#### GENOME SEQUENCING

# The Complete Genome Sequence of *Adansonia digitata* (Malvaceae, Malvales), the African Baobab

Stephen Woods<sup>1</sup>, Kathleen O'Neill<sup>2</sup>, Stacy Pirro<sup>2</sup>

 $^{\rm 1}$  California Baobabs, Simi Valley, CA,  $^{\rm 2}$  Biodiversity, Iridian Genomes

Keywords: baobab, africa, viridiplantae, genome https://doi.org/10.56179/001c.72789

#### **Biodiversity Genomes**

*Adansonia digitata*, the African Baobab, is a long-lived tree species found in sub-Saharan Africa. 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 (SRR23340274) and assembled genome (JAQSVH00000000).

#### Introduction

The African Baobab is a large tree species in the dry savannah regions of sub-Saharan Africa. They are an important source of food and shelter for many species of African wildlife, and culturally important for African communities.

## Methods

A single captive-bred 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,  $2 \times 150$ bp 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.

## Results

The genome assembly yielded a total sequence length of 686,996,791.

## 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=SRR23340274

#### ASSEMBLED GENOME

https://www.ncbi.nlm.nih.gov/nuccore/JAQSVH00000000

## 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: February 22, 2023 EST, Accepted: March 05, 2023 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. <u>https://doi.org/10.1089/cmb.2012.0021</u>.
- Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20. <u>https://doi.org/10.1093/</u> <u>bioinformatics/btu170</u>.
- Kieras, M. 2021. Zanfona, a genome finishing process for use with paired-end short reads. https://github.com/zanfona734/zanfona.