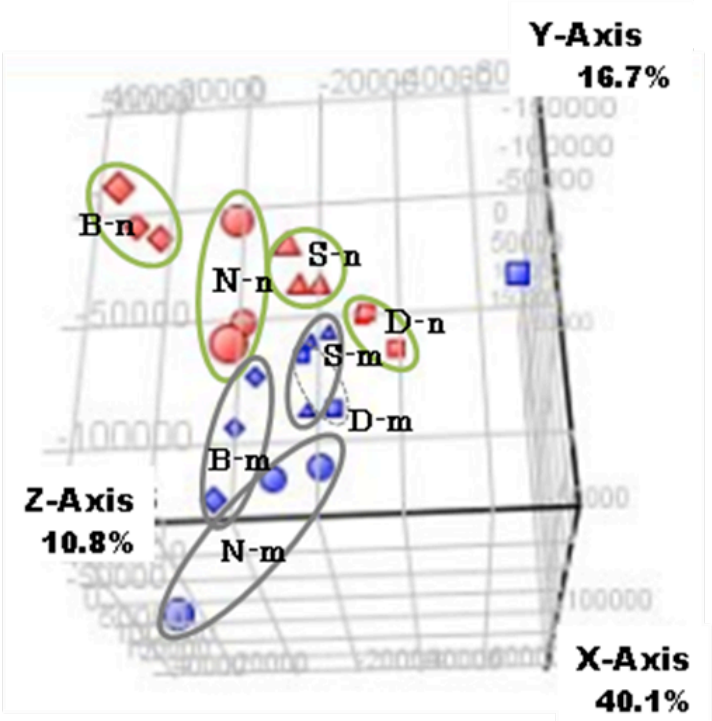

Figure 2. Principal component analysis of transcriptome data obtained from eight different conditions with three biological repeats:
-n, new leaf; -m, mature leaf;
B, before light treatment (diamond); N, normal light (circle); S, shaded (triangle); D, dark (square)

To elucidate the mechanisms of etiolation and the increase in amino acid content in tea leaves due to complete shading, a custom array was created by designing probes consisting of one 60-mer oligonucleotide for each gene. Highly reproducible data was obtained by microarray analysis using RNAs isolated from eight different lighting conditions and developmental stages (see Figure 2).

Arginine is one of the amino acids that accumulates in juvenile leaf grown in the dark. Arginase is a manganese-containing enzyme which converts L-arginine into L-ornithine and urea. The expression level of arginase gene grown under dark conditions was significantly reduced compared that under light conditions. Arginase may be a key enzyme that accounts for the high arginine content in tea leaves grown in the dark.

Acknowledgements

This study was supported by The Local Collective Research Development Program “From Shizuoka to the World: Research and Development of Next-genaration Bottled Tea Drinks and Tea Extracts”.

References

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14:1147-59

Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Chen Q, Xia T, Wan XC. (2011) Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics. 12:131