






GENOME NOTE

An updated version of the Madagascar periwinkle genome

[version 1; peer review: 2 approved]

Clément Cuello ^{1*}, Emily Amor Stander^{1*}, Hans J. Jansen ²,
Thomas Dugé De Bernonville^{1,3}, Audrey Oudin¹, Caroline Birer Williams¹,
Arnaud Lanoue¹, Nathalie Giglioli Guivarc'h¹, Nicolas Papon⁴, Ron P. Dirks²,
Michael Krogh Jensen⁵, Sarah Ellen O'Connor⁶, Sébastien Besseau¹,
Vincent Courdavault ¹

¹EA2106 Biomolécules et Biotechnologies Végétales, Université de Tours, Tours, 37200, France

²Future Genomics Technologies, Leiden, 2333BE, The Netherlands

³Present address: Centre de Recherche, Limagrain, Chappes, 07745, France

⁴IRF, SFR ICAT, Univ Angers, Univ Brest, Angers, 49000, France

⁵Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, 2800, Denmark

⁶Department of Natural Product Biosynthesis, Max Planck Institute for Chemical Ecology, Jena, 07745, Germany

* Equal contributors

V1 First published: 21 Dec 2022, 11:1541
<https://doi.org/10.12688/f1000research.129212.1>

Latest published: 21 Dec 2022, 11:1541
<https://doi.org/10.12688/f1000research.129212.1>

Abstract



The Madagascar periwinkle, *Catharanthus roseus*, belongs to the *Apocynaceae* family. This medicinal plant, endemic to Madagascar, produces many important drugs including the monoterpene indole alkaloids (MIA) vincristine and vinblastine used to treat cancer worldwide. Here, we provide a new version of the *C. roseus* genome sequence obtained through the combination of Oxford Nanopore Technologies long-reads and Illumina short-reads. This more contiguous assembly consists of 173 scaffolds with a total length of 581.128 Mb and an N50 of 12.241 Mb. Using publicly available RNAseq data, 21,061 protein coding genes were predicted and functionally annotated. A total of 42.87% of the genome was annotated as transposable elements, most of them being long-terminal repeats. Together with the increasing access to MIA-producing plant genomes, this updated version should ease evolutionary studies leading to a better understanding of MIA biosynthetic pathway evolution.


Keywords


Monoterpene indole alkaloids, *Catharanthus roseus*, Apocynaceae

Open Peer Review

Approval Status  

	1	2
version 1		
21 Dec 2022	view	view

1. **Evangelos Tatsis** , Shanghai Institute of Plant Physiology and Ecology, Shanghai, China

2. **Amit Rai** , Chiba University, Chiba, Japan

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Genomics and Genetics gateway**.

Corresponding author: Vincent Courdavault (vincent.courdavault@univ-tours.fr)

Author roles: **Cuello C:** Formal Analysis, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Stander EA:** Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Jansen HJ:** Data Curation, Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Dugé De Bernonville T:** Conceptualization, Investigation, Resources, Writing – Review & Editing; **Oudin A:** Investigation, Writing – Review & Editing; **Birer Williams C:** Investigation, Writing – Review & Editing; **Lanoue A:** Investigation, Writing – Review & Editing; **Giglioli Guivarc'h N:** Investigation, Writing – Review & Editing; **Papon N:** Investigation, Writing – Review & Editing; **Dirks RP:** Data Curation, Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Jensen MK:** Conceptualization, Funding Acquisition, Project Administration, Writing – Review & Editing; **O'Connor SE:** Conceptualization, Writing – Review & Editing; **Besseau S:** Conceptualization, Writing – Review & Editing; **Courdavault V:** Conceptualization, Funding Acquisition, Project Administration, Writing – Review & Editing

Competing interests: Ron P. Dirks and Hans J. Jansen are CEO and CTO of Future Genomics Technologies, respectively.

Grant information: This work was supported by EU Horizon 2020 research and innovation program [MIAMI project, grant number 814645; MK], SEO, VC]; ARD CVL Biopharmaceutical program of the Région Centre-Val de Loire [ETOPOCentre project, VC]; and ANR [project MIACYC – ANR-20-CE43-0010, VC].

Copyright: © 2022 Cuello C *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Cuello C, Stander EA, Jansen HJ *et al.* **An updated version of the Madagascar periwinkle genome [version 1; peer review: 2 approved]** F1000Research 2022, 11:1541 <https://doi.org/10.12688/f1000research.129212.1>

First published: 21 Dec 2022, 11:1541 <https://doi.org/10.12688/f1000research.129212.1>

Introduction

The Madagascar periwinkle, *Catharanthus roseus* (L.) G. Don, is an *Apocynaceae* plant native to Madagascar. *C. roseus* produces several specialized metabolites including monoterpene indole alkaloids (MIA; O'Connor and Maresh, 2006). These molecules are produced by plants to face biotic and abiotic pressures accounting for their wide range of bioactive properties (Dugé de Bernonville *et al.*, 2015). Above all, MIAs produced by *C. roseus* are well-known for being part of the human pharmacopoeia against cancer, such as the well-known vinblastine and vincristine, and other MIA derivatives, including vinorelbine (O'Connor and Maresh, 2006).

Due to its high economic importance, *C. roseus* has extensively been studied within the last three decades becoming the model species for MIA biosynthetic pathway studies (see Pan *et al.*, 2016 and Kulagina *et al.*, 2022 for extensive review). *C. roseus* genome was firstly sequenced in 2015 (Kellner *et al.*, 2015). Recently, a more contiguous version (v2) was generated to ease inter-species genomic comparison (Franke *et al.*, 2019). To date, *C. roseus* genome sequencing and assembly did not benefit from the development of third generation sequencing technologies that lead to more contiguous genome (Jiao and Schneeberger, 2017). Thanks to these new technologies, we present here an even more contiguous genome assembly. This updated version (v2.1) should ease inter-species studies in order to better understand the diversification of MIAs and the evolution of their biosynthetic pathways.

Methods

Sample collection, DNA extraction and sequencing

C. roseus cv 'SunStorm[®] Apricot' seeds (variety ID: 70001114, Syngenta flowers, Basel, Switzerland) were greenhouse-grown at the University of Tours for 1 month before sampling. DNA was extracted from *C. roseus* leaves using Qiagen Plant DNeasy kit (ID: 69204, Qiagen, Hilden, Germany) following the manufacturer's instructions. Illumina sequencing library were constructed using the TruSeq DNA PCR-free kit (ID: 20015962, Illumina, San Diego, USA) and sequenced in paired-end mode (2 × 150 bp) by Eurofins Genomics (Les Ulis, France) using Illumina NextSeq500 technology. Future Genomics Technologies (Leiden, The Netherlands) constructed ONT library using ONT 1D ligation sequencing kit (SQK-LSK109, Oxford Nanopore Technologies Ltd, Oxford, United-Kingdom) subsequently sequenced on Nanopore GridION flowcell and Nanopore PromethION flowcell (Oxford Nanopore Technologies Ltd, Oxford, United-Kingdom) with the GuPPy (RRID:SCR_022353) version 3.2.6 high-accuracy basecaller. A total of 114,329,683 paired-end reads were obtained from the Illumina HiSeq sequencing, 908,999 and 2,588,997 from the ONT GridION and ONT PromethION sequencing, respectively.

De novo genome assembly

The *C. roseus* genome was assembled by Future Genomics Technologies (Leiden, The Netherlands). After adapters removal using Porechop (RRID:SCR_016967) (Wick *et al.*, 2017), ONT reads were first assembled into contig using Flye (RRID:SCR_017016) assembler (v.2.5, Kolmogorov *et al.*, 2019) with the following options: `-min-overlap 10000 -i 2`. Redundant contigs were removed using Purge_haplotigs (RRID:SCR_017616) (v.1.1.0) followed by two rounds of polishing with Illumina paired-end reads using Pilon (RRID:SCR_014731) (v.1.23, Walker *et al.*, 2014).

Gene model prediction and gene functional annotation

RNA-seq data were retrieved from the NCBI Sequence Read Archive (SRA) (RRID:SCR_004891) database using the following accession numbers: ERS1229288, ERS1229289, ERS1229290, ERS1229291, ERS1229292, ERS1229293, ERS1229294, ERS1229295, ERS1229296, ERS1907920, ERS2396963, ERS2396964, ERS2396965, ERS2396966, SRR20661631. These data were individually aligned to the *C. roseus* genome using HISAT2 (RRID:SCR_015530) (v.2.2.1, Kim *et al.*, 2019). Transcripts were subsequently assembled using the resulting RNA-seq alignments and StringTie (RRID:SCR_016323) (v.2.1.7, Pertea *et al.*, 2015). These individual transcriptomes were further merged using stringtie-merge to a non-redundant set of transcripts. A combination of similarity search using BLASTX (RRID:SCR_001653) and BLASTP (v.2.6.0-1, Camacho *et al.* 2009) against UniProt (RRID:SCR_002380) database (v.2022-10-12) and hmmscan (v.3.1b2, Finn *et al.*, 2011) against the Pfam (RRID:SCR_004726) database was used to assign putative function to each gene model.

Assembly completeness assessment

The stat program from BBmap (RRID:SCR_016965) tool (v.38.94, Bushnell, 2014) was used to assess assembly quality. Benchmarking Universal Single-Copy Orthologs (BUSCO v.5.2.2, Simão *et al.*, 2015) (RRID:SCR_015008) with default settings was used to assess genome and gene models completeness using a plant-specific database of 2,326 single copy orthologs (eudicots_odb10). The agat_sp_statistics perl script from the AGAT package (v.0.8.0, Dainat *et al.*, 2022) was used to get the gene models statistics.

Transposable elements (TE) prediction and annotation

Identification and annotation of transposable elements was determined using extensive *de novo* TE annotator (EDTA v.1.9.5, [Ou et al., 2019](#)) (RRID:SCR_022063) using the sensitive mode. This pipeline annotates long-terminal repeat (LTR) using [LTR_Finder](#) (RRID:SCR_015247) (v. 1.07, [Xu and Wang, 2007](#)) and LTRharvest (RRID:SCR_018970) included in [GenomeTools](#) (RRID:SCR_016120) (v.1.5.10, [Ellinghaus et al., 2008](#)); terminal inverted repeat (TIR) using Generic repeat finder (v.1.0, [Shi and Liang, 2019](#)) and TIR-learner (v.2.5, [Su et al., 2019](#)); and Helitrons using HelitronScanner (v.1.1, [Xiong et al., 2014](#)). TE size thresholds are further used to prevent false discoveries. Hence, TIR shorter than 80 bp as well as LTR and Helitrons shorter than 100 bp are considered as tandem repeats and short sequences. To prevent false LTR discoveries, LTR are further filtered using [LTR_retriever](#) (RRID:SCR_017623) (v.2.9.0, [Ou and Jiang, 2018](#)). TIR candidates are classified as MITEs if not exceeding 600 bp. TIR and Helitrons are further filtered using EDTA advanced filters (see [Ou et al., 2019](#) for details). The genome is then masked using the obtained TE library. Unmasked part of the genome is then scanned by [RepeatModeler](#) (RRID:SCR_015027) (v.2.0.1, default parameters, [Flynn et al., 2020](#)) to identify non-LTR retrotransposons and unclassified TE missed by structure-based TE identification tools. Finally, EDTA uses the provided CDS sequences to remove gene-related sequences.

Results

Genome assembly

C. roseus genome was assembled from ONT long-reads using Flye (v.2.5) resulting in a 651.9 Mb assembly distributed across 788 contigs. This assembly was collapsed using `purge_haplotigs` into 173 scaffolds reducing length to 585,8 Mb but increasing N50 from 10.3 Mb to 12.3 Mb. Assembly polishing was performed twice using Illumina short-reads with `pilon` (v. 1.23). *C. roseus* final assembly consisted in 173 scaffolds with a total length of 581.45 Mb. Even though *C. roseus* v.2.1 displayed similar BUSCO scores compared to *C. roseus* v.2 based on *Eudicotyledons* Benchmarking Universal Single-Copy Orthologs (BUSCO), this new version v.2.1 turns out to be much more contiguous with a 12 time less contigs and a six-fold larger N50 ([Table 1](#)) ([Cuello et al., 2022](#)).

Gene annotation

RNA-seq based gene model prediction using publicly available data resulted in a total of 21,061 genes. Despite less genes were annotated; a higher BUSCO score was obtained ([Figure 1](#)). The combination of BLASTP and BLASTX against

Table 1. Genome assembly metrics.

Version	Assembly size (Mb)	No. of scaff. ^a	N50 (Mb)	BUSCO scores (genome mode) C [S; D]; F; M ^b	Protein coding genes	Ref.
<i>C. roseus</i> v.2	541.13	2,090	2.58	97.0 [95.5; 1.5]; 1.3; 1.7	34,363	Franke et al., 2019
<i>C. roseus</i> v.2.1	581.45	173	12.2	97.1 [94.2; 2.9]; 1.0; 1.9	21,061	This study

C. roseus: *Catharanthus roseus*; BUSCO: Benchmarking Universal Single-Copy Orthologs.

^aNumber of scaffolds.

^bBUSCO scores (genome mode) % Complete [% Complete and single-copy; % Complete and Duplicated]; % Fragmented; % Missing (n = 2,326).

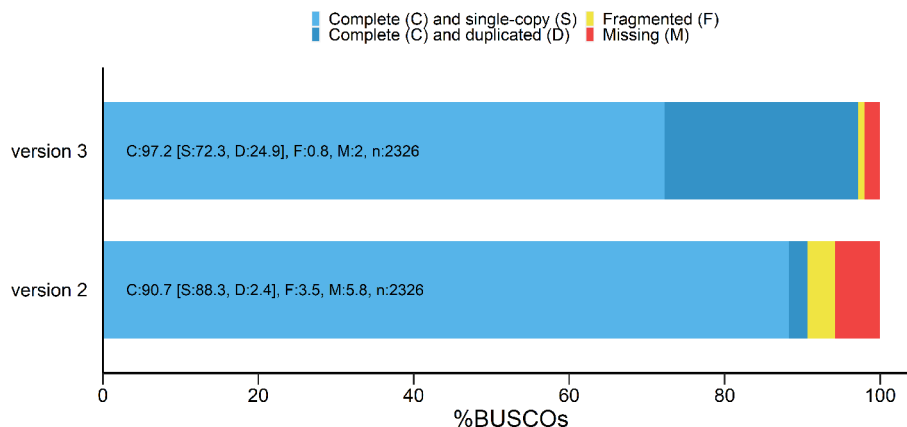


Figure 1. BUSCO scores of the predicted gene set. BUSCO: Benchmarking Universal Single-Copy Orthologs.

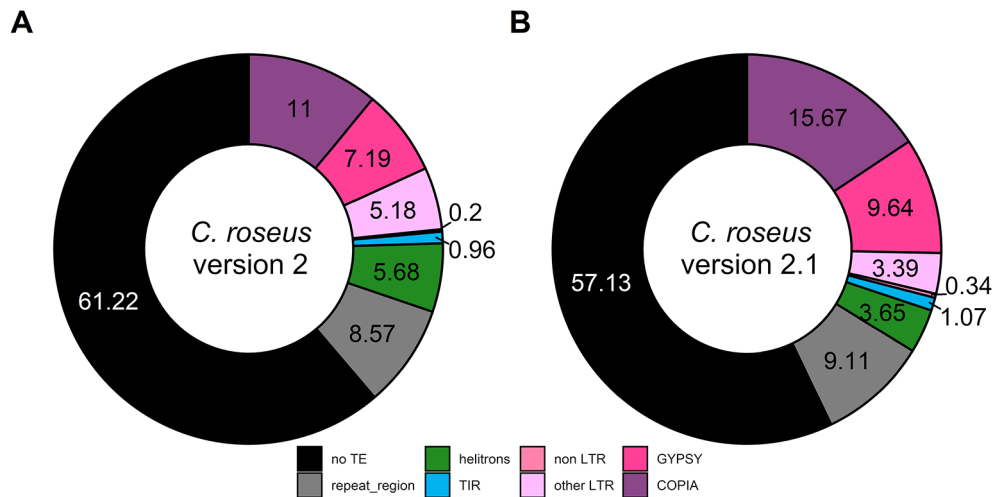


Figure 2. Proportion of transposable element (TE) in *C. roseus* assembly version 2 (A) and version 2.1 (B). TIR: terminal inverted repeat, LTR: long terminal repeat, non LTR: retrotransposons without LTR sequence, other LTR: LTR containing retrotransposons except for *Gypsy* and *Copia*.

UniProt database and hmmscan against the PFAM database led to the functional annotation of 76.5% of the predicted genes (16,118 of the 21,062 genes, Supplementary Table S1 in *Underlying data* (Cuello *et al.*, 2022)). All functionally validated MIA biosynthetic genes from *C. roseus* could be found in this new version v.2.1 of the genome with identity and coverage percentage ranging from 95 to 100% and 94 to 100%, respectively, with the exception of *G10H* and *DAT* (Supplementary Table S2-S3 in *Underlying data* (Cuello *et al.*, 2022)).

Transposable element annotation

Finally, we analyzed TE composition of this updated *C. roseus* genome. While 38.78% of the genome consisted in TE in *C. roseus* v.2, a higher proportion (42.87%) was annotated as TE in this new version (v.2.1) with similar distribution across the different TE families (Figure 2). It is worth noting that TE proportion of this v.2.1 is closer to the one in its recently sequenced closely related species *Vinca minor* (Stander *et al.*, 2022).

Data availability

Underlying data

BioProject: *Catharanthus roseus* genome sequencing. Raw sequence reads, complete genome. Accession number PRJNA907167, <https://identifiers.org/NCBI/bioproject:PRJNA907167> (Tours University, 2022a).

BioSample: Plant sample from *Catharanthus roseus*, Accession number SAMN31953452, <https://identifiers.org/NCBI/biosample:SAMN31953452> (Tours University, 2022b).

Figshare: An updated version of *Catharanthus roseus* genome. 10.6084/m9.figshare.21641111 (Cuello *et al.*, 2022). This project contains the following underlying data:

- Catharanthus_roseus_v2.1_UT.cds (Predicted CDS).
- Catharanthus_roseus_v2.1_UT.gff (Genome annotation file (GFF)).
- Catharanthus_roseus_v2.1_UT.pep (Predicted proteins).
- Catharanthus_roseus_v2.1_UT.tr (Predicted transcripts).
- Cuello et al – F1000R – SuppMat.xlsx (Supplementary tables).

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Acknowledgments

The authors benefitted from the use of the cluster at the Centre de Calcul Scientifique en région Centre-Val de Loire.

References

- Bushnell B: *BBMap: A Fast, Accurate, Splice-Aware Aligner* (No. LBNL-7065E). Berkeley, CA (United States): Lawrence Berkeley National Lab. (LBNL); 2014.
- Camacho C, Coulouris G, Avagyan V, *et al.*: **BLAST+: architecture and applications.** *BMC Bioinformatics.* 2009; **10**: 421.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cuello C, Stander E, Jansen HJ, *et al.*: An updated version of the Madagascar periwinkle genome. figshare. [Dataset]. 2022.
[Publisher Full Text](#)
- Dainat J, Hereñú DLucileSol, pascal-git: **NBISweden/AGAT: AGAT-v0.8.1.** *Zenodo.* 2022.
[Publisher Full Text](#)
- Dugé de Bernonville T, Clastre M, Besseau S, *et al.*: **Phytochemical genomics of the Madagascar periwinkle: Unravelling the last twists of the alkaloid engine.** *Phytochemistry.* 2015; **113**: 9–23.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ellinghaus D, Kurtz S, Willhoeft U: **LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons.** *BMC Bioinformatics.* 2008; **9**(1): 18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Res.* 2011; **39**: W29–W37.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Flynn JM, Hubbley R, Goubert C, *et al.*: **RepeatModeler2 for automated genomic discovery of transposable element families.** *PNAS.* 2020; **117**(17): 9451–9457.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Franke J, Kim J, Hamilton JP, *et al.*: **Gene Discovery in *Gelsemium* Highlights Conserved Gene Clusters in Monoterpene Indole Alkaloid Biosynthesis.** *ChemBioChem.* 2019; **20**: 83–87.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jiao WB, Schneeberger K: **The impact of third generation genomic technologies on plant genome assembly.** *Curr. Opin. Plant Biol.* 2017; **36**: 64–70.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kellner F, Kim J, Clavijo BJ, *et al.*: **Genome-guided investigation of plant natural product biosynthesis.** *Plant J.* 2015; **82**: 680–692.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kim D, Paggi JM, Park C, *et al.*: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.** *Nat. Biotechnol.* 2019; **37**(8): 907–915.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kolmogorov M, Yuan J, Lin Y, *et al.*: **Assembly of long, error-prone reads using repeat graphs.** *Nat. Biotechnol.* 2019; **37**: 540–546.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kulagina N, Méteignier LV, Papon N, *et al.*: **More than a Catharanthus plant: A multicellular and pluri-organelle alkaloid-producing factory.** *Curr. Opin. Plant Biol.* 2022; **67**: 102200.
[PubMed Abstract](#) | [Publisher Full Text](#)
- O'Connor SE, Maresh JJ: **Chemistry and biology of monoterpene indole alkaloid biosynthesis.** *Nat. Prod. Rep.* 2006; **23**: 532–547.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ou S, Jiang N: **LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons.** *Plant Physiol.* 2018; **176**(2): 1410–1422.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ou S, Su W, Liao Y, *et al.*: **Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline.** *Genome Biol.* 2019; **20**: 275.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pan Q, Mustafa NR, Tang K, *et al.*: **Monoterpenoid indole alkaloids biosynthesis and its regulation in *Catharanthus roseus*: a literature review from genes to metabolites.** *Phytochem. Rev.* 2016; **15**: 221–250.
[Publisher Full Text](#)
- Pertea M, Pertea GM, Antonescu CM, *et al.*: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.** *Nat. Biotechnol.* 2015; **33**(3): 290–295.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shi J, Liang C: **Generic repeat finder: a high-sensitivity tool for genome-wide de novo repeat detection.** *Plant Physiol.* 2019; **180**(4): 1803–1815.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**: 3210–3212.
[Publisher Full Text](#)
- Stander EA, Cuello C, Birer-Williams C, *et al.*: **The Vinca minor genome highlights conserved evolutionary traits in monoterpene indole alkaloid synthesis.** *G3 Genes | Genomes | Genetics.* 2022; **12**: jkac268.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Su W, Gu X, Peterson T: **TIR-learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome.** *Mol. Plant.* 2019; **12**(3): 447–460.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tours University: *Catharanthus roseus* genome. [Dataset]. *BioProject.* 2022a.
[Reference Source](#)
- Tours University: Plant sample from *Catharanthus roseus*. [Dataset]. *BioSample.* 2022b.
[Reference Source](#)
- Walker BJ, Abeel T, Shea T, *et al.*: **Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.** *PLoS One.* 2014; **9**: e112963–944 e112963.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wick RR, Judd LM, Gorrie CL, *et al.*: **Completing bacterial genome assemblies with multiplex MinION sequencing.** *Microb. Genom.* 2017; **3**(10): e000132.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Xiong W, He L, Lai J, *et al.*: **HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes.** *Proc. Natl. Acad. Sci. USA.* 2014; **111**(28): 10263–10268.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Xu Z, Wang H: **LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons.** *Nucleic Acids Res.* 2007; **35**(Web Server issue): W265–W268.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 06 February 2023

<https://doi.org/10.5256/f1000research.141880.r158636>

© 2023 Rai A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Amit Rai 

Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan

In the presented article, authors reported an updated version of genome assembly for Madagascar periwinkle, which is a valuable model plant species to study MIA biosynthesis. Compared to the previously published genome assemblies for Madagascar periwinkle, this study used long read sequencing technology and achieved an improvement in terms of contig N50.

Without a doubt, this is a better genome assembly, but authors should have considered scaffolding through HiC to achieve a chromosome-scale genome assembly as that would have allowed them to discover novel features contributing MIA biosynthesis and evolution.

Nevertheless, the resource presented here is valuable, and will inspire researchers to combine the generated datasets in this study with new sequencing data to derive a chromosome-scale genome assembly for *C. roseus*. For these reasons, I support its indexing.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome sciences, Metabolomics research. Evolutionary biology, MIA biosynthesis pathways

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 04 January 2023

<https://doi.org/10.5256/f1000research.141880.r158640>

© 2023 Tatsis E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Evangelos Tatsis 

Chinese Academy of Sciences Centre for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai, China

The current work reports an updated version of the significant medicinal plant *Catharanthus roseus*. *C. roseus* is the only source of the clinically used anticancer drug vinblastine and such work can provide resources for molecular breeding to improve the production of vinblastine and other MIAs.

The sequencing techniques and the bioinformatic analysis are appropriate. I recommend the current manuscript for indexing as it is.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant specialised metabolism

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research