



## OPEN ACCESS

## EDITED BY

Nikolai Borisjuk,  
Huaiyin Normal University, China

## REVIEWED BY

Hoang Dang Khoa Do,  
Nguyen Tat Thanh University, Vietnam  
Pedro Paulo Gattai Gomes,  
University of São Paulo, Brazil

## \*CORRESPONDENCE

Salvatore Tomasello  
✉ [salvatore.tomasello@uni-goettingen.de](mailto:salvatore.tomasello@uni-goettingen.de)

RECEIVED 08 May 2024

ACCEPTED 16 August 2024

PUBLISHED 12 September 2024

## CITATION

Tomasello S, Manzo E and Karbstein K (2024)  
Comparative plastome assembly of the  
yellow ironweed (*Verbesina alternifolia*)  
using Nanopore and Illumina reads.  
*Front. Plant Sci.* 15:1429494.  
doi: 10.3389/fpls.2024.1429494

## COPYRIGHT

© 2024 Tomasello, Manzo and Karbstein. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Comparative plastome assembly of the yellow ironweed (*Verbesina alternifolia*) using Nanopore and Illumina reads

Salvatore Tomasello<sup>1\*</sup>, Eleonora Manzo<sup>1</sup> and Kevin Karbstein<sup>2</sup>

<sup>1</sup>Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Göttingen, Göttingen, Germany, <sup>2</sup>Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany

Chloroplast genomes (plastomes) represent a very important source of valuable information for phylogenetic and biogeographic reconstructions. The use of short reads (as those produced from Illumina sequencing), along with *de novo* read assembly, has been considered the “gold standard” for plastome reconstruction. However, short reads often cannot reconstruct long repetitive regions in chloroplast genomes. Long Nanopore (ONT) reads can help bridging long repetitive regions but are by far more error-prone than those produced by Illumina sequencing. *Verbesina* is the largest genus of tribe Heliantheae (Asteraceae) and includes species of economic importance as ornamental or as invasive weeds. However, no complete chloroplast genomes have been published yet for the genus. We utilized Illumina and Nanopore sequencing data and different assembly strategies to reconstruct the plastome of *Verbesina alternifolia* and evaluated the usefulness of the Nanopore assemblies. The two plastome sequence assemblages, one obtained with the Nanopore sequencing and the other inferred with Illumina reads, were identical, except for missing bases in homonucleotide regions. The best-assembled plastome of *V. alternifolia* was 152,050 bp in length and contained 80, 29, and four unique protein-coding genes, tRNAs, and rRNAs, respectively. When used as reference for mapping Illumina reads, all plastomes performed similarly. In a phylogenetic analysis including 28 other plastomes from closely related taxa (from the *Heliantheae alliance*), the two *Verbesina* chloroplast genomes grouped together and were nested among the other members of the tribe Heliantheae s.str. Our study highlights the usefulness of the Nanopore technology for assembling rapidly and cost-effectively chloroplast genomes, especially in taxonomic groups with paucity of publicly available plastomes.

## KEYWORDS

Asteraceae, chloroplast genome, Nanopore sequencing, tribe Heliantheae, *Verbesina*

## 1 Introduction

Chloroplasts (cp) are the most emblematic organelles of plant cells, responsible for plant photosynthesis and therefore growth and reproduction. cp genomes (plastomes) are often highly conserved throughout land plants in terms of structure, size, and functionality of their genes (Bendich, 2004). Its molecule can be linear or circular (Bendich, 2004) and has a quadripartite structure consisting of two regions of unique DNA (i.e., large and small single-copy regions; LSC and SSC, respectively) and a pair of nearly identical inverted repeat regions (IRB and IRA; Kolodner and Tewari, 1979). Due to their conserved structure, low levels of recombination, and high copy numbers in plant cells, plastomes are an easily accessible source of sequence information for phylogenetic and biogeographic studies (Twyford and Ness, 2017; Tomasello et al., 2020; Karbstein et al., 2022). The advent of second-generation sequencing (e.g., Illumina sequencing) has made the assembly of entire plastomes relatively accessible, and the last decade has registered a proliferation of phylogenetic studies based on plastome data (see Tonti-Filippini et al., 2017; Pascual-Díaz et al., 2021; Dong et al., 2023). To date, more than 40,000 plant chloroplast genomes are publicly available in NCBI's GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

The genome assembly of Illumina reads has improved substantially in the past decade, and a few pipelines have been described especially dedicated to *de novo* assembly of organellar genomes (Twyford and Ness, 2017; Freudenthal et al., 2020). Fast-Plast (available at <https://github.com/mrmckain/Fast-Plast/>), NOVOplasty (Dierckxsens et al., 2017), and GetOrganelle (Jin et al., 2020) are some of the most widely used. These approaches usually need a certain amount of input data (>1 gigabases, Gbp; genome skimming data for GetOrganelle; <https://github.com/Kinggerm/GetOrganelle>) and are not always able to yield accurate assemblies when confronted with long repeat regions in chloroplast genomes (Zhou et al., 2023). In a comparison between different chloroplast genome assembly tools, GetOrganelle outperformed the others (Freudenthal et al., 2020).

Long reads, as those produced by third-generation sequencing techniques [i.e., Oxford Nanopore Technologies (ONT) and Pacific Biosciences (Pacbio)], are able to bridge long repetitive regions and therefore are helpful for plastome assembly (Pucker et al., 2022). On the other hand, these approaches are more prone to errors (Rang et al., 2018). In contrast to Pacbio, some errors in ONT sequencing seem to be non-random. Deletion errors, which are the most common errors found in Nanopore reads, increase in homonucleotide regions (Laehnemann et al., 2016). A-T miscall errors in ONT reads are less likely than all other substitution errors (Twyford and Ness, 2017; Scheunert et al., 2020).

Although several tools are available for *de novo* assembly using long-read and hybrid (both short- and long-reads) data [among others canu (Koren et al., 2017), unicycler (Wick et al., 2017), and flye (Kolmogorov et al., 2019; Syme et al., 2021)], fewer tools especially dedicated to the assembly of organellar genomes have been developed. Organelle\_pba (Soorni et al., 2017) performs *de novo* assembly of any organellar (chloroplast or mitochondrial) genomes using Pacbio reads. MitoHiFi is addressing the assembly of

mitochondrial genomes for a wide range of organisms (including plants) using Pacbio HiFi reads (Uliano-Silva et al., 2023). The newly described ptGAUL (Zhou et al., 2023) is dedicated specifically to plastomes, and it is able to use ONT and Pacbio reads.

*Verbesina* L. is the largest genus of tribe Heliantheae Cass. (Asteraceae), comprehending more than 325 species (Panero, 2007). It is a very diverse genus, including trees of montane moist forests, shrubs, and perennial (but also a few annual) herbs. It is distributed in Central America, the tropical Andes, and eastern Brazil, with a few taxa in the temperate regions of North and South America (Panero and Strother, 2021). Most of the species' diversity is concentrated in Mexico and south-western USA. A few species of *Verbesina* are of big economic value as ornamentals [e.g., *V. encelioides* (Cav.) Benth. & Hook.f. ex A.Gray, *V. alternifolia* (L.) Britton ex Kearney] or are weeds with a negative impact on ecosystems in different areas of the world (Feenstra and Clements, 2008; Fufa et al., 2022; Mehal et al., 2023).

The last revision of the genus was done in the 19<sup>th</sup> Century (Robinson and Greenman, 1899). A modern, comprehensive revision of the genus, including DNA-based phylogenetic evidence, is still missing (Panero and Strother, 2021). Recently, Lopes Moreira et al. (2023) analyzed a wide number of species of the genus using two multi-copy nuclear markers. A chloroplast genome for the genus is still missing, and so far, sequence information is available only for a few plastid regions of a number of *Verbesina* species (Funk et al., 2005).

With the present contribution, we aim to assemble the first chloroplast genome of a *Verbesina* species. For the scope, we utilized Illumina sequencing and long reads produced with Nanopore technology. We used different assembly strategies/tools and evaluated the correctness of the Nanopore assemblies. We evaluated the efficiency of the assembled plastomes as reference for reads mapping and reconstructed a phylogenetic tree with available plastome information from various members of the *Heliantheae* alliance.

## 2 Materials and methods

### 2.1 Plant material and DNA extraction

Leaf material from *Verbesina alternifolia* cultivated at the Old Botanical Garden of the University of Göttingen was collected in summer 2023 and silica-gel-dried. Genomic DNA was extracted from ~1.5-cm<sup>2</sup> leaf material of silica-dried samples using Qiagen DNeasy Plant Mini Kit<sup>®</sup> (Qiagen, Hilden, Germany). We followed the manufacturer's instructions, except for the incubation times in the lysis buffer and elution buffer, which were both increased to 30 min. DNA quality and fragment length were checked by gel electrophoresis in 1.5% agarose gel and using the Midori Green Advance DNA stain (NIPPON Genetics EUROPE, Düren, Germany) and the Quantitas Pro DNA Marker 100 bp–10 kb (Biozym Scientific GmbH, Hessisch Oldendorf, Germany). DNA concentration was estimated using 2 µL of extract and the Qubit<sup>®</sup> fluorometer with the Qubit<sup>®</sup> dsDNA HS Assay Kit (ThermoFisher Scientific, Waltham, USA).

## 2.2 Illumina library prep and sequencing

A sequencing library was prepared using the “NEBNext Ultra II FS DNA Library Prep Kit for Illumina” (E7805; New England BioLabs, Ipswich, USA). Fragmentation was carried out for 12 min at 37°C in order to obtain DNA fragments of 300–500 bp. At the end of the library preparation procedure, the sample was PCR-amplified for 14 cycles, during which sample-specific dual indices (“NEBNext Multiplex Oligos for Illumina<sup>®</sup>”, E7600; New England BioLabs) were added to the fragments. The library was purified with 50 µL of HighPrep PCR beads (MagBio, Gaithersburg, USA), following the manufacturer’s protocol. Concentrations were measured with the Qubit<sup>®</sup> fluorometer, and fragment length distributions and absence of adapter dimers were checked using a Quiagen Qiaxcel and a high-resolution cartridge (Quiagen, Hilden, Germany).

Sequencing was conducted at the NGS- Integrative Genomics Core Unit (NIG; University of Göttingen) on an Illumina NovaSeq6000 (Illumina Inc.) SP 300 cycles flow cell. The sample was mixed equimolarly with other samples in order to gather approximately 5–7.5 Gbp of data after sequencing.

## 2.3 Nanopore library prep and sequencing

Library preparation was conducted using the ONT Ligation Sequencing Kit SQK-LSK110 optimized for high-throughput and long reads (ONT, Oxford, UK) and applicable for singleplex gDNA sequencing. We adjusted the DNA concentration to 1,000 ng in 47 µL (ca. 21.5 ng/µL). We followed the manufacturer’s instructions for library preparation (protocol vers. GDE\_9141\_v112\_revH\_01Dec2021, accessible via [community.nanoporetech.com](https://community.nanoporetech.com)) with the few modifications applied in [Karbstein et al. \(2023\)](#). Accordingly, incubation times were increased up to 15 min, ethanol wash buffer concentration was increased to 80%, and we enriched the DNA fragments of more than 3 kilobases (kb).

A hardware check was performed prior to the run. We used a MinION Mk1B device and the ONT software MinKNOW vers. 21.11.9 installed on a local Linux system. We loaded the libraries into a R9.4.1 flow cell following the manufacturer’s instructions for priming and loading. Sequencing was run for 72 h.

## 2.4 Processing Nanopore data

Basecalling was done on the local HPC cluster of the University of Göttingen, (GWDG, Göttingen, Germany) using ONT software GUPPY vers. 6.0.1 and the configuration file “dna\_r9.4.1\_450bps\_hac.cfg” (i.e., “high accuracy” basecalling). Fastq files were then appended to a single file, which was submitted to PORECHOP vers. 0.2.4 for adapter trimming (available at <https://github.com/rrwick/Porechop>), with the discard\_middle option turned on. Trimmed reads were then subjected to length- and quality-filtering using CHOPPER vers. 0.2.0 ([De Coster and Rademakers, 2023](#)), discarding all reads shorter than 500 bp (–minlength 500). Two datasets were

produced with average reads quality thread equal to or higher than 8 (–q 8) and 9 (–q 9), respectively.

Assembly was performed with CANU vers. 2.2 ([Koren et al., 2017](#)) on both datasets. For the scope, the genome size was set to 155 kilobases (kb), and the minOverlapLength was set to 500 bp. The correctedErrorRate, which is the allowed difference in overlap between two corrected reads, was set to 0.134, as suggested by the developers for high-sequencing-depth Nanopore data.

Assembled contigs were blasted against the plastome of *Helianthus annuus* L. (NC\_007977.1) using BLASTN vers. 2.5.0 ([Zhang et al., 2000](#)). Contigs mapping to the *Helianthus* plastome and longer than 50 kb were aligned. Therefore, a consensus sequence was produced, and IUPAC codes were used in case of incongruities. Once this procedure was done for both the q8 and q9 datasets, the obtained sequences were aligned, and the incongruities were resolved by looking at the base called on the consensus sequence without ambiguity. In cases where the difference involved homonucleotides, the solution with the lowest number of repeated nucleotides was selected and considered the most conservative solution.

In addition to the above-mentioned procedure, we performed plastome assembly with the newly described pipeline ptGAUL ([Zhou et al., 2023](#)) (available at <https://github.com/Bean061/ptgaul>). ptGAUL is able to assemble plastid genomes using long-read data (similarly to GetOrganelle for Illumina reads). It is one of the very few available pipelines for this purpose, and it is able to work both with Nanopore and PacBio data. We used the plastome of *H. annuus* (NC\_007977.1) as reference and the default settings, apart from the coverage (–c), which was set to 500.

## 2.5 Processing Illumina reads and annotation

Illumina reads were *de novo*-assembled using the software GetOrganelle vers. 1.7.7.0 ([Jin et al., 2020](#)), specifying the “embplant\_pt” database and with the k-mer size ranging from 21 to 115. The maximum number of extension rounds was set to 30, and the maximum number of reads (–max-reads) was increased to 7.5e7. The average embplant\_pt base coverage was 1,144.

The plastome from GetOrganelle was annotated using the online tool GeSeq ([Tillich et al., 2017](#)) (available at <https://chlorobox.mpimp-golm.mpg.de/geseq.html>). The BLAT ([Kent, 2002](#)) searches were done by setting the protein search identity to 80% and the rRNA, tRNA, and DNA search identity to 85%. As reference plastome, we used the *H. annuus* plastome also used in the above-mentioned analyses (NC\_007977.1). Additionally, the MPI-MP land plant references were used for chloroplast CDS and rRNAs. For tRNA annotation, a HMMER profile search was also done using ARAGORN vers. 1.2.38 ([Laslett and Canback, 2004](#)) with the default settings. The obtained annotation was checked manually in Geneious Prime 2022.1.1 (<https://www.geneious.com>). The final annotated chloroplast genome was converted into a graphical maps using OGDRAW ([Greiner et al., 2019](#); available at <https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) with default settings.

## 2.6 Read mapping using the assembled plastomes as reference

To evaluate the performance of the plastomes obtained with the different sequencing technologies when acting as reference, we mapped the Illumina reads against the two plastome sequences (i.e., obtained from Illumina and Nanopore sequencing) using BOWTIE2 vers. 2.3.5.1 (Langmead and Salzberg, 2013) and BWA vers. 0.7.16 (Li and Durbin, 2009). Before mapping, the fastq files were processed to trim adapters and filter low-quality reads with TRIMMOMATIC vers. 0.33 (Bolger et al., 2014). Duplicate reads were also excluded using FASTUNIQU (Xu et al., 2012). After mapping, we compared the percentage of reads mapping to the reference and the average coverage.

## 2.7 Phylogenetic analyses

In order to evaluate the correct phylogenetic position of the two reconstructed plastomes (i.e., from Illumina and Nanopore sequencing) in relation to those of closely related organisms and check the possible effects of the differences between them, a phylogenetic analysis was conducted. A total of 28 chloroplast genomes of taxa closely related to *Verbesina alternifolia* (from the *Heliantheae* alliance) were downloaded from GenBank and aligned with the two obtained plastomes (i.e., from Illumina and Nanopore sequencing). A complete list of the accessions used is given in Table 1. In some cases, the orientation of the small single copy region (SSC) was inverted in order to make it match to the other accessions (see Table 1). *Senecio vulgaris* L. (NC\_046693.1) was included and used as outgroup. The sequences were processed in AliView vers. 1.20 (Larsson, 2014) and aligned with MAFFT vers. 7.305b (Kato and Standley, 2013), with the *-auto* strategy. The final alignment consisted of 163,123 characters, and the region between positions 140,777 and 141,391 was masked due to a long insertion in the plastome of *Parthenium argentatum* A.Grey (NC\_013553.1) that made an unambiguous alignment of the rest of the accessions impossible. A maximum likelihood (ML) phylogenetic tree was inferred with RAXML-NG vers. 1.2.0 (Kozlov et al., 2019), using the GTR+G as sequence evolution model and applying 1,000 bootstrap replicates.

## 3 Results

### 3.1 Nanopore sequencing

Sequencing produced 116.81 Gigabytes (Gb) of \*fast5 files (986 files in total). Furthermore, 3.94 million reads were generated with a median read fragment length of 7.11 kb. The longest read was 630 kb, and the overall data produced amounted to 11.14 Gbp. After basecalling, 1,875,797 reads were present in the fastq file, and 1,875,688 reads survived adapter trimming. After length- and quality-filtering, 1,349,347 and 1,349,205 reads were still present when using *-q* 8 and *-q* 9 as quality thresholds, respectively (corresponding to approximately 5.473 Gbp of sequence data).

TABLE 1 Detailed information on the plastomes used for the phylogenetic analysis in Figure 2.

Taxon	GenBank accession no.	Subtribe	Tribe
<i>Acmella paniculate</i>	MZ_292978	Spilanthinae	Heliantheae
<i>Aldama grandiflora</i>	MN_337894	Helianthinae	Heliantheae
<i>Ambrosia trifida</i>	NC_036810	Ambrosiinae	Heliantheae
<i>Bidens parviflora</i>	MW_691204	Coreopsidinae	Coreopsideae
<i>Dahlia pinnata</i>	NC_066129*	Coreopsidinae	Coreopsideae
<i>Echinacea pallida</i>	NC_034321	Zinniinae	Heliantheae
<i>Echinacea purpurea</i>	NC_034327	Zinniinae	Heliantheae
<i>Eclipta alba</i>	NC_039774*	Ecliptinae	Heliantheae
<i>Eclipta prostate</i>	NC_030773*	Ecliptinae	Heliantheae
<i>Gaillardia pulchella</i>	OR_124734*	Gaillardinae	Helenieae
<i>Galinsoga parviflora</i>	NC_046787	Galinsoginae	Millerieae
<i>Helianthus annuus</i>	NC_007977*	Helianthinae	Heliantheae
<i>Helianthus tuberosus</i>	MG_696658	Helianthinae	Heliantheae
<i>Iostephane heterophylla</i>	MT_700542	Helianthinae	Heliantheae
<i>Pappobolus lanatus</i>	MT_700543	Helianthinae	Heliantheae
<i>Parthenium argentatum</i>	NC_013553	Ambrosiinae	Heliantheae
<i>Parthenium hysterophorus</i>	MT_576959	Ambrosiinae	Heliantheae
<i>Rudbeckia hirta</i>	OR_124735	Rudbeckiinae	Heliantheae
<i>Rudbeckia laciniata</i> var. <i>laciniata</i>	MN_518844	Rudbeckiinae	Heliantheae
<i>Sigesbeckia orientalis</i>	MN_240004	Milleriinae	Millerieae
<i>Silphium integrifolium</i>	OM_162161	Engelmanniinae	Heliantheae
<i>Silphium perfoliatum</i>	NC_060408*	Engelmanniinae	Heliantheae
<i>Smallanthus sonchifolius</i>	NC_072537	Milleriinae	Millerieae
<i>Sphagnetocola calenduladea</i>	NC_039346*	Ecliptinae	Heliantheae
<i>Tithonia diversifolia</i>	MT_700544	Helianthinae	Heliantheae
<i>Xanthium sibiricum</i>	MH_473582	Ambrosiinae	Heliantheae
<i>Xanthium spinosum</i>	NC_054222	Ambrosiinae	Heliantheae
<i>Senecio vulgaris</i>	NC_046693		Senecioneae

The plastome of *Senecio vulgare*, not a member of the *Heliantheae* alliance, was used as outgroup. The asterisk symbols are for sequences in which the SSC was inverted in the alignment used for inferring the phylogenetic tree shown in Figure 2.

The plastome assembled using our custom procedure (i.e., assembly with *canu*, *blastn*, concatenation of plastome contigs; see “Materials and methods” section) was 151,795 bp, with 45 unresolved ambiguities, including 43 N, one M (A or C), and one Y (T or C). *ptGAUL* produced two results (paths), corresponding to the different states caused by the differing orientations of the small single copy (SSC) region [“flip-flop” (Palmer, 1983; Walker et al.,

2015)]. The plastome assembled with ptGAUL was 151,829 bp in size (see Table 2).

### 3.2 Illumina sequencing

Approximately 25.7 million pairs of reads were generated by Illumina sequencing, corresponding to 7.7 Gbp of data. GetOrganelle obtained a circular plastid genome of 152,050 bp in two different states. This plastome is available on GeneBank with accession number PP639077.

When comparing the plastome obtained from Illumina reads to those from the Nanopore sequencing, the GetOrganelle plastome was 221 and 255 bp longer than the ones obtained with ptGAUL and with our custom procedure, respectively. The fewer basepairs found in the Nanopore plastomes corresponded to the bases missing in homonucleotides regions. The same is true for the *N* in the plastome reconstructed with the custom procedure, which were all found in homonucleotide regions. The *M* in the latter plastome corresponded to a *C* in the other two plastomes, whereas the *Y* corresponded to a *T* (Table 2). In approximately 60% of the cases, bases missing from homonucleotide regions were found in the same regions of the two plastomes inferred with ONT data. In approximately 50% of cases, the number of missing bases was identical (including gaps as long as seven basepairs). The alignment with the three plastomes inferred (i.e., one from Illumina data and two from Nanopore reads) is available on Göttingen Research Online (GRO.data; doi:10.25625/QJCSC8).

### 3.3 Annotation

The annotated plastome shows a typical tripartite structure with a small single-copy region (SSC, length: 18,241 bp) and a large

TABLE 2 Summary of the assemblies obtained with reads from different sequencing technologies and with different approaches (for the Nanopore reads).

	Illumina	Nanopore	
	GetOrganelle	Custom procedure	ptGAUL
Total length	152,050	151,795	151,829
A	47,205	47,103	47,129
C	28,264	28,218	28,235
T	47,542	47,434	47,463
G	29,039	28,995	29,002
N	0	43	0
y (instead of t)		1 (at position 13,091)*	
m (instead of c)		1 (at position 110,091)*	
Total changes		300	221
% dissimilarity		0.197	0.145

single-copy region (LSC, length: 83,721 bp) separated by two IR regions (length:  $2 \times 25,044$  bp; Figure 1). The plastome contains 86 (80 unique) protein-coding genes, 36 (29 unique) tRNAs, and eight (four unique) rRNAs (duplicated ones in IRs).

In comparison with the *Helianthus annuus* plastome (NC\_007977.1), the *Verbesina alternifolia* chloroplast genome (the one here reconstructed with GetOrganelle) was 969 bp longer. It has the same numbers of unique protein-coding genes and rRNA (80 and four, respectively) and slightly more unique tRNAs (29 vs. 28).

### 3.4 Mapping performance

Mapping of Illumina reads produced comparable results when using the different references (i.e., the plastomes obtained from Illumina and ONT reads) and mapping methods (Table 3). BWA seems to have performed slightly better than BOWTIE2 in terms of mapped reads (560,311.5 and 537,273.5 reads mapped on average, respectively). When comparing the performance as reference, the plastome obtained with the Illumina reads gave slightly higher numbers of mapped reads (565,237 vs. 546,310 reads), although average per-base coverage was better when using the Nanopore reference (134.055 vs. 128.996). Detailed results are given in Table 3.

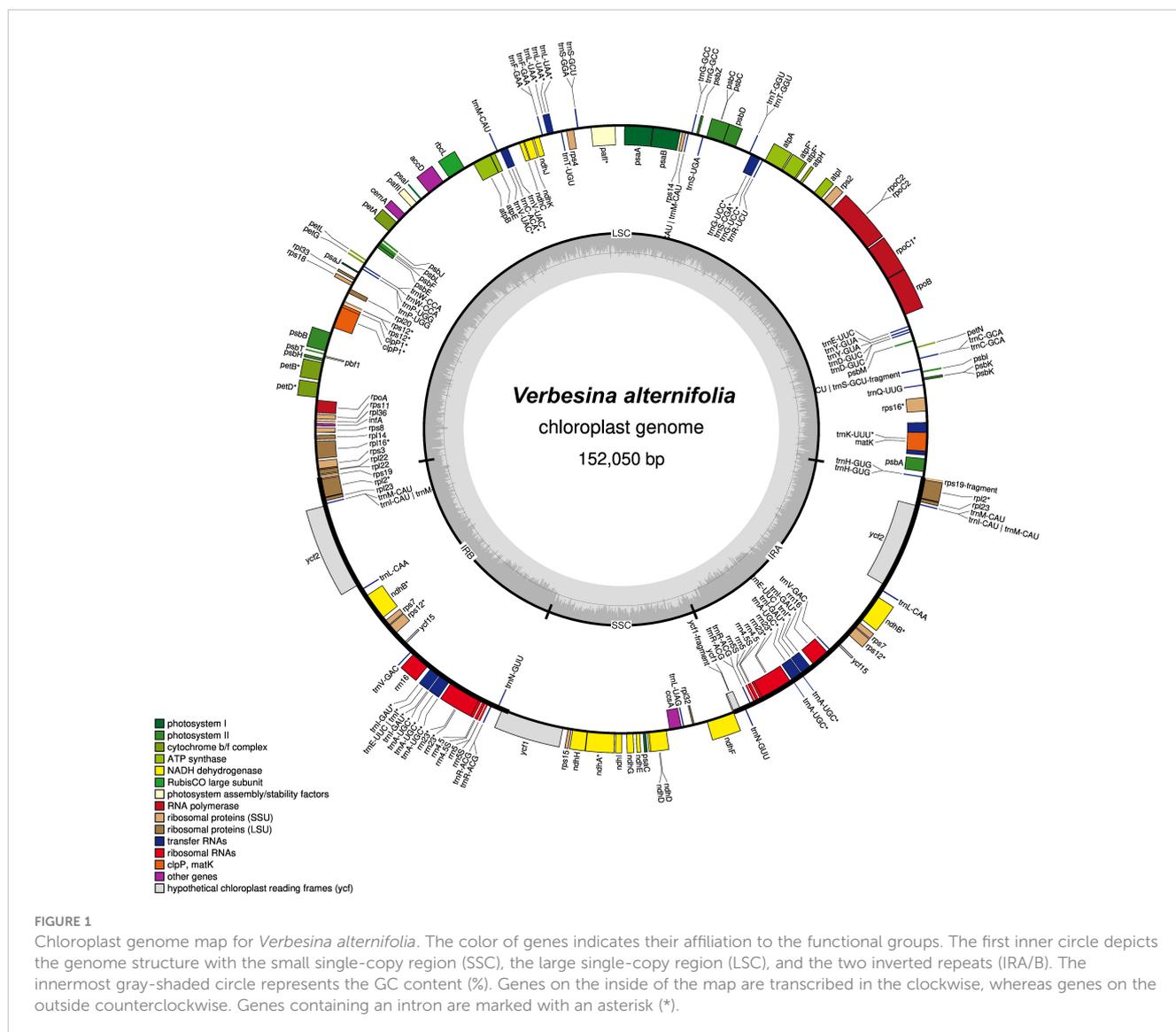
### 3.5 Phylogenetic analyses

The phylogenetic tree obtained from the ML analyses was fully resolved, with just a couple of clades receiving relatively low bootstrap (bs) support values (bs <90; Figure 2). The two *Verbesina* plastomes (i.e., Illumina and ONT assemblies) grouped together with high support (bs: 100). Those are found within tribe Heliantheae, sister to a clade including subtribes Rudbeckiinae H. Rob., Engelmanniinae Stuessy, Spilathinae Panero, Zinniinae Benth. & Hook.f., Ambrosiinae Less., and Helianthinae Dumort (clade, however, moderately low supported; bs: 78).

Additionally, the clade including Helianthinae and *Parthenium* L. (from subtribe Ambrosiinae) is not supported (bs: 68). Subtribe Ambrosiinae is not reconstructed as monophyletic, with genera of the “core” Ambrosiinae (here *Xanthium* L. and *Ambrosia* L.) grouping together with high support and *Parthenium* sister to subtribe Helianthinae (relationship, however, not supported).

## 4 Discussion

Insights gained from chloroplast genomes have enhanced our understanding of plant biology and diversity (Daniell et al., 2016). On the one hand, in the last few decades, the use of a few variable plastid regions has served to resolve phylogenies at a relatively deep taxonomic level (Wicke and Quandt, 2009; Dong et al., 2015). On the other, complete chloroplast genomes provide the high resolution necessary to differentiate closely related taxa and are therefore a valuable source of information to decipher phylogenetic relationships between them and to improve our understanding of the evolution of plant species (Daniell et al., 2016).



Third-generation sequencing, with the long reads it can produce, facilitates *de novo* genome assembly of plastid genomes, particularly in the four junctions between the inverted repeat (IR) and single-copy regions (Daniell et al., 2016). Long reads can also help in assembling plastid genomes in plant groups characterized by rampant plastome rearrangements (see Wicke et al., 2011 for a review). Nanopore technology has demonstrated to be a good solution for sequencing and assembling plastid genomes in plant groups for which this

information is still missing (Bae and Kim, 2021; Scheunert et al., 2020; among others).

With the present contribution, we assembled the first chloroplast genome for the genus *Verbesina*. *Verbesina* is the most species-rich genus of tribe Heliantheae (Asteraceae), with a number of species important as ornamental or because of their invasive attitude. For the scope, we have used both short Illumina reads and long Nanopore ones. We have assembled chloroplast genomes using different

**TABLE 3** Comparison of the mapping results for the short Illumina reads when using as reference the plastomes assembled with Illumina or Nanopore reads and two different mapping programs.

Reference plastome (mapping tool)	Nr. reads	Paired reads	Forward unpaired	Reverse unpaired	Mapped reads	% mapped	mean coverage
Illumina (bowtie2)	43,120,986	21,086,222	782,412	166,130	537,317	1.246	128.360
Illumina (bwa)					565,237	1.31	129.632
ONT custom procedure (bowtie2)					537,234	1.245	128.555
ONT custom procedure (bwa)					555,386	1.287	139.555

“Nr. reads” refers to the number of reads after quality trimming and duplicate reads removal.

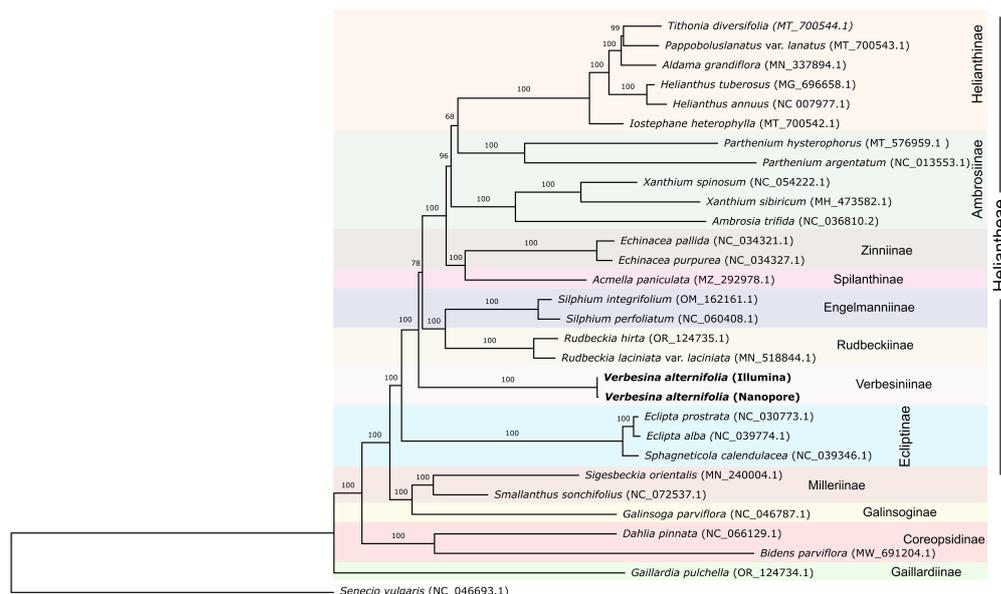


FIGURE 2

ML phylogenetic tree including *Verbesina* plastomes assembled from Illumina and Nanopore reads, along with plastomes from members of the *Heliantheae* alliance available on the GenBank (Table 1). The numbers above the branches are bootstrap support values. On the right part of the figure, the subtribal membership of the accessions is indicated. Colors are according to subtribes in the *Heliantheae* alliance.

approaches and evaluated the effectiveness of the Nanopore reads against the Illumina data. Our study highlights the importance of Nanopore technology as a tool to rapidly and cost-effectively assemble plastid genomes, particularly in taxonomic groups in which no plastomes are publicly available yet.

#### 4.1 Illumina vs. Nanopore assemblies

The plastome assembled with Nanopore data resulted very similar to the one inferred with Illumina reads by GetOrganelle. In the plastome assembled with our custom procedure (see “Materials and methods”), only two positions were not unambiguously resolved, whereas the plastome reconstructed with ptGAUL was identical to the Illumina genome, apart from missing bases in homonucleotide regions. When considering gaps (missing bases in homonucleotide regions), the sequence identity of the former and latter plastomes to the one obtained with Illumina reads is 99.80% and 99.85%, respectively (see Table 2). These values are in line with those obtained in comparable studies [e.g., 99.59% in Scheunert et al. (2020)].

All mismatches (all but two in the Nanopore plastome assembled with our custom procedure) were gaps in homonucleotide regions. In approximately 60% of cases, these gaps were found in the same homonucleotide regions in both Nanopore plastomes; in approximately 50% of cases, these gaps were identical in terms of missing nucleotide, including extreme cases of gaps with up to seven missing nucleotides (see alignment available at <https://doi.org/10.25625/QJCSC8>). If, on the one hand, this confirms what has already been found in other studies, that Nanopore sequencing accumulates deletion errors in homonucleotide regions (Laehnmann et al., 2016; Scheunert et al., 2020), it also testifies that this accumulation does not

depend on the assembly strategy used. It must be also noted that since couple of years the R10.4.1 flow cell has been released, which has contributed to improving the read quality and the results of assemblies, although homopolymer problems persist to some extent (Lerminiaux et al., 2024; Sawicki et al., 2024). However, such mismatches are unimportant in phylogenetic analyses since most of the phylogenetic methods do not take gaps into account (Husmeier, 2005; Mahadani et al., 2022). This is also reflected in our phylogenetic analyses, in which (as expected) the *Verbesina* plastomes assembled from reads obtained with different sequencing technologies clustered together and with full support in the phylogenetic tree (Figure 2).

When comparing the two approaches used for the assembly of Nanopore reads, ptGAUL clearly outperformed our custom approach (i.e., assembly with canu, blastn, and concatenation of plastome contigs). ptGAUL managed to assemble a circular molecule, whereas canu was unable to recover the plastome in a single contig. This is something already noticed in other studies (Wang et al., 2018; Scheunert et al., 2020) in which the *de novo* assembly of plastomes often resulted in two or more contigs of varying length. There might be multiple reasons for that: the high coverage of our Nanopore sequencing may result in alternative assemblies which can lead to contig fragmentation (Izan et al., 2017; Wang et al., 2018); the quality of the DNA (Bethune et al., 2019), i.e., we used silica-gel-dried leaf material and an extraction protocol not specifically designed for the extraction of ultra-long fragments.

A chloroplast genome inferred with Nanopore data might also be important as reference in taxonomic groups, in which no other plastomes are publicly available yet. For instance, *de novo* assemblers of plastomes based on Illumina data need a certain sequencing depth (>1 Gbp of genome skimming data as suggested

by the developers of GetOrganelle). The amount of data needed for *de novo* assembly may vary depending on different factors, e.g., factors intrinsic to the taxonomic group under investigation (e.g., nuclear genome size), the type of material (degraded DNA and archival DNA), or technical issues (library preparation protocol). In such circumstances, even more than 1 Gbp of genome skimming data might not be enough to reconstruct complete plastid genomes using short reads data. A reference plastome obtained with long Nanopore reads may serve for reference-based mapping of Illumina short reads with, e.g., BWA or BOWTIE2, when the amount of Illumina data available is not enough to reconstruct complete plastomes with the above-mentioned *de novo* approaches.

## 4.2 Phylogenetic considerations

Inference of phylogenetic relationships among higher-level classification lineages and the identification of clear borders of tribe Heliantheae have been challenging due to extensive variations in morphological characters registered in members of this taxonomic group and the complex patterns that sometimes result from molecular-based phylogenetic analyses (Panero, 2007; Baldwin, 2009). Heliantheae is now treated as a group (alliance), including a few major lineages recognized at tribal rank (Panero and Funk, 2002), Heliantheae s.str. being one of them. The relationships among the tribes that we included in this study (Figure 2) reflect the one depicted in Panero (2007) (also based on cp data) and in Zhang et al. (2021) (based on transcriptomic data), with Helenieae Lindl. (here represented by *Gaillardia* Foug.) as an early branch in the alliance and then Coreopsidae Lindl. (*Bidens* L. and *Dahlia* Cav.) and Millerieae Lindl. (*Galinsoga* Ruiz & Pav., *Smallanthus* Mack., *Sigesbeckia* L.) as sister to the Heliantheae s.str. Within tribe Heliantheae, the position of subtribes is again congruent with the cpDNA tree shown in Panero (2007), with Verbesiniinae being sister to a relatively big clade including Rudbeckiinae, Engelmanniinae, Spilanthinae, Zinniinae, Ambrosiinae, and Helianthinae (Figure 2). The supposed sister relationship between Verbesiniinae and Engelmanniinae based on transcriptomic data (Zhang et al., 2021) is not supported here.

Interestingly, subtribe Ambrosiinae resulted paraphyletic, with *Parthenium* nested rather in the clade with members of subtribe Helianthinae (Figure 2). The non-monophyly of subtribe Ambrosiinae was already supposed (Panero, 2005) and supported by cp DNA data (Panero et al., 2001; Panero, 2007; Tomasello et al., 2019). Nuclear DNA-based analyses, however, seem to sustain the

monophyly of the subtribe (Tomasello et al., 2019; Zhang et al., 2021).

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

## Author contributions

ST: Conceptualization, Data curation, Formal Analysis, Investigation, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. EM: Formal Analysis, Investigation, Validation, Writing – review & editing. KK: Investigation, Validation, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the “Seed Money for Post-Doc Grants” of the DFG priority program 1991 “Taxon-Omics”. We gratefully acknowledge the support of the Open Access Publication Funds of the University of Göttingen.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Bae, S. H., and Kim, C. K. (2021). The complete chloroplast genome of *Adenophora triphylla* (Asterales: Campanulaceae). *Mitochond. DNA Part B: Resour.* 6, 82–83. doi: 10.1080/23802359.2020.1847613
- Baldwin, B. G. (2009). “Heliantheae Alliance,” in *Systematics, Evolution, and Biogeography of Compositae*. Eds. V. A. Funk, A. Susanna, T. F. Stuessy and R. J. Bayer (IAPT, Vienna), 689–711.
- Bendich, A. J. (2004). Circular chloroplast chromosomes: The grand illusion. *Plant Cell* 16, 1661–1666. doi: 10.1105/tpc.160771
- Bethune, K., Mariac, C., Couderc, M., Scarcelli, N., Santoni, S., Ardisson, M., et al. (2019). Long-fragment targeted capture for long-read sequencing of plastomes. *Appl. Plant Sci.* 7, e1243. doi: 10.1002/aps3.1243
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134. doi: 10.1186/s13059-016-1004-2

- De Coster, W., and Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 39, btad311. doi: 10.1093/bioinformatics/btad311
- Dierckxens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: *De novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18. doi: 10.1093/nar/gkw955
- Dong, W., Gao, L., Xu, C., Song, Y., and Pocza, P. (2023). Editorial: Rise to the challenges in plastome phylogenomics. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1200302
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* 5, 8348. doi: 10.1038/srep08348
- Feenstra, K. R., and Clements, D. R. (2008). Biology and impacts of Pacific Island invasive species. 4. *Verbesina encelioides*, Golden Crownbeard (Magnoliopsida: Asteraceae). *Pacific Sci.* 62, 161–176. doi: 10.2984/1534-6188(2008)62[161:BAIOPI]2.0.CO;2
- Freudenthal, J. A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M. J., and Förster, F. (2020). A systematic comparison of chloroplast genome assembly tools. *Genome Biol.* 21, 254. doi: 10.1186/s13059-020-02153-6
- Fufa, A., Tessema, T., Bekeko, Z., and Mesfin, T. (2022). Distribution and Abundance of Wild Sunflower (*Verbesina encelioides*) and its Impacts on Plant Biodiversity in The Central Rift Valley of Ethiopia. *Ethiop. J. Crop Sci.* 10, 89–106.
- Funk, V. A., Bayer, R. J., Keeley, S., Chan, R., Watson, L., Gemeinholzer, B., et al. (2005). Everywhere but Antarctica: Using a super tree to understand the diversity and distribution of the Compositae. *B5* 55, 343–374.
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Husmeier, D. (2005). “Introduction to Statistical Phylogenetics,” in *Probabilistic Modeling in Bioinformatics and Medical Informatics*. Eds. D. Husmeier, R. Dybowski and S. Roberts (Springer, London), 83–145. doi: 10.1007/1-84628-119-9\_4
- Izan, S., Esselink, D., Visser, R. G. F., Smulders, M. J. M., and Borm, T. (2017). *De novo* assembly of complete chloroplast genomes from non-model species based on a K-mer frequency-based selection of chloroplast reads from total DNA sequences. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01271
- Jin, J. J., Yu, W., Yang, J. B., Song, Y., Depamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: A fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21, 241. doi: 10.1186/s13059-020-02154-5
- Karbstein, K., Tomasello, S., Hodač, L., Wagner, N., Marinček, P., Barke, B. H., et al. (2022). Untying Gordian knots: unraveling reticulate polyploid plant evolution by genomic data using the large *Ranunculus auricomus* species complex. *New Phytol.* 235, 2081–2098. doi: 10.1111/nph.18284
- Karbstein, K., Tomasello, S., Wagner, N., Barke, B. H., Paetzold, C., Bradican, J. P., et al. (2023). Efficient hybrid strategies for assembling the plastome, mitochondriome, and large nuclear genome of diploid *Ranunculus cassubicifolius* (Ranunculaceae). *bioRxiv*, 552429. doi: 10.1101/2023.08.08.552429
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kent, W. J. (2002). BLAT - The BLAST-like alignment tool. *Genome Res.* 12, 256–264. doi: 10.1101/gr.229202
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 240–246. doi: 10.1038/s41587-019-0072-8
- Kolodner, R., and Tewari, K. K. (1979). Inverted repeats in chloroplast DNA from higher plants. *Proc. Natl. Acad. Sci.* 76, 41–45. doi: 10.1073/pnas.76.1.41
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive  $\kappa$ -mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAXML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305
- Laehnemann, D., Borkhardt, A., and McHardy, A. C. (2016). Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief. Bioinform.* 17, 154–179. doi: 10.1093/bib/bbv029
- Langmead, B., and Salzberg, S. (2013). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152
- Lerminiaux, N., Fakhruddin, K., Mulvey, M. R., and Mataseje, L. (2024). Do we still need Illumina sequencing data? Evaluating Oxford Nanopore Technologies R10.4.1 flow cells and the Rapid v14 library prep kit for Gram negative bacteria whole genome assemblies. *Can. J. Microbiol.* 70, 178–189. doi: 10.1139/cjm-2023-0175
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lopes Moreira, G., Panero, J. L., Inglis W., P., Zappi C., D., and Cavalcanti, B. T. (2023). A time-calibrated phylogeny of *Verbesina* (Heliantheae – Asteraceae) based on nuclear ribosomal ITS and ETS sequences. *Edinburgh J. Bot.* 80, 1–22. doi: 10.24823/ejb.2023.1953
- Mahadani, A. K., Awasthi, S., Sanyal, G., Bhattacharjee, P., and Pippal, S. (2022). Indel-K2P: a modified Kimura 2 Parameters (K2P) model to incorporate insertion and deletion (Indel) information in phylogenetic analysis. *Cyber-Phys. Syst.* 8, 32–44. doi: 10.1080/23335777.2021.1879274
- Mehal, K. K., Sharma, A., Kaur, A., Kalia, N., Kohli, R. K., Singh, H. P., et al. (2023). Modelling the ecological impact of invasive weed *Verbesina encelioides* on vegetation composition across dryland ecosystems of Punjab, northwestern India. *Environ. Monit. Assess.* 195, 175. doi: 10.1007/s10661-023-11299-2
- Palmer, J. D. (1983). Chloroplast DNA exists in two orientations. *Nature* 301, 92–93. doi: 10.1038/301092a0
- Panero, J. L. (2005). New combinations and infrafamiliar taxa in the Asteraceae. *Phytologia* 87, 1–14.
- Panero, J. L. (2007). “Compositae: Tribe Heliantheae,” in *Families and Genera of Vascular Plants, vol. VIII, Flowering Plants, Eudicots, Asterales*. Eds. J. W. Kadereit and C. Jeffrey (Springer-Verlag, Berlin), 440–477.
- Panero, J. L., Baldwin, E. G., Schilling, E. E., and Clevinger, J. A. (2001). “Molecular phylogenetic studies of members of tribes Helenieae, Heliantheae, and Eupatorieae (Asteraceae),” in *Botany 2001 Abstracts* (Botanical Society of America, Albuquerque, New Mexico), 132.
- Panero, J. L., and Funk, V. A. (2002). Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). *Proc. Biol. Soc. Washingt.* 115, 909–922.
- Panero, J. L., and Strother, J. L. (2021). Chromosome numbers in *verbesina* (Asteraceae, heliantheae, verbesininae). *Lundellia* 24, 1–10. doi: 10.25224/1097-993x-24.1.1
- Pascual-Diaz, J. P., Garcia, S., and Vitales, D. (2021). Plastome diversity and phylogenomic relationships in asteraceae. *Plants* 10, 2699. doi: 10.3390/plants10122699
- Pucker, B., Irisarri, I., De Vries, J., and Xu, B. (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quant. Plant Biol.* 3, e5. doi: 10.1017/qpb.2021.18
- Rang, F. J., Kloosterman, W. P., and de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 19, 90. doi: 10.1186/s13059-018-1462-9
- Robinson, B. L., and Greenman, J. M. (1899). Synopsis of the genus *verbesina*, with an analytical key to the species. *Proc. Am. Acad. Arts Sci.* 34, 534–566. doi: 10.2307/20020930
- Sawicki, J., Krawczyk, K., Pauksztó, Ł., Maździarz, M., Kurzyński, M., Szablińska-Piernik, J., et al. (2024). Nanopore sequencing technology as an emerging tool for diversity studies of plant organellar genomes. *Diversity* 16, 173. doi: 10.3390/d16030173
- Scheunert, A., Dorfner, M., Lingl, T., and Oberprieler, C. (2020). Can we use it? On the utility of *de novo* and reference-based assembly of Nanopore data for plant plastome sequencing. *PLoS One* 15, e0226234. doi: 10.1371/journal.pone.0226234
- Soorni, A., Haak, D., Zaitlin, D., and Bombarely, A. (2017). Organelle\_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC Genomics* 18, 49. doi: 10.1186/s12864-016-3412-9
- Syme, A. E., McLay, T. G. B., Udovicic, F., Cantrill, D. J., and Murphy, D. J. (2021). Long-read assemblies reveal structural diversity in genomes of organelles – an example with *Acacia pycnantha*. *GigaByte* 2021, gigabyte36. doi: 10.46471/gigabyte.36
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Tomasello, S., Karbstein, K., Hodač, L., Paetzold, C., and Hörandl, E. (2020). Phylogenomics unravels Quaternary vicariance and allopatric speciation patterns in temperate-montane plant species: A case study on the *Ranunculus auricomus* species complex. *Mol. Ecol.* 29, 2031–2049. doi: 10.1111/mec.15458
- Tomasello, S., Stuessy, T. F., Oberprieler, C., and Heubl, G. (2019). Ragweeds and relatives: Molecular phylogenetics of Ambrosiinae (Asteraceae). *Mol. Phylogenet. Evol.* 130, 104–114. doi: 10.1016/j.ympev.2018.10.005
- Tonti-Filippini, J., Nevill, P. G., Dixon, K., and Small, I. (2017). What can we do with 1000 plastid genomes? *Plant J.* 90, 808–818. doi: 10.1111/tj.13491
- Twyford, A. D., and Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.* 17, 858–868. doi: 10.1111/1755-0998.12626
- Uliano-Silva, M., Ferreira, J. G. R. N., Krashenninnikova, K., Blaxter, M., Mieszowska, N., Hall, N., et al. (2023). MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinf.* 24, 288. doi: 10.1186/s12859-023-05385-y
- Walker, J. F., Jansen, R. K., Zanis, M. J., and Emery, N. C. (2015). Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. *Am. J. Bot.* 102, 1751–1752. doi: 10.3732/ajb.1500299
- Wang, W., Schalamun, M., Morales-Suarez, A., Kainer, D., Schwessinger, B., and Lanfear, R. (2018). Assembly of chloroplast genomes with long- and short-read data: A comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC Genomics* 19, 977. doi: 10.1186/s12864-018-5348-8
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *Biol. Comput.* 13, e1005595. doi: 10.1371/journal.pcbi.1005595

- Wicke, S., and Quandt, D. (2009). Universal primers for the amplification of the plastid "trnK/matK" region in land plants. *Anales del Jardín Botánico Madrid* 66, 285–288. doi: 10.3989/ajbm.2231
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., et al. (2012). FastUniq: A fast *de novo* duplicates removal tool for paired short reads. *PLoS One* 7, e0052249. doi: 10.1371/journal.pone.0052249
- Zhang, C., Huang, C. H., Liu, M., Hu, Y., Panero, J. L., Luebert, F., et al. (2021). Phylotranscriptomic insights into Asteraceae diversity, polyploidy, and morphological innovation. *J. Integr. Plant Biol.* 63, 1273–1293. doi: 10.1111/jipb.13078
- Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* 7, 203–214. doi: 10.1089/10665270050081478
- Zhou, W., Armijos, C. E., Lee, C., Lu, R., Wang, J., Ruhlman, T. A., et al. (2023). Plastid genome assembly using long-read data. *Mol. Ecol. Resour.* 23, 1442–1457. doi: 10.1111/1755-0998.13787