Transcriptome analysis of *Taraxacum kok-saghyz* using RNA-Seq and identification of candidate SNP markers related to rubber biosynthesis

Zinan Luo, Brian Iaffaldano, Xiaofeng Zhuang, Katrina Cornish*, Department of Horticulture and Crop Science, The Ohio State University

ABSTRACT

RNA-Seq was used to detect sequence variants (i.e. single nucleotide polymorphisms, SNPs) between three high rubber and three low rubber plants. A total of 55,532 contigs were assembled using *de novo* assembly, and 16,891 SNPs were detected. Of those 16,891 SNPs, 77 SNPs of 18 genes were involved in the terpenoid biosynthesis pathway, which is part of the rubber biosynthetic pathway. Forty-two SNPs were finally selected and converted to functional SNP markers using KASP (Kompetitive Allele Specific PCR) technology. These SNPs were validated and then used for simple marker-trait association analysis in an F1 population with 84 individuals. A total of 37 out of 42 SNP markers (88.1%) were polymorphic, 1 was monomorphic (2.38%) and 4 (9.52%) were counted as failed reactions. The marker-trait association analysis identified two SNP markers that were significantly related to rubber content. Both of them were located in the gene encoding 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase in the MEP pathway, suggesting a potential linkage between SNP markers and quantitative trait loci (QTL) controlling rubber production. To conclude, we have established a significant genomic resource for TK, providing a comprehensive transcriptomic reference. The power of RNA-Seq to detect SNPs was validated and putative SNP markers tightly linked to QTL controlling rubber biosynthesis were identified.

INTRODUCTION

Taraxacum kok-saghyz (TK) is a potential alternative for natural rubber production due to its high molecular weight rubber, short life cycle, and diverse environmental adaptation[1]. However, its inability to compete with weeds (e.g. *Taraxacum officinale*) results in low rubber production per acre. In order to improve rubber yield, breeding efforts are necessary. Until now, only limited breeding efforts have been carried out due to TK's less competitiveness and self-incompatibility. The need to grow TK to maturity for accurate rubber yield estimation makes it an ideal crop for marker-assisted selection(MAS), a strategy integrating molecular genetics with traditional breeding efforts in attempt to select for desirable phenotypic traits, such as high rubber, in a short timeframe2. However, limited genomic resources currently available for this species make it difficult to implement MAS. A search on the NCBI EST database with the search term "*Taraxacum kok-saghyz*" only returned 16,441 results. For non-model species like TK, RNA-Seq is considered to be an effective and efficient method for generating a reduced representation of a species' genome, specifically targeting the genome transcripts. In this study, we sequenced the root transcriptome of TK with the aim to identify and functionally annotate a large amount of expressed genes as well as identify SNPs between high rubber and low rubber TK populations for development as molecular markers. Our transcriptomic data also will expand and enrich the publically available TK root transcriptome sequences as well as identify the candidate genes regulating the rubber biosynthetic pathway.

MATERIALS & METHODS

➢ Plant materials

A total of 20 random TK plants were collected from MUCK farm (Willard, OH) in December, 2013. The harvested root tissues were frozen in liquid nitrogen and stored at -80℃ until RNA extraction. An F1 population with 84 individuals was derived from crossing a high rubber plant (TK009) and a low rubber plant (TK069). The rubber content of the plants was quantified using a combined method involving Accelerated Solvent Extraction (ASE) and Near Infrared spectroscopy (NIR).

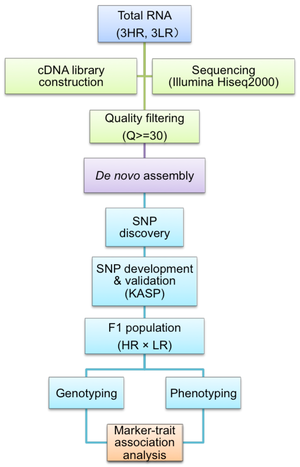

Fig 1. Overall Workflow

➢ RNA and DNA extractions

After rubber quantification, RNA was extracted from root tissues of three high rubber plants (TK6, TK14, TK21) and three low rubber plants (TK9, TK10, TK18). cDNA library was constructed for each RNA sample, and Illumina Hiseq2000 was used for RNA-Seq. Genomic DNA of the RNA-Seq samples was extracted from root tissues following a published protocol[3], including treatment with RNase during homogenization. Genomic DNA of F1 progenies was extracted from leaf samples using the CTAB Geno/Grinder^{TM} procedure with modifications (B. Iaffaldano, unpublished).

➢ SNP marker development and validation

Filtered raw reads from RNA-Seq data were assembled into contigs by *de novo* assembly (Fig1). After RNA-Seq data analysis (raw reads assembly, mapping and SNP calling), SNPs assigned to the terpenoid biosynthesis pathway were selected for further research using KASP genotyping. For each SNP, two allele-specific forward primers and one common reverse primer were designed (LGC Genomics, Teddington, UK). Genotyping reactions were performed in a total volume of 10 µl containing 5 µl 2× Master mix, 0.14 µl Assay mix, and 5 µl genomic DNA (1-10ng/µl). The following cycling conditions were used: 15 min at 94 °C; 10 touchdown cycles of 20 s at 94 °C, 60 s at 61–55°C (dropping 0.6°C per cycle); and 26 cycles of 20s at 94°C, 60s at 55°C. Fluorescence detection of the reactions were performed using a Bio-Rad CFX96 RT-PCR and data were analyzed using the CFX96 manager software.

➢ Preliminary marker-trait association analysis

All the markers were scored as "A", "B" or "H" format for F1 individuals and their parents. The genotypic scores then were combined with rubber content and analyzed using a simple marker-trait model in the R program: Yjk = m+ Mj +Gk(Mj) with the script lm(L~ as.factor(marker)).

RESULTS & DISCUSSION

➢ RNA–Seq

- 55,532 contigs were assembled from raw reads using *de novo* assembly. Of these contigs, 37,317 (67.2%) showed significant BLASTx matches in the nr database. A total of 31,800 contigs were annotated using Blast2GO. KEGG pathway analysis identified 102 contigs involved in the terpenoid backbone biosynthesis pathway.

➢ SNP discovery, marker development & validation

- By SNP calling, 92,150 SNPs were detected among all three low rubber plants, while 78,407 among all three high rubber plants. Finally, 16,891 candidate SNPs were filtered with read depths (DP) >=10 in each sample (Fig.2).

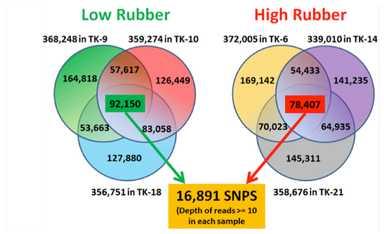

Fig 2. SNP discovery between three high rubber plants and three low rubber plants

- A total of 77 SNPs in 34 contigs (representing 18 enzymes) were involved in terpenoid backbone biosynthesis pathway (Table1).

Table 1. SNPs involved in the terpenoid backbone biosynthesis pathway

Enzymes	# contigs	# contigs w/ SNPs	# SNPs	Developed SNP markers
MVA pathway				
acetyl-CoA C-acetyltransferase (AACT)	7	2	2	SNP1348/ SNP1092
hydroxymethylglutaryl-CoA synthase (HMGS)	3	0	0	/
hydroxymethylglutaryl-CoA reductase (NADPH)	17	5	13	SNP146, SNP19, SNP269/ SNP84/ SNP354,SNP415, SNP117, SNP342/ SNP1249
mevalonate kinase (MVK)	2	0	0	/
phosphomevalonate kinase (PMK)	6	1	1	SNP843
diphosphomevalonate decarboxylase (MVD)	1	0	0	/
MEP pathway				
1-deoxy-D-xylulose-5-phosphate synthase (DXS)	5	1	1	SNP1493
1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)	3	1	1	SNP131
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT)	3	3	19	SNP285, SNP375, SNP770, SNP822
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK)	1	1	4	SNP1113, SNP1245
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS)	2	2	4	SNP332, SNP437, SNP455
(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS)	8	3	8	SNP1440/ SNP215, SNP263, SNP1006, SNP1193
4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR)	4	2	7	SNP1440/ SNP215, SNP263, SNP1006, SNP1193
Downstream				
farnesyl diphosphate synthase	16	5	6	SNP483/ SNP141
geranylgeranyl diphosphate synthase	12	3	3	SNP521
heptaprenyl diphosphate synthase	8	4	4	SNP213/ SNP52
geranyl-diphosphate synthase	23	10	19	SNP483/ SNP141/ SNP213/ SNP1770, SNP389, SNP925/ SNP52/ SNP140
delta-isomerase	5	0	0	/
prenylcysteine oxidase	3	2	3	SNP1451
all-trans-nonaprenyl-diphosphate synthase [geranylgeranyl-diphosphate specific]	4	4	7	SNP213/ SNP52/ SNP140
all-trans-nonaprenyl-diphosphate synthase [geranyl-diphosphate specific]	4	4	7	SNP213/ SNP52/ SNP140
protein-S-isoprenylcysteine O-methyltransferase	4	0	0	/
NADP+-dependent farnesol dehydrogenase (NADP+)	3	1	1	SNP435

- A total of 42 SNP markers were finally selected and converted to KASP genotyping SNP assays (Fig3). Thirty-seven out of 42 SNPs (88.1%) were polymorphic, 1 was monomorphic (2.38%) and 4 (9.52%) failed reactions. Of the 38 assays worked, 32 (84.2%) had an MAF>0.1, 24 (63.2%) had an MAF>0.2, with an average MAF= 0.28.

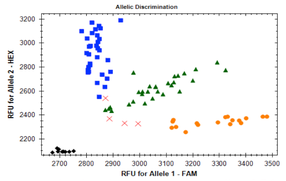

Fig 3. An example showing genotyping calls of 84 genotypes for SNP1245 using KASP chemistry on a Bio-Rad CFX96 RT-PCR

➢ Simple marker-trait association analysis

- Two SNP markers (Fig4 & 5) located in 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK) in MEP pathway were found to be significantly related to rubber content, suggesting a potential QTL controlling rubber production, but more reliable data will be obtained after genotyping-by-sequencing (GBS) and QTL mapping.

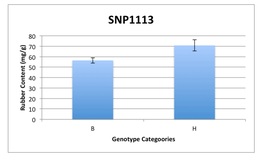

Fig 4. SNP1113 significantly associated with rubber content

The heterozygous genotypes (A/G or H) have an average of 14.35mg/g rubber higher than the homozygous genotypes (G/G or B).

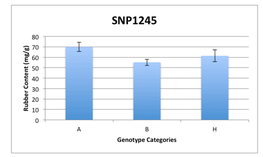

Fig 5. SNP1245 significantly associated with rubber content

The homozygous genotypes (T/T or A) have an average of 69.87mg/g rubber, followed by the heterozygous genotypes (T/C or H) with 61.36mg/g and another homozygous genotypes (C/C or B) with 55.14mg/g rubber.

CONCLUSIONS

- This study applies RNA-Seq to provide a comprehensive TK root transcriptome reference, which serves as an extensive genetic resource for a non-model species (i.e.TK) currently lacking such resources.
- SNPs involved in the terpenoid backbone biosynthesis pathway were identified and developed to SNP markers, two of which were identified to be significantly related to rubber content.
- The developed SNP markers can be used in further molecular breeding research.

BIBLIOGRAPHY

1. Van Beilen, J.B. & Poirier, Y. (2007). Guayule and Russian dandelion as alternative sources of natural rubber. Critical Reviews in Biotechnology, 27(4), 217–31.
2. Collard, B.C.Y et al. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica, 142(1-2), 169–196.
3. Tel-zur, N et al. (1999). Modified CTAB Procedure for DNA Isolation from Epiphytic Cacti of the Genera *Hylocereus* and *Selenicereus* (Cactaceae). Plant Molecular Biology Reporter, 17: 249-254.

ACKNOWLEDGEMENTS

- This project was supported by the Center of Applied Plant Sciences of The Ohio State University.
- Special thanks to all members of the Francis lab, Dorrance lab, Cardina lab and Sneller lab.

THE OHIO STATE UNIVERSITY
COLLEGE OF FOOD, AGRICULTURAL, AND ENVIRONMENTAL SCIENCES

Contact:
Ohio Agricultural Research and Development Center
212A Williams Hall
1680 Madison Avenue
Wooster OH 44691
Phone: 614-980-6116
Email: luo.356@osu.edu