РАЗДЕЛ I

БИОИНФОРМАТИКА

DOI: 10.25205/978-5-4437-1843-9-1

ASSEMBLY OF COMPLETE SEA BUCKTHORN GENOME USING NANOPORE SEQUENCING AND THE HIFIASM ONT ALGORITHM *

A. A. Arkhipov[1], E. N. Pushkova[1], N. L. Bolsheva[1], V. V. Babenko[2], K. M. Klimina[2], Yu. A. Zubarev[3], N. V. Melnikova[1], A. A. Dmitriev[1]

[1] *Engelhardt Institute of Molecular Biology RAS, Moscow*
[2] *Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, FMBA of Russia, Moscow*
[3] *Federal Altai Scientific Center of Agrobiotechnologies, Barnaul*

✉ arkhipov.aleksandr2.0@gmail.com

Abstract

Sea buckthorn is used in food and pharmaceuticals. We sequenced the DNA of sea buckthorn on the Oxford Nanopore Technologies platform and assembled the genome using the Hifiasm ONT algorithm. The assembly was 1.17 Gb long, and 11 of 12 chromosomes were complete. The BUSCO completeness was 96.8 %. Telomeric repeats were found at both ends of 11 of 12 chromosomes. Thus, a near-telomere-to-telomere (T2T) assembly of the sea buckthorn genome was obtained.

Sea buckthorn (*Hippophae rhamnoides* L.) is a valuable woody oil plant, which is used in food and pharmaceuticals. However, from a genetic point of view, sea buckthorn is an understudied species. In recent years, several genomes of *H. rhamnoides* were assembled using third-generation sequencing and the Hi-C approach, but they are still far enough from the telomere-to-telomere (T2T) level. Meanwhile, T2T genome assemblies are becoming the new standard in plant genomics. We aimed to obtain a T2T assembly of the *H. rhamnoides* ssp. *mongolica* genome, which is cultivated in Russia.

We obtained pure high-molecular-weight DNA of the variety Triumf of *H. rhamnoides* ssp. *mongolica* using the nuclei isolation with subsequent CTAB extraction and column purification. Then DNA was prepared using the SQK-LSK114 kit and sequenced on PromethION (Oxford Nanopore Technologies (ONT)) with R10.4.1 flow cells.

About 155 Gb of raw ONT reads with an N50 of 31.4 kb were obtained. Basecalling was performed using the Dorado tool (v.1.0.2, dna_r10.4.1_e8.2_400bps_sup@v5.2.0). Reads with a length less than 10 kb and an average quality less than Q10 were filtered out. Genome assembly was performed using the Hifiasm ONT algorithm (v.0.25.0) with parameters optimized for ONT data. Assembly parameters, such as length, number of contigs, and others were evaluated using QUAST (v.5.3.0) and BUSCO (v.5.8.3). The TIDK software package (v.0.2.64) was used to detect telomeric sequences in the assembly.

The obtained genome assembly of the variety Triumf was presented by 13 contigs: 11 complete chromosomes and one chromosome consisting of two contigs. The assembly had a high BUSCO completeness of 96.8%. We compared the Triumf genome assembly with the previously obtained assemblies of *H. rhamnoides*: NCBI, GCA_033030585.1 and CNGB, CNP0001846. The Triumf genome assembly was longer than the other genome assemblies (1.17 vs 0.92 and 0.73 Gb) and had a higher contig N50 value (83.7 vs 14.8 and 3.5 Mb).

Eleven of twelve chromosomes of the Triumf genome assembly had telomeric repeats at both ends and one chromosome (Chr 3) had telomeric repeats at one end. We also searched for telomeric repeats in other available sea buckthorn genome assemblies: NCBI, GCA_033030585.1 and CNGB, CNP0001846. Absence of telomeric repeats at chromosome ends or their incorrect location were frequent, indicating incomplete or erroneous assemblies.

* The study is supported by the Russian Science Foundation (project #23-46-00026).

Thus, the near-T2T genome of the variety Triumf was assembled. The assembly surpassed the previously obtained genomes of *H. rhamnoides* in terms of length, contiguity, and presence of telomeric repeats at chromosome ends. The genome assembly of the variety Triumf is an essential tool for a range of basic and applied studies on sea buckthorn, including genome editing.