RNA-Seq: Identification of rubber-related genes and genetic polymorphisms associated with yield

Zinan Luo, Brian Iaffaldano, Xiaofeng Zhuang, Jonathan Fresnedo and Katrina Cornish*

Department of Horticulture and Crop Science, The Ohio State University, Wooster, Ohio 44691

ABSTRACT

A comprehensive *Taraxacum kok-saghyz* (TK) transcriptomic reference was generated. A large set of SNPs can be used in genetic mapping and genotyping efforts, and among them, molecular markers related to rubber biosynthesis can be developed. A total of 55,532 contigs with lengths over 200 base pairs were *de novo* assembled from TK by RNA-Seq, Genome-wide SNPs, TK rubber-related homologues (472) and differentially expressed genes (158) between high and low rubber genotypes were identified. Two out of 42 developed SNP markers were significantly associated with rubber content. This transcriptomic resource provides a solid foundation for further QTL mapping and MAS in TK.

INTRODUCTION

Taraxacum kok-saghyz (TK) is a potential alternative plant for natural rubber production [1] due to its high molecular weight rubber, short life cycle, and diverse environmental adaptation. However, improvements in rubber yield and agronomic relevant traits are still required before it can become a commercially-viable crop. Marker-assisted selection (MAS) seems an appropriate strategy [2] given TK's heterogeneous-self-incompatible genetic base and annual breeding cycle. The development of RNA-Seq transcriptome datasets and identification of exonic markers linked to rubber biosynthesis related loci will provide a definitive tool for subsequent application of MAS in the pursuit of TK breeding.

MATERIALS AND METHODS

➢ **RNA extraction and RNA-Seq**

RNA was extracted from root tissues of six samples (3 high rubber and 3 low rubber). A cDNA library was constructed for each RNA sample, and Illumina Hiseq2000 was used for RNA-Seq.

➢ **TK homologue identification**

After *de novo* assembly, the assembled contigs were annotated for protein functions using the NCBI non-redundant (nr) protein database. Then a local database of 50 rubber-related proteins was established and a blastx search with a cut-off e-value of $1e^{-10}$ for the homologue identification in our transcripts was performed.

➢ **Differentially expressed genes**

Differentially expressed genes between high rubber and low rubber groups were identified using EBSeq, and finally visualized using heatmap2 package in the R package.

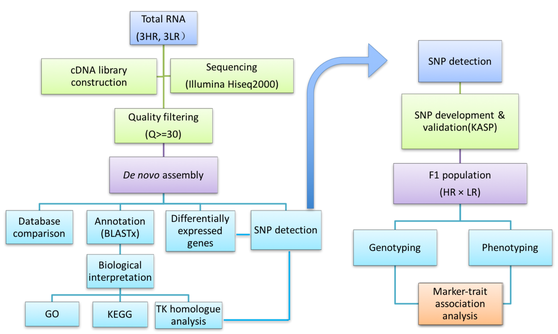

Fig. 1. Overall workflow.

RESULTS

➢ **SNP detection and SNP marker development**

Filtered raw reads from RNA-Seq data were assembled into contigs by *de novo* assembly (Fig. 1). After RNA-Seq data analysis (raw reads assembly, mapping and SNP calling), Genome-wide SNPs, as well as the SNPs assigned to the TK homologues and differentially expressed genes, were selected for further research using KASP genotyping technology.

➢ **Genotyping and phenotyping of an F_1 segregating family**

Genomic DNA of F_1 progenies derived from crossing a high and low rubber parent was extracted from leaf samples using the modified CTAB Geno/GrinderTM procedure [3]. KASP genotyping reactions were performed in the F1 individuals following the manufacturer's instructions. At the same time, field trials were conducted in three different locations and the F1 individuals were then phenotyped with rubber content and related traits.

➢ **Preliminary marker-trait association analysis**

By combining the genotypic data with rubber content, followed by a simple marker-trait model in the R package, $Y_{jk} = m+M_j+G_k(M_j)$ with the script *lm(L~ as.factor(marker))*, then SNP markers with significantly p-values ($p<0.05$) were considered as associated with rubber content.

➢ **TK homologues related to rubber biosynthesis**

Forty-seven out of 50 publically available rubber-related proteins generated a total of 472 homologues transcripts in TK. These transcripts are the copy version of rubber-related proteins previously identified in other rubber-producing species. The homologues with the highest similarity are shown in Table 1.

Table 1. TK homologues with the highest similarity to rubber-related proteins in other rubber-producing crops or members of Asteraceae.

Homologue sequences	Identity percentage	Protein name	Species	No. of SNPs
REF & SRPP				
comp27426_c0_seq2	55.294	REF	*H. brasiliensis*	2
	51.22	SRPP	*H. brasiliensis*	2
comp32369_c0_seq2	62.5	GHS (Guayule homolog of SRPP)	*P. argentatum*	2
CPT				
comp31947_c1_seq8	100	CPT3	*T. brevicorniculatum*	3
comp34122_c0_seq1	91.026	CPT1	*L. sativa*	2
CPTL				
comp20754_c0_seq2	98.008	CPTL1	*T. officinale*	1
MVA pathway				
comp35735_c0_seq22	84.034	AACT	*H. brasiliensis*	4
comp24631_c1_seq1	90.698	HMG CoA reductase (HMGR)	*H. brasiliensis*	4
comp33776_c0_seq1	72.021	Mevalonate kinase (MVK)	*H. brasiliensis*	0
comp27904_c1_seq7	72.586	Phosphomevalonate kinase (PMK)	*H. brasiliensis*	1
comp32262_c0_seq14	78.604	HMG CoA synthase (HMGS2)	*H. brasiliensis*	0
comp28398_c1_seq1	79.137	MVD	*H. brasiliensis*	0
MEP pathway				
comp32604_c0_seq1	59.673	DXS1	*H. brasiliensis*	1
comp30628_c0_seq10	83.262	DXR	*H. brasiliensis*	1
comp26131_c0_seq2	72.457	CMK	*H. brasiliensis*	1
comp23703_c0_seq1	92.547	MCS2	*H. brasiliensis*	2
	91.925	MCS1	*H. brasiliensis*	2
comp33968_c0_seq22	86.198	HDS	*H. brasiliensis*	0
comp19447_c0_seq1	90	CMS1	*H. brasiliensis*	0
comp32227_c0_seq7	86.387	HDR	*H. brasiliensis*	1
Downstream				
comp10115_c0_seq2	76.757	FPP synthase	*H. brasiliensis*	1
comp35494_c0_seq13	91.111	GPP synthase	*H. brasiliensis*	3
comp30039_c0_seq1	80.347	GGPP synthase	*H. brasiliensis*	0
comp32029_c1_seq15	83.691	IDP isomerase	*H. brasiliensis*	0
Inulin biosynthesis				
comp31611_c0_seq2	97.271	1-SST	*T. officinale*	1
comp32671_c0_seq5	75	1-FFT	*L. sativa*	2
comp35385_c0_seq3	79.741	1-FEH	*H. tuberosus*	4
Others				
comp34751_c1_seq3	72.917	Patatin-like latex allergen/rubber biosynthesis inhibitor	*H. brasiliensis*	1
comp23840_c0_seq2	91.824	Rubber biosynthesis stimulator protein	*H. brasiliensis*	1

➢ **Differentially expressed genes**

A total of 158 differentially expressed genes between high rubber and low rubber groups were identified (Fig. 2). Ninety-four of them were upregulated in high rubber group and 64 were upregulated in low rubber group.

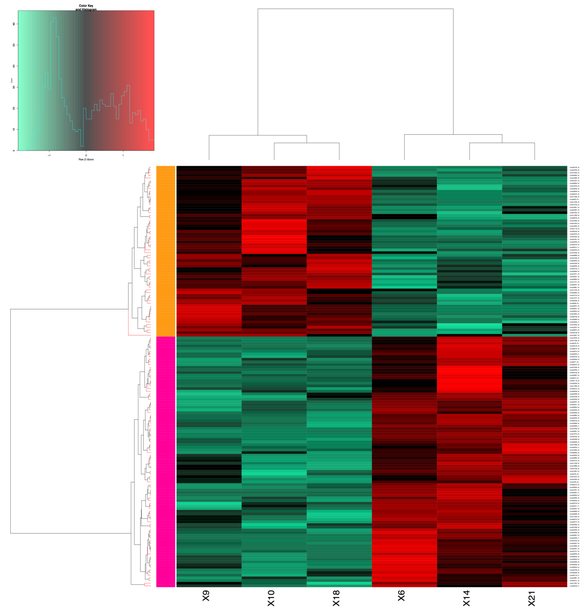

Fig. 2. Differentially expressed genes between high rubber and low rubber groups. The red and green cells are up- and down-regulated genes, respectively. X9, X10, and X18 are low rubber, while X6, X14, X21 are high rubber samples.

➢ **SNP detection and SNP marker development**

A total of 21,036 genome-wide SNPs were identified (Fig. 3), among which 112 and 117 SNPs were assigned on the TK homologues (Fig.2) and differentially expressed genes, respectively. A total of 42 SNPs were finally developed as KASP markers for F_1 genotyping (Fig. 4).

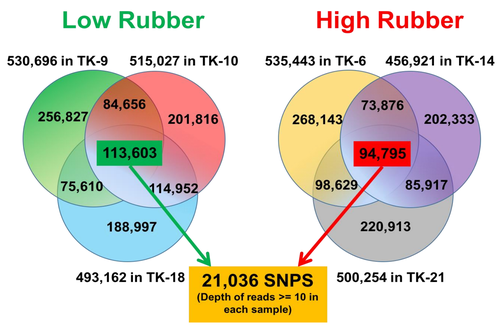

Fig. 3. Genome-wide SNPs identified after SNP calling.

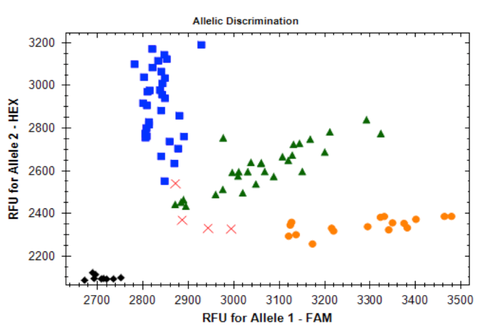

Fig. 4. An example showing genotyping calls of 84 genotypes for SNP1245 using KASP chemistry on a Bio-Rad CFX96 RT-PCR.

➢ **Simple marker-trait association analysis**

Two SNP markers (Fig. 5 and 6) located in 4-diphospho-cytidyl-2-C-methyl-D-erythritol kinase (CMK) in the MEP pathway were significantly associated to rubber content in the F_1 family, suggesting a potential QTL-controlling rubber production. However, more reliable data will be obtained after genotyping-by-sequencing (GBS) and QTL mapping.

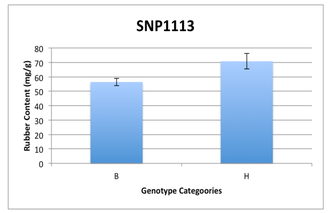

Fig. 5. SNP1113 significantly associated with rubber content. *The heterozygous genotypes (A/G or H) have an average of 14.35 mg/g rubber higher than the homozygous genotypes (G/G or B).*

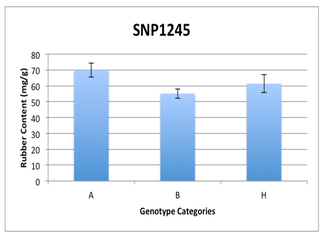

Fig. 6. SNP1245 significantly associated with rubber content. *The homozygous genotypes (T/T or A) have an average of 69.87mg/g rubber, followed by the heterozygous genotypes (T/C or H) with 61.36mg/g, and another homozygous genotypes (C/C or B) with 55.14mg/g rubber.*

CONCLUSIONS

- A significant genomic resource for TK has been developed. It provides a comprehensive transcriptomic reference, a large set of SNPs to be used in genetic mapping and genotyping efforts, and among them, putative markers related to rubber biosynthesis. This resource provides a solid foundation for further QTL mapping and MAS in TK.
- SNPs in the rubber-related genes as well as other differentially expressed genes were identified and two SNP markers were significantly associated with rubber content. These SNPs now are KASP markers that can be used in further molecular breeding research.
- The KASP genotyping platform proved to be an efficient method for large-scale population genotyping.

ACKNOWLEDGEMENTS

- We thank the Center of Applied Plant Sciences, the PENRA consortium, OARDC, Ohio Third Frontier, and USDA NIFA Hatch project 230837 for their support.
- Special thanks to all members of the Francis, Dorrance, Cardina and Sneller labs.

REFERENCES

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.