

# Construction and Application of the Draft Genome of *Taraxacum kok-saghyz*, an Alternative Natural Rubber Resource

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

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

## ABSTRACT

Rubber dandelion (*Taraxacum kok-saghyz*, TK) is being developed as an alternative natural rubber resource in response to increasing global rubber market demand. However, the lack of genomic resources specific to TK hampered its domestication from a wild species to a crop. Here we present the TK draft genome. The nuclear genome consists of 188,608 scaffolds covering 1,440 Mb, with a N50 scaffold size of 55.2 Kb. The genome contains 63,912 protein-coding genes by preliminary annotation. In addition, the complete Chloroplast Genome (151,338 bp) contains 134 genes including 82 protein-coding genes; The TK Mitochondrial Genome (351 Kb, incomplete) consists of 96 scaffolds, with a N50 size of 27 Kb. Currently, the TK genome has been used as a platform on the study of functional genomics, comparative genomics, Quantitative Trait Locus (QTL) mapping and Marker-assisted selection (MAS) TK breeding.

## INTRODUCTION

Natural rubber is a high molecular weight polymer that cannot be replaced by synthetics in most applications [1]. Rubber is currently harvested from the Para rubber tree (*Hevea brasiliensis*), but is faced with serious challenges, such as potential shortages of supply and pathogen threats [2].

*Taraxacum kok-saghyz* (TK) is an alternative rubber plant, which produces high molecular weight rubber very similar to Hevea rubber. Also, TK's short life cycle, and adaptation to diverse environments make it a potentially ideal rubber-producing crop. However, difficulties with conventional breeding along with limited genome-based information have impeded efficient TK crop improvement.

Here, we describe assembly of the TK reference genome, which will aid accelerated domestication TK and production of high-yielding clones.

## MATERIALS AND METHODS

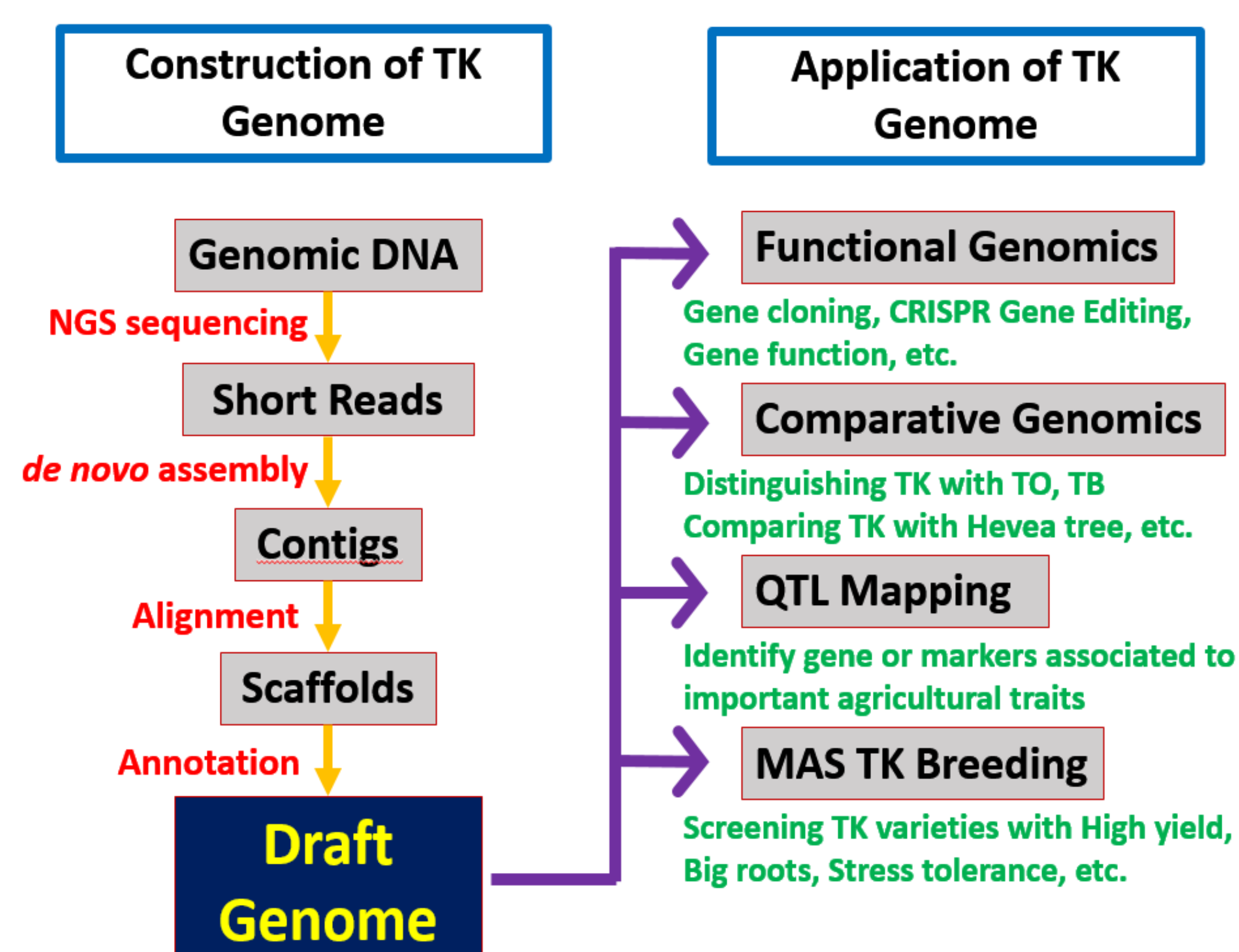


Figure 1. Workflow of TK whole genome sequencing project

## RESULTS

### Construction of the TK genome

#### TK Nuclear Genome

- More than 2 billion high quality reads (sequencing depth 192 X) were produced by the Illumina HiSeq2500 platform.
- *De novo* assembly by the SOAPdenovo2 [3 software] yielded 188,608 scaffolds, covering 1.4 Gb, with an N50 scaffold size of 55.2 Kb. The assembly results are better than the draft genome of the *Hevea* rubber tree (Table 1) [4].
- A total of 63,912 protein-coding genes were predicted by whole genome annotation using program Augustus-3.2.1 [5].

Table 1. Comparison of assembly results between TK and the Hevea rubber tree [4]

	<i>T. kok-saghyz</i>	Rubber tree
Scaffold number	188,608	608,017
Total scaffold length	1.4 Gb	1.1 Gb
Average scaffold length	7,631 bp	1,840 bp
Longest scaffold	1,077,596 bp	531,465 bp
N50	55,183 bp	2,972 bp
Predicted genes	63,912	68,955

#### TK Chloroplast Genome

- The TK Chloroplast Genome is 151,338 bp. A total of 134 genes have been identified, including 82 protein-coding genes, 8 rRNA genes, 36 tRNA genes, as well as 8 pseudogenes and Open Reading Frames (ORFs) [6].

#### TK Mitochondrial Genome

- The incomplete TK Mitochondrial genome consists of 96 Scaffolds, covering 351 Kb, with a N50 Scaffold size of 27 Kb.

### Application of the TK genome

#### Functional Genomics

- CRISPR Gene Editing of TK can be used to confer describable domestication traits and enhance rubber yield [7].
- Genes involved in important agronomic traits have been identified and cloned [Jonathan Fresnedo; Kyle A. Benzle, unpublished].

#### Comparative Genomics

- Genomic signatures have been identified which effectively differentiate *T. officinale* (common dandelion), *T. brevicorniculatum* (seed contaminant) as well as interspecific hybrids with TK [6].

#### QTL Mapping

- Two SNP markers associated with the rubber biosynthesis genes have been found indicative of TK rubber content [Luo et al. in prep].

#### MAS TK Breeding

- More than 16k SNP markers were developed after comparing three high and three low rubber content TK genotypes [Luo et al. in prep], and will be useful for MAS of high rubber content.

## DISCUSSION

Genome sequencing is now affordable, but assembling plant genomes *de novo* remains challenging for heterozygous crops such as TK, for a variety of biological, and computational reasons [8]. Although improvement and completeness of TK organellar genomes has been successful, the TK nuclear genome is heavily fragmented and non-cytologically correspondent with the chromosomal composition of TK, due to the high rates of heterozygosity and repetitive sequences in the genome. In order to address this issue, we have started a second round of sequencing using PacBio technology in order to produce longer reads (average > 10 Kb, some reads > 60 Kb) and then have a better scaffolding (backbone) for alignment of short reads. Combined with the draft nuclear genome we present here, a improved, finest-assembled, and well-annotated genome will be constructed by the end of this year.

In order to make TK commercialization a reality, domestication, genetic improvement and breeding must be carried out to develop high rubber TK germplasm, which is agroecologically adapted to conventional US farming systems. Whole genome sequencing of TK can provide valuable information to accelerate the introduction of this economically important new industrial crop.

## CONCLUSIONS

These genomic resources will assist in the genetic dissection of rubber related traits, the identification of marker-trait associations, genome-wide and genome-targeted studies, and finally, a genomics platform upon which the genetic improvement of TK may rely. Accelerated breeding and domestication will help empower the US to produce, rather than import, 1.2 million metric tons natural rubber per year, satisfy the internal demand, and eventually lead to rubber exports.

## ACKNOWLEDGEMENTS

This work was funded by the Center for Applied Plant Sciences, the PENRA consortium, OARDC, Ohio Third Frontier, and USDA NIFA Hatch project 230837. The authors thank the Ohio Supercomputer Center for providing hardware support for partial bioinformatics analysis.

## REFERENCES

1. Collins-Silva, J. et al. (2012). *Phytochemistry* 79,46-56.
2. van Beilen, J.B., and Poirier, Y. (2007). *Critical reviews in biotechnology* 27, 217-231.
3. Luo, R.B., et al. (2012). *Gigascience* 1.
4. Rahman, A.Y. et al. (2013). *BMC genomics* 14,75.
5. Stanke, M. et al. (2006). *Genome Biology* 7, S11.
6. Zhang Y. et al. (2017). *BMC Plant Biology*, 17:34
7. Iaffaldano B. et al. (2016). *Ind. Crop. and Prod.* 89: 356-362
8. Schatz M.C. et al. (2012) *Genome Biology*, 13:243