Badad, Oussama, Stacy Pirro, Qamar Lahlimi, and Hassan Ghazal. 2023. "The Complete Genome Sequence of *Verbascum Thapsus* (Scrophulariaceae, Lamiales), the Common Mullein." *Biodiversity Genomes*, March. https://doi.org/10.56179/001c.73050.

GENOME SEQUENCING

The Complete Genome Sequence of *Verbascum thapsus* (Scrophulariaceae, Lamiales), the Common Mullein

Oussama Badad[1], Stacy Pirro[2], Qamar Lahlimi[1], Hassan Ghazal[3,4,5]

[1] Laboratory of Plant and Microbial Biotechnology, Biodiversity and Environment, Faculty of Sciences, Mohammed V University in Rabat, 10000, Rabat, Morocco, [2] Biodiversity, Iridian Genomes, [3] Laboratory of Genomics, Bioinformatics and Digital Health, Mohammed VI University of Sciences and Health, Mohammed VI University of Sciences and Health, Casablanca, Morocco, [4] Laboratory of Genomics, Bioinformatics and Digital Health, Mohammed VI Center for Research and Innovation, Casablanca, Morocco, [5] National Center for Scientific and Technical Research (CNRST), Rabat, Morocco

Keywords: mullein, viridiplantae, genome

https://doi.org/10.56179/001c.73050

Biodiversity Genomes

Verbascum thapsus is a biennial plant native to Europe, northern Africa, and Asia and introduced in the Americas and Australia. We present the whole genome sequence of this species. 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 (SRR18183247) and assembled genome (JAOXOC000000000).

Introduction

Verbascum thapsus, or the Common Mullein, is a biennial plant that can grow to 2 m tall or more. Its small, yellow flowers are densely grouped on a tall stem, which grows from a large rosette of leaves. It grows in a wide variety of habitats but prefers well-lit areas and can grow from long-lived seeds that persist in the soil. It is a common weedy plant that spreads by prolifically producing seeds and has become invasive in temperate world regions.

Methods

A single wild-collected specimen was used for sequencing. 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 × 150 bp format. The resulting fastq files were trimmed of adapter/primer sequences 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.

Results

The genome assembly yielded a total sequence length of 560,917,718 bp.

……

Data availability

Raw and assembled data are publicly available via GenBank.

RAW GENOME DATA

https://www.ncbi.nlm.nih.gov/sra/?term=SRR18183247

ASSEMBLED GENOME

https://www.ncbi.nlm.nih.gov/nuccore/JAOXOC000000000

Funding

Funding was provided by Iridian Genomes, grant# IRGEN_RG_2021-1345. Hassan Ghazal is a US NIH grant recipient through the H3abionet/H3africa consortium U24HG006941.

Submitted: February 14, 2023 EDT, Accepted: March 07, 2023 EDT

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.

Kieras, M. 2021. *Zanfona, a genome finishing process for use with paired-end short reads*. https://github.com/zanfona734/zanfona.