

DATA NOTE

Open Access



# Chromosome-scale assembly of the Verbenaceae species Queen's Wreath (*Petrea volubilis* L.)

John P. Hamilton<sup>1,2\*</sup>, Brieanne Vaillancourt<sup>1</sup>, Joshua C. Wood<sup>1</sup> and C. Robin Buell<sup>1,2,3\*</sup>

## Abstract

**Objectives** *Petrea volubilis*, a member of the Order Lamiales and the Verbenaceae family, is an important horticultural species that has been used in traditional folk medicine. To provide a genome sequence for comparative studies within the Order Lamiales that includes important families such as Lamiaceae (mints), we generated a long-read, chromosome-scale genome assembly of this species.

**Data description** Using a total of 45.5 Gb of Pacific Biosciences long read sequence, we generated a 480.2 Mb assembly of *P. volubilis*, of which, 93% is chromosome anchored. Representation of genic regions was robust with 96.6% of the Benchmarking of Universal Single Copy Orthologs present in the genome assembly. A total of 57.8% of the genome was annotated as a repetitive sequence. Using a gene annotation pipeline that included refinement of gene models using transcript evidence, 30,982 high confidence genes were annotated. Access to the *P. volubilis* genome will facilitate evolutionary studies in the Lamiales, a key order of Asterids that includes significant crop and medicinal plant species.

**Keywords** Queen's wreath, Genome assembly, Lamiales, Comparative genomics, *Petrea volubilis*

## Objective

The Asterid species, *Petrea volubilis* L., also known as Queen's Wreath, Purple Wreath, Bluebird vine or Sandpiper vine, is a member of the Verbenaceae family within the Order Lamiales. As a perennial woody vine, *P. volubilis* is a key ornamental species due to its intense violet flowers. Historically, leaves of *P. volubilis* have been used

in Mexico as folk medicine to remedy kidney stones, rheumatism, diarrhea, and urinary infections [1] and as an abortifacient in Jamaica [2]. *P. volubilis* extracts have been found to have antipyretic, analgesic, and antimicrobial [3, 4] and insecticidal activities [4]. Recently, *P. volubilis* was included as one of four outgroup species in a study that revealed the evolutionary basis of chemical diversity in the Lamiaceae [5]. In this project, we sequenced and annotated the *P. volubilis* genome to facilitate our understanding of genome and chemodiversity evolution within the Lamiales.

## Data description

High molecular weight DNA was isolated using a modified cetyl trimethylammonium bromide method (2% CTAB, 100 mM Tris, 1.4 M Sodium Chloride, 20 mM EDTA) [6] followed by RNase treatment and cleanup using the DNeasy PowerClean Pro Cleanup Kit (Qiagen). Pacific

\*Correspondence:

John P. Hamilton  
John.Hamilton@uga.edu  
C. Robin Buell  
Robin.Buell@uga.edu

<sup>1</sup> Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, USA

<sup>2</sup> Department of Crop & Soil Sciences, University of Georgia, Athens, GA 30602, USA

<sup>3</sup> Institute of Plant Breeding, Genetics, & Genomics, University of Georgia, Athens, GA 30602, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Table 1** Overview of data files and data sets used in this study

| Label       | Data file/Data set name   | File types                     | Data repository and identifier (DOI or accession number)   |
|-------------|---|--------------------------------|--|
| Data file 1 | <i>Petrea volubilis</i> libraries used in this study                            | Spreadsheet (.xlsx)            | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 2 | Genomescope k-mer frequency distribution plot                                   | Portable Document Files (.pdf) | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 3 | Hi-C contact map  | Portable Document Files (.pdf) | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 4 | Assembly metrics for the <i>Petrea volubilis</i> assembly                       | Spreadsheet (.xlsx)            | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 5 | Pseudomolecule lengths and gap content for the <i>Petrea volubilis</i> assembly | Spreadsheet (.xlsx)            | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 6 | KAT k-mer comparison plot   | Portable Document Files (.pdf) | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 7 | BUSCO results on the <i>Petrea volubilis</i> assembly and annotation            | Spreadsheet (.xlsx)            | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 8 | Repetitive sequence content in the <i>Petrea volubilis</i> assembly             | Spreadsheet (.xlsx)            | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data file 9 | <i>Petrea volubilis</i> gene annotation summary                                 | Spreadsheet (.xlsx)            | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 1  | Pac Bio reads from high molecular weight DNA, SRR11516643                       | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR11516643">https://identifiers.org/ncbi/insdc.sra:SRR11516643</a> ) [7, 21] |
| Data set 2  | Pac Bio reads from high molecular weight DNA, SRR11516644                       | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR11516644">https://identifiers.org/ncbi/insdc.sra:SRR11516644</a> ) [7, 22] |
| Data set 3  | Illumina whole genome shotgun reads, SRR11516645                                | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR11516645">https://identifiers.org/ncbi/insdc.sra:SRR11516645</a> ) [7, 11] |
| Data set 4  | Illumina Hi-C DNA sequence reads, SRR15904679                                   | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR15904679">https://identifiers.org/ncbi/insdc.sra:SRR15904679</a> ) [7, 15] |
| Data set 5  | Illumina Hi-C DNA sequence reads, SRR15904680                                   | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR15904680">https://identifiers.org/ncbi/insdc.sra:SRR15904680</a> ) [7, 16] |
| Data set 6  | Illumina RNA-Seq—Root, SRR8937863   | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR8937863">https://identifiers.org/ncbi/insdc.sra:SRR8937863</a> ) [7, 23]   |
| Data set 7  | Illumina RNA-Seq—Petiole, SRR8937861  | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR8937861">https://identifiers.org/ncbi/insdc.sra:SRR8937861</a> ) [7, 24]   |
| Data set 8  | Illumina RNA-Seq—Stem, SRR8937862   | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR8937862">https://identifiers.org/ncbi/insdc.sra:SRR8937862</a> ) [7, 25]   |
| Data set 9  | Illumina RNA-Seq—Immature leaf, SRR8937859                                      | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR8937859">https://identifiers.org/ncbi/insdc.sra:SRR8937859</a> ) [7, 26]   |
| Data set 10 | Illumina RNA-Seq—Mature leaf, SRR8937860  | Fastq file (.fastq.gz)         | NCBI ( <a href="https://identifiers.org/ncbi/insdc.sra:SRR8937860">https://identifiers.org/ncbi/insdc.sra:SRR8937860</a> ) [7, 27]   |
| Data set 11 | Genome assembly of <i>Petrea volubilis</i>                                      | fasta file (.fa)               | NCBI ( <a href="https://identifiers.org/assembly:GCA_026212405.1">https://identifiers.org/assembly:GCA_026212405.1</a> ) [7, 28]     |
| Data set 12 | High Confidence <i>Petrea volubilis</i> Gene Models cDNA                        | fasta file (.fa)               | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 13 | High Confidence <i>Petrea volubilis</i> Gene Models CDS                         | fasta file (.fa)               | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 14 | High Confidence <i>Petrea volubilis</i> Gene Models GFF3                        | GFF3 file (.gff3)              | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 15 | High Confidence <i>Petrea volubilis</i> Gene Models Proteins                    | fasta file (.fa)               | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 16 | Representative High Confidence <i>Petrea volubilis</i> Gene Models cDNA         | fasta file (.fa)               | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 17 | Representative High Confidence <i>Petrea volubilis</i> Gene Models CDS          | fasta file (.fa)               | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 18 | Representative High Confidence <i>Petrea volubilis</i> Gene Models GFF3         | GFF3 file (.gff3)              | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |
| Data set 19 | Representative High Confidence <i>Petrea volubilis</i> Gene Models List         | text file (.txt)               | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7]       |

**Table 1** (continued)

| Label       | Data file/Data set name   | File types        | Data repository and identifier (DOI or accession number)   |
|-------------|---|-------------------|--|
| Data set 20 | Representative High Confidence <i>Petrea volubilis</i> Gene Models Proteins | fasta file (.fa)  | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7] |
| Data set 21 | <i>Petrea volubilis</i> Working Gene Models cDNA                            | fasta file (.fa)  | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7] |
| Data set 22 | <i>Petrea volubilis</i> Working Gene Models CDS                             | fasta file (.fa)  | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7] |
| Data set 23 | <i>Petrea volubilis</i> Working Gene Models GFF3                            | GFF3 file (.gff3) | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7] |
| Data set 24 | <i>Petrea volubilis</i> Working Gene Models Proteins                        | fasta file (.fa)  | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7] |
| Data set 25 | <i>Petrea volubilis</i> Working Gene Models Functional Annotation           | text file (.txt)  | Figshare ( <a href="https://doi.org/10.6084/m9.figshare.21429219.v3">https://doi.org/10.6084/m9.figshare.21429219.v3</a> ) [7] |

Biosciences (PacBio) SMRTbell Express Template libraries were constructed and sequenced on a PacBio Sequel instrument generating 45.5 Gb of total sequence (Table 1, Data file 1, Data sets 1 & 2, [7]). Reads less than 5 kb were filtered out and the remaining reads were assembled using Canu v1.8 [8] with the options: minOverlapLength=2000 minReadLength=5000 genomeSize=450 m resulting in an initial assembly of 630.0 Mb with 6,515 contigs and an N50 contig length of 369,179 bp. The genome was polished with two rounds of GCpp (v1.9.0) [9], followed by three rounds of polishing with Pilon (v1.23) [10] using Illumina whole genome shotgun reads (Table 1, Data file 1, Data set 3, [7, 11]). A k-mer distribution plot using GenomeScope [12] revealed the genome was heterozygous (Table 1, Data file 2, Data set 3, [7]). Haplotigs were removed using two rounds of purge\_dups using the default parameters (v1.0.0) [13, 14] and Hi-C libraries constructed by Phase Genomics (Table 1, Data file 1, Data sets 4 & 5, [7, 15, 16]) were used to place the final scaffolds into 17 chromosomes using the Juicer (v1.6)/3D-DNA pipeline (git commit: 529ccf4; Table 1, Data file 3) [7, 17, 18]. The final assembly size is 480.2 Mb (478.8 Mb ungapped, 93% chromosome-anchored), consistent with the size estimated by flow cytometry of 455 Mb per 1C [5] (Table 1, Data files 4 & 5, [7]). A comparison of k-mers in the Illumina whole genome shotgun reads vs the genome assembly using KAT (v2.4.1) [19] with a k-mer size of 21 revealed that *P. volubilis* is heterozygous (estimated heterozygosity rate 1.45%) and the assembly is near-complete (estimated completeness, 98.8%;(Table 1, Data file 6, [7]). The majority of k-mers in the reads are present in one copy indicating the haplotigs were successfully purged from the final assembly (Table 1, Data files 1 & 6, Data set 3, [7]). Assessment of representation of genic regions using the Benchmarking of Universal Single Copy Orthologs [20] (BUSCO; v5.4.3 with embryophyta\_odb10) revealed 96.6% of the BUSCO genes

present in the genome assembly (Table 1, Data file 7, [7]). While the scaffold N50 was 25.6 Mb, the contig N50 was 0.53 Mb due potentially to heterozygosity that reduced the ability of the assembler to generate longer contigs (Table 1, Data file 6, [7]; see Limitations).

The *P. volubilis* genome was annotated as described previously [29]. In brief, repetitive sequences were identified in the unscaffolded contigs using RepeatModeler (v2.0.1) [30] and protein-coding genes removed from the library using ProtExcluder (v1.2) [31]. The custom repetitive sequences were then added to the Repbase Viridiplantae repeats (v20150807) [32] and used to mask repeats using RepeatMasker (v4.1.0) [30] with the parameters -s -nolow -no\_is -gff (Table 1, Data file 8, [7]); 57.8% of the genome was masked. RNA-seq reads from five libraries (Table 1, Data file 1, Data sets 6, 7, 8, 9, & 10, [7, 23–27]) were cleaned with Cutadapt (v2.9) [33] using a quality cutoff of 10 and a minimum length 100 nt and then aligned using HISAT2 (v2.2.0) [34] with a maximum intron length of 5000 bp. Gene predictions were generated with BRAKER2 (v2.1.5) [35] using the RNA-seq alignments as hints. Final gene models were refined using the RNA-seq transcript assemblies using two rounds of PASA2 (v2.4.1) [36, 37] and genome-guided transcript assemblies created from the RNA-seq alignments using Stringtie (v2.1.1) [38]. Gene models were annotated using alignments to the predicted *Arabidopsis thaliana* proteome, Pfam database, and transcript evidence as described previously [29]; a total of 49,169 high confidence models (30,982 genes) within the 56,052 working models (37,610 genes) were annotated (Table 1, Data file 9, [7]). High confidence models within the working model set were defined by either protein evidence (alignment to Arabidopsis or Pfam domain and/or expression evidence (TPM > 0)). Representative models, both working and high confidence, were defined as

the model for each locus (gene) with the longest CDS. BUSCO assessments (v5.4.3 and embryophyta\_odb10) of the annotation revealed 89.9% and 88.5% of BUSCO genes in the working gene model and representative high confidence gene model set, respectively (Table 1, Data file 7, [7]). The final genome annotation was transferred from the scaffolds to the chromosomes using Liftoff (v1.6.3) [39] with the parameters -a 0.9 -s 0.95 -exclude\_partial -cds -polish.

## Limitations

*Petrea volubilis* is heterozygous and we purged haplotigs in the assembly process. This likely contributed to the reduced N50 contig size (0.53 Mb) and the slightly larger assembly size (480.2 Mb) compared to the estimated genome size from flow cytometry (445 Mb). However, based on BUSCO scores, a mere 4.3% of the orthologs were duplicated in the assembly suggestive that we removed the majority of alternative haplotigs. Future efforts using near-perfect long genomic reads such as PacBio HiFi or Oxford Nanopore Technologies Q20+ platforms would permit a haplotype-resolved genome assembly.

## Abbreviations

BUSCO Benchmarking Universal Single Copy Orthologs  
PacBio Pacific BioSciences

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-023-01110-z>.

**Additional file 1: Data file 1.** *Petrea volubilis* libraries used in this study. **Data file 2.** Genomescope k-mer frequency distribution plot. **Data file 3.** Hi-C contact map. **Data file 4.** Assembly metrics for the *Petrea volubilis* assembly. **Data file 5.** Pseudomolecule lengths and gap content for the *Petrea volubilis* assembly. **Data file 6.** KAT k-mer comparison plot. **Data file 7.** Benchmarking universal single copy orthologs (BUSCO) results on the *Petrea volubilis* assembly and annotation. **Data file 8.** Repetitive sequence content in the *Petrea volubilis* assembly. **Data file 9.** *Petrea volubilis* gene annotation summary.

## Acknowledgements

We acknowledge the efforts of Dr. Dongyan Zhao in preliminary genome assembly efforts of the genome. We acknowledge the sequencing performed at the Michigan State University Research Technology Support Facility and the University of Georgia Genomics and Bioinformatics Core. We thank Pamela and Doug Soltis of the University of Florida for providing a *Petrea volubilis* plant.

## Authors' contributions

B.V. and J.C.W. generated sequence, performed quality assessments, and performed data management. J.P.H. assembled and annotated the genome. J.P.H. and C.R.B. wrote the manuscript. C.R.B. conceived of the study and obtained project funding. All authors approved the manuscript.

## Funding

Funding for this work was provided via grants to CRB from the National Science Foundation (IOS-1444499), the Georgia Research Alliance, and the

University of Georgia. The funders had no role in the design, execution, interpretation, or written summary of this study.

## Availability of data and materials

All raw sequence data is available in the National Center for Biotechnology Information under BioProject ID PRJNA534065 (<https://identifiers.org/bioproject:PRJNA534065>; [11, 15, 16, 21–27]). The assembled genome is available in Genbank under the accession JAOWBU0000000000 ([https://identifiers.org/assembly:GCA\\_026212405.1](https://identifiers.org/assembly:GCA_026212405.1); [28]) and in Figshare (<https://doi.org/10.6084/m9.figshare.21429219.v3>, [7]). A summary of data sets is available in Table 1 and are available on Figshare (<https://doi.org/10.6084/m9.figshare.21429219.v3>, [7]).

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 4 November 2022 Accepted: 1 February 2023

Published online: 03 March 2023

## References

- Josabad Alonso-Castro A, Jose Maldonado-Miranda J, Zarate-Martinez A, Jacobo-Salcedo MDR, Fernández-Galicia C, Alejandro Figueroa-Zuñiga L, et al. Medicinal plants used in the Huasteca Potosina, México J Ethnopharmacol. 2012;143:292–8.
- Mitchell SA, Ahmad MH. A review of medicinal plant research at the University of the West Indies, Jamaica, 1948–2001. West Indian Med J. 2006;55:243–69.
- Abdelwahab M, Abdel-Lateff A, Fouad M, Desoukey S, Kamel M. Phytochemical and biological study of *Petrea volubilis* L. (Verbenaceae). Bull Pharm Sci. 2011;34:9–20.
- El-Hela AA, Al-Amier H, Craker LE. Phytochemical and Biological Investigation of Bluebird Vine (*Petrea volubilis*). Planta Med. 2009;75:P-56.
- Mint Evolutionary Genomics Consortium. Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. Mol Plant. 2018;11:1084–96.
- Doyle JJ, Doyle LJ. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.
- Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Data files and Data sets for Hamilton et al. "Chromosome-scale assembly of the Verbenaceae species Queen's Wreath (*Petrea volubilis* L.)." 2023. <https://doi.org/10.6084/m9.figshare.21429219.v3>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–36.
- GCp. 2022. <https://github.com/PacificBiosciences/gcpp>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9:e112963.
- Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina whole genome shotgun reads, SRR11516645. Illumina whole genome shotgun reads, SRR11516645. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR11516645>.
- Vurtture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33:2202–4.
- purge\_dups. 2022. [https://github.com/dfguan/purge\\_dups](https://github.com/dfguan/purge_dups).

14. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 2020;36:2896–8.
15. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina Hi-C DNA sequence reads, SRR15904679. Illumina Hi-C DNA sequence reads, SRR15904679. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR15904679>.
16. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina Hi-C DNA sequence reads, SRR15904680. Illumina Hi-C DNA sequence reads, SRR15904680. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR15904680>.
17. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356:92–5.
18. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. 2016;3:95–8.
19. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*. 2017;33:574–6.
20. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol*. 2018;35:543–8.
21. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Pac Bio reads from high molecular weight DNA, SRR11516643. Pac Bio reads from high molecular weight DNA, SRR11516643. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR11516643>.
22. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Pac Bio reads from high molecular weight DNA, SRR11516644. Pac Bio reads from high molecular weight DNA, SRR11516644. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR11516644>.
23. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina RNA-Seq - Root, SRR8937863. Illumina RNA-Seq - Root, SRR8937863. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR8937863>.
24. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina RNA-Seq - Petiole, SRR8937861. Illumina RNA-Seq - Petiole, SRR8937861. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR8937861>.
25. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina RNA-Seq - Stem, SRR8937862. Illumina RNA-Seq - Stem, SRR8937862. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR8937862>.
26. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina RNA-Seq - Immature leaf, SRR8937859. Illumina RNA-Seq - Immature leaf, SRR8937859. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR8937859>.
27. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Illumina RNA-Seq - Mature leaf, SRR8937860. Illumina RNA-Seq - Mature leaf, SRR8937860. 2023. <https://identifiers.org/ncbi/insdc.sra:SRR8937860>.
28. Hamilton JP, Vaillancourt B, Wood JC, Buell CR. Chromosome-scale assembly of the Verbenaceae species Queen's Wreath (*Petrea volubilis* L.) Genome Assembly. *Petrea volubilis* L. genome assembly. 2023. [https://identifiers.org/assembly:GCA\\_026212405.1](https://identifiers.org/assembly:GCA_026212405.1).
29. Pham GM, Hamilton JP, Wood JC, Burke JT, Zhao H, Vaillancourt B, et al. Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience*. 2020;9:gjaa100.
30. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020;117:9451–7.
31. Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, et al. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol*. 2014;164:513–24.
32. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6:11.
33. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10–2.
34. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–15.
35. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-Genome Annotation with BRAKER. In: Kollmar M, editor. *Gene Prediction: Methods and Protocols*. Springer, New York: New York, NY; 2019. p. 65–95.
36. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 2003;31:5654–66.
37. Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics*. 2006;7:327.
38. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. 2019;20:278.
39. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2020;37:1639–43.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

