

The first high-quality chromosomal genome assembly of a medicinal and edible plant *Arctium lappa*

Liang Xu¹, Shengnan Li², Yanyun Yang¹, Yanping Xing¹, Zhongren Zhang³, Tao Liu¹, Wuliji Ao¹, Guihua Bao¹, Zhilai Zhan¹, Rong Zhao¹, Tingting Zhang¹, Dachuang Zhang¹, Tingguo Kang¹, Yueyue Song¹, and Che Bian¹

¹Affiliation not available

²Liaoning University of Traditional Chinese Medicine - Dalian Campus

³Novogene Bioinformatics Institute

May 4, 2021

Abstract

Arctium lappa has a long medicinal and edible history with great economic importance. We combined Illumina and PacBio sequences to generate the first high-quality chromosome-level draft genome of *A. lappa*. The assembled genome is approximately 1.79 Gb with a N50 contig size of 6.88 Mb. Approximately 1.70 Gb (95.4%) of the contig sequences were anchored onto 18 chromosomes using Hi-C data; the scaffold N50 was improved to be 91.64 Mb. Furthermore, we obtained 1.12 Gb (68.46%) of repetitive sequences and 32,771 protein-coding genes; 616 positively selected candidate genes were identified. Additionally, we compared the transcriptomes of *A. lappa* roots at three different developmental stages and identified 8,943 differentially expressed genes (DEGs) in these tissues. Among candidate genes related to lignan biosynthesis, the following were found to be highly correlated with the accumulation of arctiin: 4-coumarate-CoA ligase (4CL), dirigent protein (DIR), and hydroxycinnamoyl transferase (HCT). These data can be utilized to identify genes related to *A. lappa* quality or provide a basis for molecular identification and comparative genomics among related species.

Hosted file

manuscript.pdf available at <https://authorea.com/users/411779/articles/520689-the-first-high-quality-chromosomal-genome-assembly-of-a-medicinal-and-edible-plant-arctium-lappa>