El Ouafi, Zainab, Stacy Pirro, Najib Al Idrissi, and Hassan Ghazal. 2023. “The Complete Genome Sequence of *Curcuma Longa* (Zingiberaceae, Zingiberales), Turmeric.” *Biodiversity Genomes*, March. https://doi.org/10.56179/001c.73625.

GENOME SEQUENCING

The Complete Genome Sequence of *Curcuma longa* (Zingiberaceae, Zingiberales), Turmeric

Zainab El Ouafi[1,2], Stacy Pirro[3], Najib Al Idrissi[1,2], Hassan Ghazal[1,2,4]

[1] Laboratory of Genomics, Bioinformatics and Digital Health, Mohammed VI University of Sciences and Health, Casablanca, Morocco, [2] Laboratory of Genomics, Bioinformatics and Digital Health, Mohammed VI Center for Research and Innovation, Rabat, Morocco, [3] Iridian Genomes, Bethesda, MD 20817, USA, [4] National Center for Scientific and Technical Research (CNRST), Rabat, Morocco

Keywords: turmeric, genome, viridiplantae

https://doi.org/10.56179/001c.73625

Biodiversity Genomes

Curcuma longa is a perennial native to India and Southeast Asia. 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 (SRR11229490) and assembled genome (JAOBBC000000000).

Introduction

Curcuma longa, commonly known as Turmeric, is a perennial native to India and Southeast Asia. The rhizomes are used as a coloring and flavoring agent in many Asian cuisines and for dyeing cloth a deep orange-yellow.

Methods

A single cultivated specimen 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 sequences and low-quality regions with Trimmomatic v0.33 (Bolger, Lohse, and Usadel 2014). The trimmed sequences were 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 581,175,068bp.

……

Data availability

Raw and assembled data is publicly available via GenBank:

RAW GENOME DATA

https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR11229490

ASSEMBLED GENOME

https://www.ncbi.nlm.nih.gov/nuccore/JAOBBC000000000

Funding

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

Conflict of Interest Statement

The authors declare they have no conflicts of interest.

Submitted: March 08, 2023 EDT, Accepted: March 17, 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.