

GENOME SEQUENCING

The Complete Genome Sequence of *Ceratonia siliqua* L. (Fabaceae, Fabales), Carob

Omar Chlyah¹, Meriem Alaoui Mdarhi², Douae El Ghouali³, Michael Kieras⁴, Stacy Pirro⁴, Hassan Ghazal^{3,4,5}

¹ Laboratory of Plant Physiology and Biotechnology, Mohammed V University in Rabat, ² Biotechnology Unit, National Institute of Agricultural Research,

³ Laboratory of Genomics, Bioinformatics and Digital Health, Mohammed VI University of Science and Health, ⁴ Biodiversity, Iridian Genomes, ⁵ Department of Sports Sciences, Laboratory of Sports Sciences and Performance Optimization, Royal Institute of Executive Management for Youth and Sport

<https://doi.org/10.56179/001c.92058>

Biodiversity Genomes

We present the whole genome sequence of *Ceratonia siliqua* L. Illumina paired-end reads were assembled by a de novo method followed by a finishing step. The raw and assembled data are publicly available via GenBank: Sequence Read Archive (SRR24502586) and assembled genome (JASKGM000000000).

Introduction

Ceratonia siliqua L., or carob, is a flowering evergreen tree or shrub in the legume family, Fabaceae. It is widely cultivated for its edible fruit pods, and as an ornamental tree in gardens and landscapes. The carob tree is native to the Mediterranean region and the Middle East. Portugal is the largest producer of carob, followed by Italy and Morocco (Brassesco et al. 2021).

In the Mediterranean Basin, extended to the southern Atlantic coast of Portugal, carob pods are often used as animal feed. The ripe, dried, and sometimes toasted pods are often grounded into carob powder, which can be used as an alternative to cocoa powder.

Methods

A single leaf from a Moroccan cultivated tree was used for this study. DNA extraction was performed using the Qiagen DNAeasy genomic extraction kit using the standard process. A paired-end sequencing library was constructed using the Illumina TruSeq kit, according to the manufacturer's instructions. The library was sequenced on an Illumina Hi-Seq platform in paired-end, 2 × 150bp format. The resulting fastq files were trimmed of adapter/primer sequence and low-quality regions with Trimmomatic v0.33 (Bolger, Lohse, and Usadel 2014). The trimmed sequence was assembled by SPAdes v2.5 (Bankevich et al. 2012) followed by a finishing step using Zanfona v1.0 (Kieras 2021) to make additional contig joins based on conserved regions in related species.

Data availability

Raw reads (SRR24502586) and the assembled genome (JASKGM000000000) are available in Genbank.

Funding

Funding was provided by Iridian Genomes, grant# **IRGEN_RG_2021-1345**
Genomic Studies of Eukaryotic Taxa

Conflict of Interest Statement

The authors declare they have no conflicts of interests.

Submitted: January 04, 2024 EST, Accepted: January 06, 2024 EST

REFERENCES

- Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology* 19 (5): 455–77. <https://doi.org/10.1089/cmb.2012.0021>.
- Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brassesso, María Emilia, Teresa R.S. Brandão, Cristina L.M. Silva, and Manuela Pintado. 2021. “Carob Bean (*Ceratonia Siliqua* L.): A New Perspective for Functional Food.” *Trends in Food Science & Technology* 114: 310–22. <https://doi.org/10.1016/j.tifs.2021.05.037>.
- Kieras, M. 2021. *Zanfona, a genome finishing process for use with paired-end short reads*. <https://github.com/zanfona734/zanfona>.