GENOME SEQUENCING

The Complete Genome Sequence of *Chlorophytum comosum* (Asparagaceae, Asparagales), the Spider Plant

Abdellah Idrissi Azami[1], Kathleen O'Neill, Stacy Pirro[2], Hassan Ghazal[1,3]

[1] Laboratory of Genomics and Bioinformatics, School of Pharmacy, Mohammed VI University of Health Sciences, [2] Biodiversity, Iridian Genomes, [3] National Center for Scientific and Technical Research

Biodiversity Genomes

Chlorophytum comosum is a species of evergreen perennial flowering plant native to tropical and southern Africa but has become naturalized in other parts of the world, including western Australia and Bangladesh. We present the whole genome sequence of *Chlorophytum comosum*. 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 (SRR11638255) and assembled genome (GCA_025212335).

Introduction

Chlorophytum comosum is a species of evergreen perennial flowering plant native to tropical and southern Africa, but has become naturalized in other parts of the world, including western Australia and Bangladesh. The species is also popular as a houseplant. *Chlorophytum comosum* grows to approximately 60 cm tall, although as a hanging plant, it can descend many feet. It has fleshy, tuberous roots, each approximately 5–10 cm long. The long narrow leaves reach a length of 20–45 cm and are around 6–25 millimeters wide. Flowers are produced in a long, branched inflorescence, which can reach a length of up to 75 cm and eventually bend downward to meet the earth. Flowers initially occur in clusters of 1–6 at intervals along the stem of the inflorescence. Each cluster is at the base of a bract, which ranges from 2–8 cm in length, becoming smaller toward the end of the inflorescence.

Methods

A single live cultivated plant 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.

Results

The genome assembly yielded a total sequence length of 375,177,971 bp over 389,500 scaffolds.

Data availability

Raw and assembled data is publicly available via GenBank:

Raw genome data

https://www.ncbi.nlm.nih.gov/sra/?term=SRR11638255

Assembled genome

https://www.ncbi.nlm.nih.gov/assembly/GCA_025212335

Conflict of Interest Statement

The authors declare they have no conflicts of interests.

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.