





# Chromosome-scale reference genome of *Pectocarya recurvata*, the species with the smallest reported genome size in Boraginaceae

Poppy C. Northing  | Jessie A. Pelosi  | D. Lawrence Venable  |  
Katrina M. Dlugosch 

Department of Ecology and Evolutionary Biology,  
University of Arizona, Tucson, Arizona  
85721, USA

## Correspondence

Poppy C. Northing, Department of Ecology and  
Evolutionary Biology, University of Arizona,  
Tucson, Arizona 85721, USA.  
Email: [poppynorthing@gmail.com](mailto:poppynorthing@gmail.com)

## Abstract

**Premise:** *Pectocarya recurvata* (Boraginaceae, subfamily Cynoglossoideae), a species native to the Sonoran Desert (North America), has served as a model system for a suite of ecological and evolutionary studies. However, no reference genomes are currently available in Cynoglossoideae. A high-quality reference genome for *P. recurvata* would be valuable for addressing questions in this system and across broader taxonomic scales.

**Methods:** Using PacBio HiFi sequencing, we assembled a reference genome for *P. recurvata* and annotated coding regions with full-length transcripts from an Iso-Seq library. We assessed genome completeness with BUSCO and *k*-mer analysis, and estimated the genome size of six individuals using flow cytometry.

**Results:** The chromosome-scale genome assembly for *P. recurvata* was 216.0 Mbp long (N50 = 12.1 Mbp). Previous observations indicated *P. recurvata* is  $2n = 24$ . Our assembly included 12 primary contigs (158.3 Mbp) containing 30,655 genes with telomeres at 23 out of 24 ends. Flow cytometry measurements from the same population included two plants with  $1C = 196.9$  Mbp, the smallest measured for Boraginaceae, and four with  $1C = 385.8$  Mbp, which is consistent with tetraploidy in this population.

**Discussion:** The *P. recurvata* genome assembly and annotation provide a high-quality genomic resource in a sparsely represented area of the angiosperm phylogeny. This new reference genome will facilitate answering open questions in ecophysiology, biogeography, and systematics.

## KEYWORDS

Boraginaceae, cytotype variation, *Pectocarya recurvata*, reference genome assembly, small genome size, Sonoran Desert winter annual plant

## Resumen

**Premisa:** *Pectocarya recurvata* (Boraginaceae, subfamilia Cynoglossoideae), una especie nativa del desierto de Sonora, ha servido como un sistema modelo para estudios ecológicos y evolutivos. Sin embargo, Cynoglossoideae no cuenta con ningún genoma de referencia. Un genoma de alta calidad para *P. recurvata* sería valioso para el estudio de este sistema así como estudios a escalas taxonómicas más amplias.

**Métodos:** Utilizando secuenciación PacBio HiFi, ensamblamos un genoma de referencia para *P. recurvata* y anotamos las regiones codificantes con transcritos de tamaño completo empleando una librería de Iso-Seq. Evaluamos la integridad del genoma usando BUSCO y análisis *k*-mer, y también usamos citometría de flujo en seis individuos para estimar el tamaño del genoma.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Applications in Plant Sciences* published by Wiley Periodicals LLC on behalf of Botanical Society of America.

**Resultados:** Nuestro ensemble de referencia para *P. recurvata* mide a nivel de cromosoma 216.0 Mbp (N50 = 12.1 Mbp). Observaciones previas indican que *P. recurvata* es  $2n = 24$ . Nuestro ensemble incluye 12 contig primarios (158.3 Mbp) que contiene 30,655 genes con telómeros en 23 de 24 de los extremos. Medidas de citometría de flujo en la misma población incluyeron dos plantas con  $1C = 196.9$  Mbp, el valor más pequeño medido para Boraginaceae, y cuatro plantas con  $1C = 385.8$  Mbp, cual es consistente con la tetraploidía de esta población.

**Discusión:** El ensemble y la anotación del genoma de *P. recurvata* provee un recurso genómico de alta calidad en una área escasamente representada de la filogenia de las angiospermas. Este nuevo genoma de referencia facilitará el estudio de preguntas abiertas en ecofisiología, biogeografía, y sistemática.

*Pectocarya recurvata* I.M. Johnst. (curvenut combseed) is a winter annual forb in the Cynoglossidae subfamily and Amsinckiinae subtribe of Boraginaceae (Boraginales) (Figure 1) (Chacón et al., 2016; Luebert et al., 2016; Simpson et al., 2017a). *Pectocarya recurvata* is distributed across the Sonoran, Chihuahuan, and Mojave deserts of northern Mexico and the American Southwest, regions known for unpredictable winter precipitation regimes (Noy-Meir, 1973; Huxman et al., 2004). As a winter annual plant in this region, *P. recurvata* relies on cues from late fall and early winter rains to appropriately time its germination, growth, and reproduction to take advantage of the cooler, wetter conditions that occur immediately after rain events (Mulroy and Rundel, 1977; Huxman et al., 2008). The Sonoran Desert winter annual plant community, including *P. recurvata*, has served as an important model system in understanding how species adapt to variable environments and subsequently how these environments promote species coexistence (Pake and Venable, 1995, 1996; Chesson et al., 2004; Angert et al., 2009; Huxman et al., 2013). These investigations have been largely facilitated by the long-term vegetation plots at the Desert Laboratory at Tumamoc Hill (Tucson, Arizona, USA), where detailed vital rates of the most abundant species (including *P. recurvata*) have been recorded since 1982 (Venable, 2007; Huxman et al., 2013). Despite its importance to long-term eco-evolutionary studies, there is no published genome or transcriptome for *P. recurvata* or any species in the Cynoglossidae subfamily.

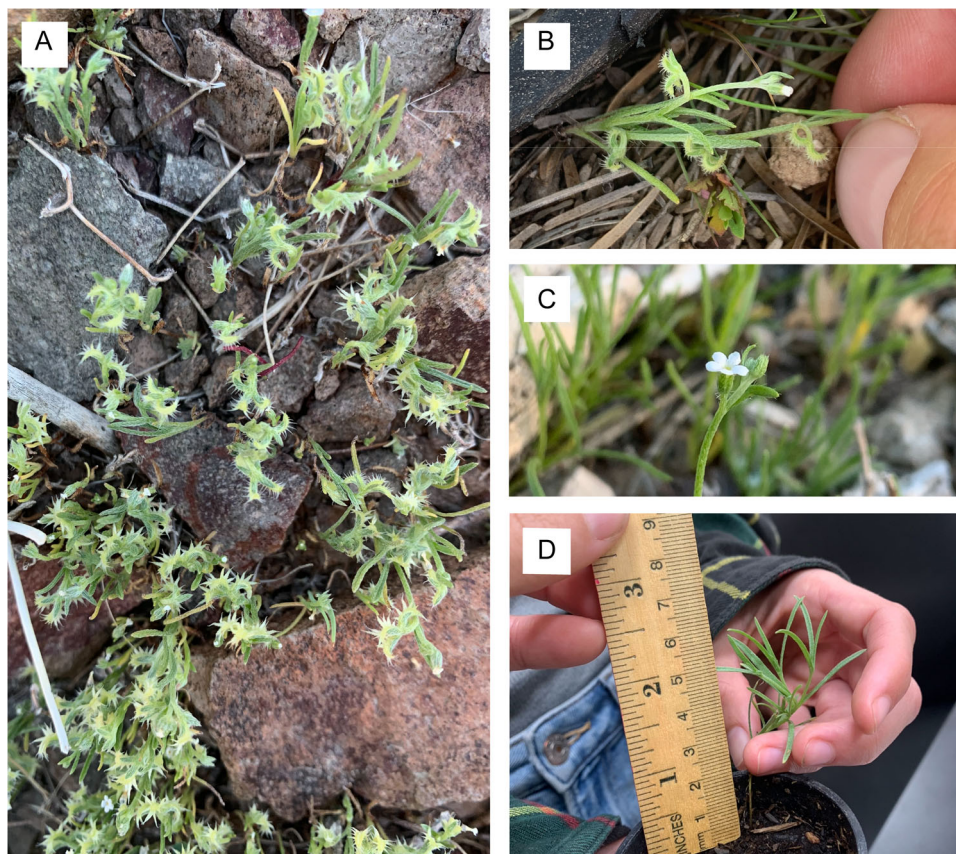
Stable coexistence within the Sonoran Desert winter annual plant community is maintained through temporal variability in precipitation and temperature acting on variation in a fundamental physiological trade-off between growth and reproduction (Angert et al., 2007, 2009; Huxman et al., 2008). Species are separated along a specific trade-off between water use efficiency (the amount of carbon fixed per unit of water lost) and relative growth rate (Angert et al., 2007; Huxman et al., 2008). As a more resource-conservative member of this community, *P. recurvata* exhibits a slow growth rate and highly efficient water use, the latter of which is associated with the maintenance of photosynthesis at lower temperatures to optimize growth after winter rains (Huxman et al., 2008; Barron-Gafford et al., 2013). Two areas of active interest in this system are the genetic architecture underlying the physiological trade-off between relative growth rate and water use efficiency, and the potential for these traits

to be involved in adaptive responses to climate change (Kimball et al., 2010; Angert et al., 2014).

More broadly, evolutionary relationships among taxa in Boraginaceae have served as models to investigate the biogeographical processes underlying commonly observed disjunctions in plant distributions (Gottschling et al., 2004; Moore et al., 2006; Williams et al., 2017; Mabry and Simpson, 2018). Several genera in the Amsinckiinae subtribe, including *Pectocarya*, contain species distributed across the subtropical and temperate regions of North and South America (but not within the tropics themselves), a distribution pattern referred to as the American amphitropical disjunction (AAD) (Raven, 1963; Williams et al., 2017; Simpson et al., 2017b). It is inferred that long-distance dispersal events cause AAD distributions (Raven, 1963; Wen and Ickert-Bond, 2009), yet the types of propagule morphology that might promote long-distance dispersal and the features associated with successful establishment after colonization are less clear (Nathan, 2006; Chacón et al., 2017; Harris et al., 2018). In Amsinckiinae alone, at least 18 distinct long-distance dispersal events across the American tropics have occurred in the relatively recent past (Chacón et al., 2017; Williams et al., 2017).

In Boraginaceae, there are currently (as of October 2024) three publicly available genome assemblies: *Lithospermum erythrorhizon* Siebold & Zucc. (Auber et al., 2020), *Echium plantagineum* L. (Tang et al., 2020), and *Pentaglottis sempervirens* (L.) Tausch ex L.H.Bailey (Darwin Tree of Life Project Consortium, 2022), which are all taxa placed within the Boraginoideae subfamily. There are no genome assemblies currently available within the remaining subfamilies of Echiochiloideae and Cynoglossoideae, despite the Cynoglossoideae being the largest subfamily, with over 1000 species across 50 genera (Chacón et al., 2016). Moreover, the relationships of the major orders that comprise the core Lamiidae phylogeny (Boraginales, Gentianales, Solanales, Lamiales) remain unresolved (Bremer et al., 2002; Soltis et al., 2011; Zhang et al., 2020). Therefore, expanding the availability of nuclear genomic resources into additional areas of the Lamiidae phylogeny could also help clarify the relationships between these orders.

Here, we present a chromosome-scale genome assembly for *P. recurvata*, the first reference genome in the Cynoglossoideae subfamily of Boraginaceae. The *P. recurvata* genome



**FIGURE 1** The morphology of *Pectocarya recurvata*. (A) Individuals of *P. recurvata* growing among individuals of *P. platycarpa* in the crevices of small rocks near Lake Havasu City, Arizona, USA. (B) *Pectocarya recurvata* with many fruits (four fused nutlets). (C) A characteristic *P. recurvata* flower. (D) The individual sampled for genomic and transcriptomic sequencing (ARIZ 444648). Photos taken by Poppy C. Northing.

assembly is a near telomere-to-telomere chromosome-level assembly accompanied by structural and functional gene annotation as well as a complete chloroplast genome assembly and annotation. We also report the smallest genome size to date for a species in Boraginaceae (reported 1C-values range from 270–16,000 Mbp), as well as putative ploidy variation in *P. recurvata*. Additionally, we find evidence of a whole genome duplication (WGD) in the evolutionary history of *P. recurvata* congruent with previously described WGDs in Boraginaceae (Ren et al., 2018; Tang et al., 2020). We analyze gene synteny to compare genome structure between *E. plantagineum* and our assembled *P. recurvata* chromosomes. Our annotated *P. recurvata* genome and plastome assemblies provide essential genomic tools for future opportunities to investigate outstanding questions in physiological ecology, biogeography, and phylogenetics.

## METHODS

### Species description

*Pectocarya recurvata* is a small herbaceous forb with basally branching, strigose stems arranged with alternate, strigose

to bristle-haired linear leaves (Figure 1A) (Veno, 1979; Guilliams and Kelley, 2021). It has racemose inflorescences consisting of self-pollinating, inconspicuous flowers with small, white, funnellform corollas and bilateral calyces (Figure 1C). *Pectocarya recurvata* is distinguished from other species in the genus by its characteristic fruit, which are four fused, recurved nutlets with toothed margins that enable epizoochoric dispersal (Figure 1A, B) (Veno, 1979; Guilliams and Kelley, 2021). Within its range spanning the Chihuahuan, Mojave, and Sonoran deserts of the Southwestern United States and northwestern Mexico, *P. recurvata* commonly occurs in patches at low elevations (10–1600 m), often growing abundantly in the shelter of rocks (Figure 1A) and larger plants such as *Larrea tridentata* (Sessé & Moc. ex DC.) Coville (creosote) (Guilliams and Kelley, 2021). The chromosome number of *P. recurvata* is  $2n = 24$ , as counted by Veno (1979) in individuals from five widely distributed populations in Arizona and California.

### Genome size estimation

The genome size of *P. recurvata* was estimated using flow cytometry with a FACSCanto II flow cytometer (BD

Biosciences, San Jose, California, USA). The 2C DNA content was measured from six individuals grown from seed sourced from the Desert Laboratory at Tumamoc Hill (32°13'N, 111°01'W; Tucson, Arizona, USA) in 2016 and 2019.

Sample preparation and analysis by flow cytometry was conducted following a modified two-step protocol described in Doležel et al. (2007) and Cang et al. (2024). In brief, nuclei were isolated from approximately 20 mg of freshly collected leaf tissue from each *P. recurvata* sample. *Raphanus sativus* 'Saxa' (2C DNA content = 1.11 pg, provided by the Institute of Experimental Botany, Prague, Czech Republic; Doležel et al., 2007) was used as an internal reference standard. Leaf tissue from both species was chopped simultaneously on an ice-cold glass Petri dish in 200 µL of ice-cold Otto I buffer, using a new razor blade for each plant. The nuclear suspension was filtered through 18-µm nylon mesh into a flow cytometry tube and kept on ice. Subsequently, 280 µL of Otto II buffer was added to the suspension, followed by 20 µL of room-temperature 1 mg/mL RNase and 23 µL of ice-cold 1 mg/mL propidium iodide (PI) stain. Each sample was gently vortexed and incubated for 5 min before analysis on the FACSCanto II with a medium flow rate. Measurements of the six samples took place across four different days (Table S9 in Appendix S1; see Supporting Information with this article). The haploid (1C) genome size was calculated using the G1 peak mean PI fluorescence of the sample and standard, following the formula described by Doležel and Bartoš (2005).

## Reference tissue

Plant tissue for the reference genome was sourced from a *P. recurvata* seed collected in spring 2019 at the Desert Laboratory. The seed was germinated in soil in a D25L deepot (Steuwe and Sons, Tangent, Oregon, USA) at 10°C and grown in a greenhouse at the University of Arizona (Tucson, Arizona, USA) in ambient light conditions and hand-watered twice daily. After 77 days, 0.21 g of stem and leaf tissue was collected from the individual (Figure 1D), flash frozen in liquid nitrogen, and stored at -80°C at the Arizona Genomics Institute (AGI); this tissue was used to prepare genomic sequencing libraries. After 66 days of additional growth, 0.52 g of leaf, stem, and reproductive tissue was flash frozen in liquid nitrogen and stored at -80°C at the AGI; this tissue was used to prepare the transcriptome sequencing library. A reference voucher of the remaining plant tissue from this individual is archived in the University of Arizona herbarium (ARIZ 444648).

## Genome and transcriptome sequencing

Genome sequencing was conducted by the AGI following PacBio's HiFi protocols for library preparation and sequencing as follows. After high-molecular-weight DNA was extracted following the Doyle and Doyle (1987)

protocol with minor modifications and assessed for quality control with a Femto Pulse System (Agilent Technologies, Santa Clara, California, USA), DNA was sheared to 10–20 kbp using a Megaruptor 3 sonicator (Diagenode, Denville, New Jersey, USA) and treated with SMRTbell cleanup beads (PacBio, Menlo Park, California, USA). The final HiFi libraries were constructed using a SMRTbell Prep Kit v3.0 (PacBio), size selected to 10–25 kbp using a PippinHT size selector (Sage Science, Beverly, Massachusetts, USA), and validated with the Femto Pulse System (Agilent Technologies). This final, size-selected library was prepared for sequencing using the PacBio SequelII Sequencing Kit v3.1 for HiFi libraries, loaded onto one PacBio Revio SMRT cell, and sequenced for 24 h in circular consensus sequencing mode, yielding 65.6 Gbp of HiFi sequencing data.

An Iso-Seq library was also constructed by the AGI to enable gene identification and annotation of the *P. recurvata* genome assembly. Total RNA was isolated from the flash frozen sample of pooled leaf, stem, and reproductive tissue using the PureLink Plant RNA Reagent kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA), and the RNA integrity number (RIN) was measured for quality control using an Agilent 2100 Bioanalyzer (sample RIN = 7.8). The Kinnex Iso-Seq library was prepared by first synthesizing cDNA from 300 ng of total RNA using the Iso-Seq express 2.0 kit (PacBio), and the size distribution of the resulting cDNA was checked with a Bioanalyzer (Agilent Technologies). The final Iso-Seq library was constructed following the PacBio Kinnex PCR protocol and sequenced on one Revio SMRT cell, generating 75.7 Gbp of HiFi reads comprising 85 million transcripts that reduced to 1.65 million isoform transcripts after clustering was performed using the *IsoSeq* software toolkit (<https://isoseq.how/>, PacBio).

## Assembly

A *k*-mer analysis of the HiFi reads was used to estimate genome size, abundance of repetitive elements, heterozygosity, and ploidy level using KMC v3.1.0 (Kokot et al., 2017), GenomeScope v1.0 (Vurture et al., 2017), and SmudgePlot v1.0 (Ranallo-Benavidez et al., 2020) with a maximum *k*-mer coverage of 10,000 to count and analyze 17-mers, respectively. Values of *k* = 21, 25, 31, 41, and 71 were investigated in the same manner, and the results of these analyses are reported in Table S1 in Appendix S1 and in Figures S2 and S8 in Appendix S2. The initial genome assembly was constructed de novo from the PacBio HiFi reads using Hifiasm v0.16.1 (Cheng et al., 2021). Based on low estimated heterozygosity from the *k*-mer analysis (see Results) and the selfing mating system of *P. recurvata*, purging of haplotypic duplications was skipped in the assembly (using the *-l0* flag). The resulting assembly graph was visualized using Bandage v0.8.1 (Wick et al., 2015) assembly errors using Inspector with the *--pacbio-hifi* flag (Chen et al., 2021). Foreign contaminants were identified in the assembly using BlobToolKit v4.3.5

(Challis et al., 2020), which identifies potential contaminants from raw read coverage, GC content, and BLAST hits. The HiFi reads were aligned to the initial assembly to calculate coverage using Minimap2 v2.28 with the *-map-hifi* flag (Li, 2018), and BLAST hits from the nucleotide database were generated using BLASTN with BLAST v2.13.0 (Altschul et al., 1990). This initial assembly (i.e., every contig >10 kbp) was assessed for completeness using 2326 benchmarking universal single-copy orthologs (BUSCOs) from the eudicots dataset *eudicots\_odb10* (1 August 2024) using BUSCO v5.6.1 (Manni et al., 2021).

The initial assembly consisted of 12 large contigs and 6449 smaller, trailing contigs (see Results; Table 1). Three approaches were taken to identify the sequence content of these contigs. First, the trailing contigs were mapped to our assembled *P. recurvata* reference plastome (see below) using Geneious Prime 2024.0.5 (<https://www.geneious.com>); any contigs aligned to our reference plastome were removed from the assembly. Second, publicly available mitochondrial DNA sequences from species in Boraginaceae in the National Center for Biotechnology Information (NCBI) nucleotide database (accessed June 2024) were mapped to the trailing contigs to identify contigs containing mitochondrial DNA. Finally, the trailing contigs were mapped to the 12 large contigs to identify any repetitive genomic sequences in the contigs. While the contigs that mapped to the chloroplast were removed from the initial assembly (generating the final assembly), the contigs that mapped to mitochondrial sequences and those that mapped to the genome were retained due to ambiguity in their placements. Upon uploading this final assembly to NCBI, the NCBI contamination screen was used to detect and remove any remaining foreign DNA (Astashyn et al., 2024).

The chloroplast genome was constructed de novo using the PacBio HiFi reads. Potential chloroplast reads were identified by aligning the raw HiFi reads to the *E. plantagineum* (Boraginaceae) reference plastome (GenBank accession: OL335188.1; Carvalho Leonardo et al., 2022) using Minimap2 v2.28 with default settings (Li, 2018). Reads that did not align to the reference were removed using SAMtools v1.10 (Li et al., 2009). The remaining, mapped reads had an exceptionally high estimated coverage (159,683×; see Results); accordingly, reads were randomly subsampled with SeqKit v2.8.1 (Shen et al., 2016) to approximately 150× coverage to mitigate assembly errors. These subsampled reads were then used to generate a circular *P. recurvata* chloroplast genome assembly using Hifiasm v0.16.1 with default settings (Cheng et al., 2021). The assembly was evaluated using Bandage v0.8.1 (Wick et al., 2015) and Geneious Prime 2024.0.5.

## Annotation

Repetitive elements were identified, classified, and soft-masked in the final assembly. First, a custom species-specific library of identified and classified repetitive element families was generated with RepeatModeler v2.0.3 using default

parameters (Flynn et al., 2020), which identifies known and unknown repetitive elements de novo using RECON v1.08 (Bao and Eddy, 2002) and RepeatScout v1.0.6 (Price et al., 2005). Distinct repeats were identified and soft-masked (using the *-xsmall* flag) in the assembly iteratively with RepeatMasker v4.1.3 (<https://www.repeatmasker.org/RepeatMasker/>), starting with simple repeats (specifying the *-noint* flag), then repeats identified from the *nrTEplants2020* curated repetitive sequence library (Contreras-Moreira et al., 2021), followed by known and unknown repeats from the *P. recurvata* repeat library. After repeats were masked in the final assembly, telomeric regions of the chromosomes were predicted using TeloExplorer from QuarTeT v1.1.6 with *-m 85* (Lin et al., 2023).

Gene structural annotation of the *P. recurvata* genome assembly was performed on the soft-masked assembly using the BRAKER3 annotation pipeline (Hoff et al., 2016, 2019; Brúna et al., 2021; Gabriel et al., 2021, 2024), which uses evidence from homologous proteins and transcripts to train AUGUSTUS (Stanke et al., 2006) and GeneMark-EP+ (Lomsadze et al., 2005; Brúna et al., 2020) to produce high-confidence gene models. Whole transcripts from the Iso-Seq library were aligned to the *P. recurvata* reference assembly using Minimap2 v2.28 with *-ax splice:hq* specified (Li, 2018). A custom protein database containing 456,224 peptides was constructed from the proteomes of 10 lamiids species with well-annotated genomes (Table S7 in Appendix S1). We used the *braker3\_lr* singularity container to accommodate the use of the *P. recurvata* Iso-Seq transcriptome as evidence (alongside the custom protein database). The resulting set of predicted proteins was assessed for completeness against the *eudicots\_odb10* (8 January 2024) benchmarking ortholog database with BUSCO v5.6.1 in *euk\_tran* mode (specifying *-m 'tran'*; Manni et al., 2021). Summary statistics describing the protein annotation were generated using the *agat\_sp\_statistics.pl* script from AGAT v1.0.0 (Dainat, 2022). The final protein set was functionally annotated using InterProScan v5.66-98.0 (Jones et al., 2014) with the *--goterms* flag to include gene ontology terms.

The chloroplast assembly was annotated using GeSeq v2.03 (Tillich et al., 2017). GeSeq utilizes HMMER v3.4 (<http://hmmer.org>), Chloë (Zhong, 2020), and BLAST (Altschul et al., 1990) to annotate the inverted repeat regions, protein-coding sequences, rRNAs, and tRNAs in plastid genomes. The map of the annotated *P. recurvata* chloroplast genome (Figure S7 in Appendix S2) was generated using OGDRAW v1.3.1 (Greiner et al., 2019).

## Synteny analysis

Regions of conserved gene synteny were identified among the assembled *P. recurvata* chromosomes and with the *E. plantagineum* chromosomes (GCA\_003412495.2), the only other chromosome-level assembly in Boraginaceae that has been annotated (Tang et al., 2020). While the *E. plantagineum* annotation described by Tang et al. (2020) is not

**TABLE 1** *Pectocarya recurvata* assembly and annotation statistics.

Metric	Assembly (initial)	Assembly (final)	12 Putative chromosomes
Coverage	264×	—	—
GC content (%)	37.8%	38.4%	35.1%
<b>Contiguity and completeness</b>			
Total length (bp)	393,226,675	215,950,562	158,303,007
Total length of contigs > 1 Mbp	158,303,007	158,303,007	158,303,007
Total number of contigs	6461	1531	12
Number of contigs > 1 Mbp	12	12	12
Length of longest contig (bp)	18,090,075	18,090,075	18,090,075
N50 (bp)	61,471	12,148,414	—
L50	451	8	—
N90 (bp)	27,701	36,944	—
L90	4656	561	—
Eudicots complete BUSCOs	2208 (94.9%)	2208 (94.9%)	2208 (94.9%)
Eudicots complete and single-copy BUSCOs	2051 (88.2%)	2052 (88.2%)	2054 (88.3%)
Eudicots duplicated BUSCOs	157 (6.7%)	156 (6.7%)	154 (6.6%)
Eudicots fragmented BUSCOs	25 (1.1%)	25 (1.1%)	25 (1.1%)
<b>Annotation</b>			
Total repetitive DNA	277,362,488 (70.54%)	99,482,072 (46.07%)	41,832,491 (26.43%)
Transposable elements	224,364,318 (57.06%)	75,748,802 (35.08%)	38,159,365 (24.11%)
Simple repeats	3,299,783 (0.84%)	2,223,469 (1.03%)	2,137,440 (1.35%)
Total number of gene models	—	47,030	30,655
Total number of proteins	—	49,558	33,020
Total coding region (bp)	—	51,374,723	40,679,859
Average gene length (bp)	—	1987	2538
Eudicots complete BUSCOs	—	2234 (96.0%)	2232 (95.9%)
Eudicots complete and single-copy BUSCOs	—	1920 (82.5%)	1913 (82.2%)
Eudicots duplicated BUSCOs	—	314 (13.5%)	319 (13.7%)
Eudicots fragmented BUSCOs	—	10 (0.4%)	10 (0.4%)
<b>De novo TE library</b>			
Known repeats	118	113	103
Unknown repeats	538	473	438
Total repeats	656	686	541

Note: BUSCOs = benchmarking universal single-copy orthologs; TE = transposable element.

publicly available, the raw transcripts used to annotate the reference genome are available on NCBI. Three of these available transcriptome libraries (SRR4034891, SRR4034890, and SRR7076848) were used to generate a draft gene structural annotation of the *E. plantagineum* assembly. Using the standard *braker3* singularity container, the same annotation protocol described above was followed

to incorporate short-read RNA-Seq transcriptome data as evidence of gene structure (Gabriel et al., 2024). Groups of orthologous genes from the resulting proteomes from each genome annotation were identified using OrthoFinder v2.5.4 with default settings (Emms and Kelly, 2019). The groups of shared orthologs identified by OrthoFinder were used to identify blocks of conserved gene order that cluster

into regions of synteny using the GENESPACE v1.4 R package with default settings (Lovell et al., 2022). The syntenic depth ratio of each syntenic block (collinear sets of genes >5) was calculated using MCscan (Python version), which depends on LAST (Tang et al., 2008; Kielbasa et al., 2011).

Additionally, evidence for ancient WGDs in the *P. recurvata* and *E. plantagineum* genomes was identified by analyzing the frequency distribution of synonymous divergence ( $K_s$ ) between paralogs for each genome individually.  $K_s$  was calculated for each pair of duplicate genes using wgd v2 (Chen et al., 2024). The  $K_s$  plot was filtered by only retaining syntenic duplicates, i.e., those duplicates that most represent duplication from a genome-wide multiplication rather than small-scale multiplications such as tandem duplicates, within wgd using i-ADHoRe v3.0 (Proost et al., 2012).

## RESULTS

### Assembly

The  $k$ -mer analysis using the HiFi reads predicted a haploid genome size of 152,502,236 bp, genome-wide heterozygosity of 0.0475%, and a repeat content length of 34,877,869 bp (comprising 22.9% of the genome) for the sequenced *P. recurvata* individual. The Hifiasm assembly contained 6573 contigs covering 394,090,231 bp with no structural or small-scale errors detected by Inspector (Chen et al., 2021). After removing contigs smaller than 10 kbp, 6461 contigs covering 393,226,675 bp remained in the initial assembly (Table 1). Twelve of these contigs were greater than 1 Mbp (total length = 158,303,007 bp), corresponding to the expected number of haploid chromosomes ( $n = 12$ ) in *P. recurvata* (Veno, 1979). Of the 6449 trailing contigs in the initial assembly, 4930 (76.4%) aligned to our *P. recurvata* plastome assembly (with 94.1% identical sequence) and were subsequently removed from the initial assembly, creating the final assembly. The Boraginaceae mitochondrial sequences from NCBI mapped to 678 unique contigs (10.5% of the trailing contigs), and 268 other contigs (4.4% of the trailing contigs) mapped to the 12 large contigs, all to the end of chromosome 12. No external contamination was detected in our final assembly through our BlobToolKit screen (Figure S9 in Appendix S2). The NCBI contamination screen detected one trailing contig that contained a small section (47 bp) of foreign DNA sequence on the end of one trailing contig, which was subsequently trimmed.

The final assembly contains 1531 contigs covering 215,950,562 bp and has an N50 of 12,148,414 bp and an L50 of 8, indicating a high level of contiguity (Table 1). The total length of the 12 leading contigs is 158,303,007 bp, which is similar to the estimated genome size (152.5 Mbp) from the  $k$ -mer analysis and accounts for 73.3% of the final assembly. A total of 94.9% of complete eudicot BUSCOs were found in these 12 contigs, and none of the BUSCOs that were

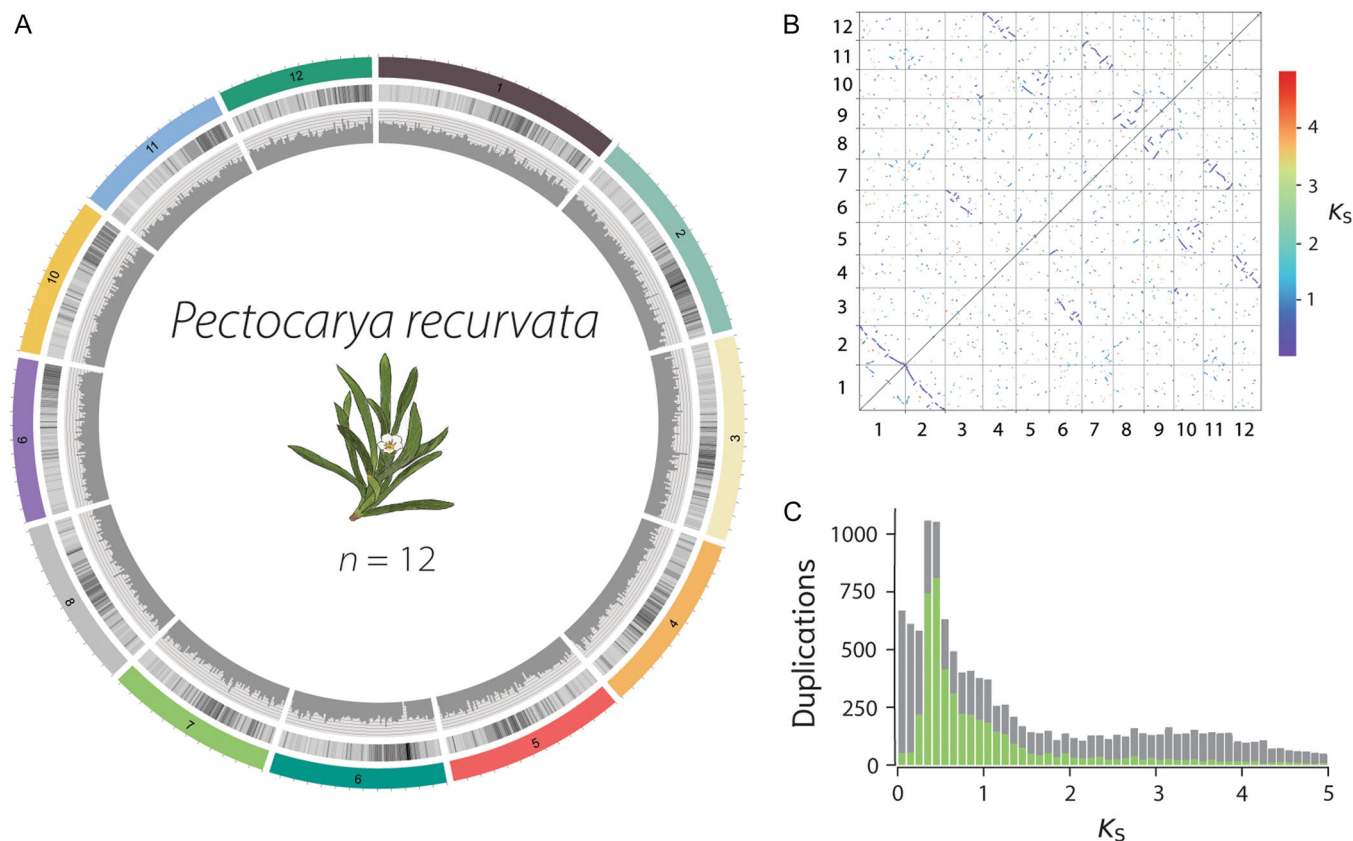
duplicated, fragmented, or missing from either BUSCO dataset were identified in the trailing contigs of the assembly (Table 1). Hereafter, we refer to these 12 leading contigs as the putative *P. recurvata* chromosomes, numbered 1–12 by length (Figure 2A, Table S2 in Appendix S1).

Telomeric regions consisting of greater than 85 tandem repeats of CCCTAAA, the typical plant telomeric repeat sequence (Peska and Garcia, 2020), were identified at both ends of the putative chromosomes 1–11, while 64 repeats were identified on one end of chromosome 12 and none were found on the other end. No other telomeric regions were identified in the assembly.

The chromosomes (158.3 Mbp) consist of 26.43% repetitive content identified by RepeatMasker (<https://www.repeatmasker.org/RepeatMasker/>), where 24.11% is attributed to transposable elements, 1.66% to tandem repeats, and 0.17% to small RNAs. The most common transposable elements are long terminal repeats (LTRs), which make up 6.29% of the chromosomes, while long interspersed nuclear elements (LINEs) make up 0.81%, short interspersed nuclear elements (SINES) make up 0.22%, and 15.38% are unclassified. DNA transposons comprise 1.40% of the chromosomes (Table S5 in Appendix S1). Notably, the repetitive content of the *P. recurvata* chromosomes (26.43%) is much lower than that reported in either of the described genome assemblies from Boraginaceae: the *L. erythrorhizon* assembly is 366.7 Mbp, of which 51.78% is repetitive content, and similarly the *E. plantagineum* assembly is 348.9 Mbp and 43.30% repetitive content (Table S6 in Appendix S1) (Auber et al., 2020; Tang et al., 2020).

### Genome size estimation: Flow cytometry, $k$ -mer analysis, and inference of ploidy level

Our genome size measurements of six *P. recurvata* individuals by flow cytometry revealed putative cytotype variation in our sampled population. Two of the individuals had an estimated genome size of  $1C = 0.201 \pm 0.008$  pg ( $196.9 \pm 7.4$  Mbp; Table S9 in Appendix S1), which is somewhat larger than (but generally consistent with) our  $k$ -mer-based genome size estimate (152.5 Mbp) and the length of the 12 chromosomal contigs in our assembly (158.3 Mbp). However, the four other samples analyzed had a mean genome size of  $1C = 0.395 \pm 0.001$  pg ( $385.8 \pm 1.3$  Mbp; Table S9 in Appendix S1), which is approximately twice the smaller estimates. We ran SmudgePlot (Ranallo-Benavidez et al., 2020) to detect ploidy level in our sequenced genome from relative  $k$ -mer coverages, and this indicated that the individual sequenced for the genome assembly was likely tetraploid ( $2n = 48$ ) (Figures S2 and S8 in Appendix S2). Because the  $k$ -mer analysis also estimated a very low level of heterozygosity, consistent with high levels of self-fertilization, this raises the possibility that our assembly represents a haploid assembly of *P. recurvata* that collapsed two, putatively highly similar, subgenomes of a recent autotetraploid.



**FIGURE 2** Analysis of self-self synteny with the *Pectocarya recurvata* reference chromosomes suggests a recent whole genome duplication in Boraginaceae. (A) The assembled *P. recurvata* chromosomes ( $n = 12$ ) labeled 1–12 by length. Transposable element density is illustrated by the gray heatmaps in the first inner circle. Gene density is depicted by the histograms in the next circle. Figure produced with Circos (Krzywinski et al., 2009). (B) A dot plot of self-self synteny in the *P. recurvata* chromosomes, colored according to the synonymous divergence ( $K_s$ ) between each pair of duplicated genes. (C) A frequency distribution of the  $K_s$  between paralogs in the *P. recurvata* chromosome. The gray background distribution includes every paralog, while the green distribution only includes syntenic paralogs.

## Annotation

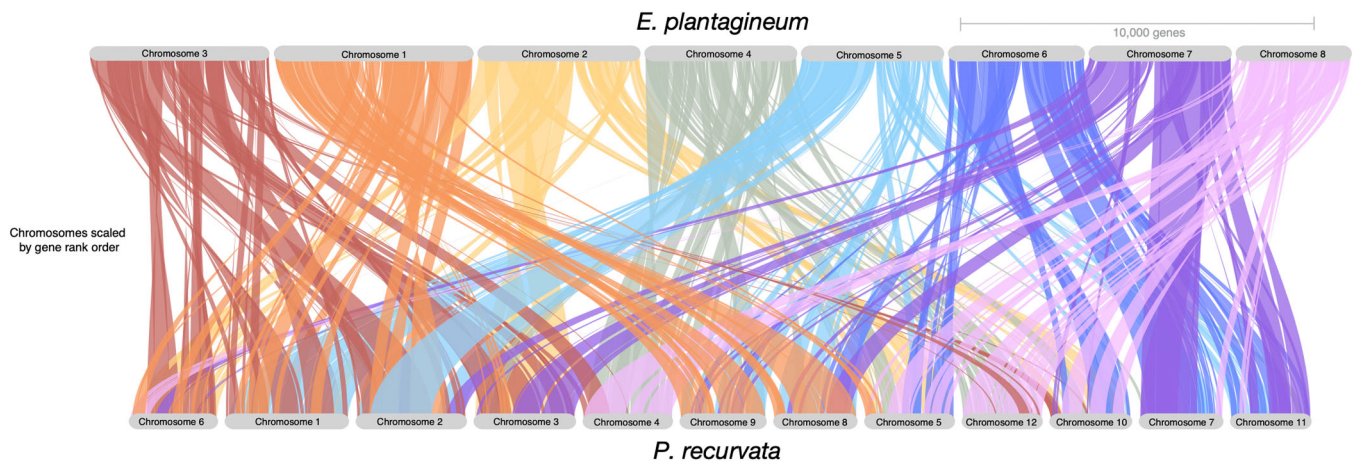
The structural gene annotation of the *P. recurvata* reference chromosomes contains 30,655 predicted gene models (mean gene length = 2538 bp) that code for 33,020 proteins. On average, the predicted genes are 2538 bp long and contain five exons. In total, the predicted genes encompass 77,831,286 bp, of which 40,679,859 bp is attributed to coding sequence. The comparison of the *P. recurvata* protein annotation against the *eudicot\_odb10* BUSCO dataset found 95.5% complete BUSCOs ( $n = 2326$ ), of which 1913 (85.7%) are single copy (Table 1). Of the 33,020 predicted proteins, 93.8% (30,979) were functionally annotated with InterProScan (Jones et al., 2014).

## Gene synteny

Self-self comparison of gene order and content in the *P. recurvata* chromosomes revealed several duplicated collinear gene segments (Figure 2B). The  $K_s$  between paralogs in the *P. recurvata* genome revealed a

distinctive peak in the number of syntenic paralogs at  $K_s = 0.4–0.5$ , reflective of an ancient WGD in the evolutionary history of *P. recurvata* (Figure 2C). There is also a distinctive peak at  $K_s = 0.4–0.5$  in the  $K_s$  distribution for *E. plantagineum*, as well as a distinctive peak in syntenic paralogs at  $K_s = 0.1$ , reflective of an additional WGD event in the evolutionary history of *E. plantagineum* (Figure S3 in Appendix S2).

By comparing the gene order of the *P. recurvata* and *E. plantagineum* genomes, we identified a 2:3 syntenic relationship between the genomes of these species (Figure 3, Figure S6 in Appendix S2). Together, the *P. recurvata* and *E. plantagineum* reference chromosomes have 70,183 annotated genes. Nearly all of these genes were placed in 19,150 orthogroups (89.0% of all genes), of which 15,906 orthogroups were shared between the two species. We identified 6339 single-copy orthologs (accounting for 40% of the *P. recurvata* genome). The shared orthogroups form 1252 syntenic blocks (collinear sets of genes >5) that cluster into 690 syntenic regions (Figure 3); nearly half (47%) of the syntenic blocks in the *P. recurvata* genome are found in duplicate in the *E. plantagineum* genome. In the *E. plantagineum* genome, 33% of syntenic blocks are found in



**FIGURE 3** Synteny between the *Pectocarya recurvata* and *Echium plantagineum* reference chromosomes.

duplicate in the *P. recurvata* genome and 28% are found in triplicate (Figures S5 and S6 in Appendix S2). The excess number of duplicate genes shared in both genomes suggests that these species share a WGD in their evolutionary history. The excess of triplicate genes in the *E. plantagineum* genome indicates that an additional, independent WGD has occurred in *E. plantagineum*.

### Chloroplast assembly and annotation

Across 1,869,123 reads (mean length = 12,814 bp), 23,952,585,969 bp successfully mapped to the *E. plantagineum* reference plastome (149,776 bp) with an estimated 159,683 $\times$  coverage. The initial Hifiasm assembly from these reads, subsampled to 150 $\times$  per-base coverage, yielded three contigs totaling 187,628 bp, consisting of one 150,172 bp circular contig and two smaller linear contigs (24,314 bp and 13,142 bp). To investigate these linear contigs, both were aligned to the large circular contig using Geneious Prime 2024.0.5. The linear contigs each share 100% sequence identity with regions on the circular contig and were subsequently removed from the assembly. Thus, the final *P. recurvata* chloroplast genome (hereafter “plastome”) assembly solely consists of the large circular contig from the initial assembly.

The *P. recurvata* plastome has 37.4% GC content and a quadripartite structure typical of chloroplast genomes, consisting of a small single-copy region (17,143 bp; 31.1% GC), two inverted repeats (IRa: 25,446 bp; IRb: 25,467 bp; 43.1% GC), and a large single-copy region (82,096 bp; 35.3% GC). GeSeq annotation of this circular contig predicted 130 unique genes in the plastome including eight rRNAs, 37 tRNAs, and 85 protein-coding genes (Figure S7 in Appendix S2). The length and gene content of the *P. recurvata* plastid genome assembly are comparable to that of the *E. plantagineum* plastome (Table S10 in Appendix S1) (Carvalho Leonardo et al., 2022).

### DISCUSSION

We assembled and annotated the first chromosome-level reference genome of *P. recurvata*. The final genome assembly (216.0 Mbp) is highly contiguous (N50 = 12.1 Mbp) and complete (94.9% eudicot BUSCOs identified). This completeness is reflected in the 12 reference chromosomes, which make up 158.3 Mbp of the assembly. These chromosomes have 23 out of 24 putative telomeres and contain 30,655 predicted genes that code for 33,020 proteins that comprise highly complete structural and function annotations (95.9% eudicot BUSCOs were identified; 93.8% were assigned functions by InterProScan). Many large syntenic regions were identified between *P. recurvata* and *E. plantagineum*, further validating the quality of our gene annotation.

The  $K_s$  of putative paralogs in the *P. recurvata* and *E. plantagineum* genomes suggests that a WGD event occurred in recent evolutionary history at  $K_s \approx 0.4$ – $0.5$  (Figure 2C). Our finding is congruent with a previously described WGD in the core Boraginales (from analyses of the *E. plantagineum* and *Cordia subcordata* Lam. genomes) at  $K_s \approx 0.417$ , which putatively occurred before the divergence of the Boraginoideae and Cynoglossoideae subfamilies in Boraginaceae (Tang et al., 2020; Chen et al., 2023). This WGD is likely the source of the large number of genes found reciprocally in duplicate between the *P. recurvata* and *E. plantagineum* genomes. While Ren et al. (2018) and Tang et al. (2020) also found evidence of another, older WGD in the core Boraginales at  $K_s \approx 0.939$ , there is no distinctive peak at  $K_s = 0.9$  in our analyses. Interestingly, some recent analyses using a combination of available genomes and transcriptomes do not identify the more recent WGD at  $K_s \approx 0.417$  that is evidenced by the *P. recurvata* genome and other analyses (Ren et al., 2018; Auber et al., 2020; Tang et al., 2020; Zhang et al., 2020). However, Ren et al. (2018) did find evidence from the *Trigonotis peduncularis* (Trevir.) Benth. ex Baker & S. Moore transcriptome (Myosotidae subtribe, Cynoglossoideae tribe) of a very recent WGD at  $K_s \approx 0.13$ , while Tang et al. (2020) did not. Moreover,

we find evidence that *E. plantagineum* has undergone a more recent, independent WGD from *P. recurvata* (Figures S5 and S6 in Appendix S2). Whole genome duplication has played an instrumental role in the evolution the angiosperms—on average, every angiosperm has experienced 3.5 rounds of WGD (and subsequent re-diploidization) in their evolutionary history (One Thousand Plant Transcriptomes Initiative, 2019; McKibben et al., 2024). It remains a prominent goal in plant biology to determine the phylogenetic placement of WGDs and their role in plant evolution (Van de Peer et al., 2017; Barker et al., 2024). Additional genomic resources in Boraginales will be helpful for confidently placing these disputed WGDs and understanding their impact on the evolution of this clade.

In estimating the genome size and sequencing the *P. recurvata* genome, we discovered putative cytotype variation in *P. recurvata* that we infer is the result of very recent autopolyploidy. Our genome size estimate from flow cytometry was  $1C = 196.9$  Mbp for two individuals, which is larger but consistent with the size of our highly complete assembly of  $n = 12$  chromosomes (159.3 Mbp), yet we also observed  $1C = 385.8$  Mbp for four different individuals from the same population. SmudgePlot analysis of our genomic data indicated that the sequenced individual was likely a very recent tetraploid with very low divergence among duplicated  $k$ -mers. If this is true, we infer that the assembly collapsed the two highly similar subgenomes, resulting in a reference assembly similar in size to the observed diploids. Prior karyological work by Veno (1979) observed only  $2n = 24$  in individuals of *P. recurvata* from five populations across Arizona and Southern California (also see Ward, 1983). Species in the genus *Pectocarya* follow a polyploid series of  $2n = 24$  (e.g., *P. recurvata*, *P. heterocarpa* I.M. Johnst.),  $2n = 48$  (e.g., *P. platycarpa* Munz & I.M. Johnst., *P. anisocarpa* Veno), and  $2n = 72$  (*P. linearis* (Ruiz & Pav.) DC.), suggesting that recent WGD has played a role in the evolutionary history of this genus (Veno, 1979). Autopolyploidy is common among plants, yet autopolyploid individuals and species are often undistinguished because they appear morphologically very similar to their diploid progenitors (Soltis and Soltis, 2000; Soltis et al., 2007; Barker et al., 2016; Lv et al., 2024). Moreover, the short-term ecological consequences of autopolyploidy remain elusive (Stebbins, 1940; Soltis et al., 2014; Assour et al., 2024). Quantifying the geographic extent of cytotype variation in *P. recurvata* would lend insight into the distribution of genetic variation in this species and also provide the opportunity to explore associations between autopolyploidy and relevant ecological factors (Servick et al., 2015; Visger et al., 2016; Barker et al., 2024).

In addition to our discovery of putative ploidy variation in *P. recurvata*, we uncovered one of the smallest recorded genome sizes in Boraginaceae to date. There are currently 57 primary genome size estimates for taxa in Boraginaceae recorded in the Kew Plant DNA C-values Database (Release 7.1; <https://cvalues.science.kew.org/> [accessed September

2024]), ranging from  $1C = 313.0$  Mbp (*Echium bonnetii* Coincy) to  $1C = 10,878.0$  Mbp (*Alkanna leiocarpa* Rech.f.) (mean  $1C = 1599.07$  Mbp). A more recent comprehensive survey of Boraginaceae genome size estimates from species occurring in the Czech Republic (Kobrlóvá and Hroneš, 2019) recorded genome sizes that range from approximately  $1C = 270$  Mbp for *Myosotis sylvatica* Ehrh. to  $1C = 16,000$  Mbp for *Lycopsis arvensis* L. At  $1C = 196.9$  Mbp, our flow cytometry measurement is smaller than all current observations in the family. Genome size can vary greatly between related species, in part due to WGD and subsequent re-diploidization (Li et al., 2021) as well as transposable element activity (Vitte and Panaud, 2005; El Baidouri and Panaud, 2013). In Boraginaceae, taxa in the Boraginoideae subfamily are more often diploid than taxa in Cynoglossoideae, yet have larger chromosomes and genome sizes, suggesting that transposable elements might have played an important role in generating the 35-fold variation in genome sizes across clades within this group (Kobrlóvá and Hroneš, 2019). Notably, we found that the repetitive content of the *P. recurvata* chromosomes (26.43%) is much lower than that reported in either of the described genome assemblies from Boraginaceae, both members of the Boraginoideae subfamily: the *L. erythrorhizon* assembly is 366.7 Mbp, of which 51.78% is repetitive content, and similarly the *E. plantagineum* assembly is 348.9 Mbp, with 43.30% repetitive content (Table S6 in Appendix S1) (Auber et al., 2020; Tang et al., 2020).

The adaptive significance of plants' genome sizes remains under investigation (Gregory, 2001; Knight and Beaulieu, 2008; Veselý et al., 2012; Bureš et al., 2024). The large genome constraint hypothesis asserts that larger genome sizes are maladaptive due to the physiological consequences associated with larger cell sizes (Knight et al., 2005; Veselý et al., 2020); for example, larger stomatal guard cells open and close more slowly, resulting in lower water use efficiency (Drake et al., 2013). Such small changes in physiology can compound and decrease fitness in different climates (Bureš et al., 2024). Herbaceous angiosperms with larger genomes are associated with heightened extinction risk, regardless of climate (Soto Gomez et al., 2024). Moreover, smaller genome sizes are frequently associated with faster rates of cell division that lead to faster development and flowering times (Grime and Mowforth, 1982; Bilinski et al., 2018; Cacho et al., 2021; Cang et al., 2024). Desert annual plants, such as *P. recurvata*, are under strong selective pressure to rapidly grow and reproduce after germinating to take advantage of ephemeral moisture following precipitation events (Huxman et al., 2008; Barron-Gafford et al., 2013). Past selection has favored a *P. recurvata* phenotype that is highly water use efficient relative to other members of the Sonoran Desert winter annual community (Angert et al., 2007; Huxman et al., 2008). Our finding of a small genome size in this species raises the question of whether selection might also favor a faster development time through genome size reduction.

While the Sonoran Desert winter annual plants are a well-developed study system in ecology and evolutionary biology, only a handful of reference genomes have been assembled for these plants. Of the 51 most common Sonoran Desert winter annuals found at Tumamoc Hill (Tucson, Arizona, USA; see Ge et al., 2019), only *Eschscholzia californica* Cham. (Hori et al., 2018), *Erodium texanum* A. Gray (NCBI accession: GCA\_036897725.1), *Sisymbrium irio* L. (Haudry et al., 2013), and now *P. recurvata*, have reference genomes. Much is known about how species coexistence in this community is shaped by variation in species' positions along a physiological trade-off between water use efficiency and relative growth rate (Chesson, 2000; Angert et al., 2009); however, little is known about the genetic architecture underlying these traits and how genetic differences may combine to manifest these ecological dynamics (Kimball et al., 2013; Angert et al., 2014). In the context of anthropogenic climate change, it is critical to understand the genetic basis of relevant ecological traits being selected on by climate change in order to assess species' extinction risk and downstream ecological consequences (Scheffers et al., 2016; Whiting et al., 2024). The Sonoran Desert winter annuals have experienced overall declines in abundance due to recent anthropogenic change, and the community is simultaneously shifting to favor species with higher water use efficiency (Kimball et al., 2010; Huxman et al., 2013). The addition of a *P. recurvata* genome will facilitate investigations into the genetic mechanisms driving responses to climate in these plants.

## Conclusions

*Pectocarya recurvata* is an important model system for a suite of ecological and evolutionary biology studies. Our generation of a new chromosome-level reference genome assembly and annotation for *P. recurvata* presents the possibility of new lines of inquiry, including exploring the effect of polyploidy and genome size variation on ecological interactions and climate adaptation in this system, as well as elucidating the genetic basis of the physiological trade-off between water use efficiency and relative growth rate exhibited by the Sonoran Desert winter annuals. Establishing such mechanistic links between genetic changes and ecology is not only a key goal in evolutionary ecology but an urgent task to evaluate responses to global climate change.

## AUTHOR CONTRIBUTIONS

K.M.D. and D.L.V. conceived of the study. P.C.N. assembled and annotated the reference genome. P.C.N. and J.A.P. conducted the synteny and WGD analyses. P.C.N. wrote the manuscript and J.A.P., K.M.D., and D.L.V. provided feedback. All authors approved the final version of the manuscript.

## ACKNOWLEDGMENTS

The authors thank the Arizona Genomics Institute for input on sequencing strategy and methodology. The authors also

thank members of the Dlugosch and Barker labs for feedback on the manuscript and Ulises Hernández (Department of Ecology and Evolutionary Biology, University of Arizona) for providing the Spanish translation of our abstract. This work was supported by funding from the U.S. National Science Foundation (grants #1256792 to D.L.V., #1750280 to K.M.D., and #2022055 providing support for P.C.N.) and from the U.S. Department of Agriculture (grants #2023-67013-40169 to K.M.D. and #2024-67012-43394 to J.A.P.).

## DATA AVAILABILITY STATEMENT

The raw genomic sequences (SRR29849658), genome assembly (GCA\_046736585.1), and plant tissue sample description (SAMN42375859) are publicly available under the NCBI BioProject PRJNA1133066. Raw transcriptomic sequences (SRR30591463) are linked to the same BioProject, and the chloroplast genome is included in the full genome assembly. The structural annotation and functional annotation of the final assembly and chromosomes, as well as relevant data and output from repeat masking, *k*-mer analysis, flow cytometry, OrthoFinder, GENESPACE, and wgd, are archived in a public repository on Zenodo (DOI: 10.5281/zenodo.14714947; Northing, 2025). The scripts used to generate the assembly and annotation as well as conduct the analyses are available on GitHub ([https://github.com/poppynorthing/PERE\\_genome/](https://github.com/poppynorthing/PERE_genome/)).

## ORCID

Poppy C. Northing  <https://orcid.org/0009-0005-1665-7685>

Jessie A. Pelosi  <https://orcid.org/0000-0002-2861-3445>

D. Lawrence Venable  <https://orcid.org/0000-0001-8413-8872>

Katrina M. Dlugosch  <https://orcid.org/0000-0002-7302-6637>

## REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Angert, A. L., T. E. Huxman, G. A. Barron-Gafford, K. L. Gerst, and D. L. Venable. 2007. Linking growth strategies to long-term population dynamics in a guild of desert annuals. *Journal of Ecology* 95: 321–331.
- Angert, A. L., T. E. Huxman, P. Chesson, and D. L. Venable. 2009. Functional tradeoffs determine species coexistence via the storage effect. *Proceedings of the National Academy of Sciences, USA* 106: 11641–11645.
- Angert, A. L., S. Kimball, M. Peterson, T. E. Huxman, and D. L. Venable. 2014. Phenotypic constraints and community structure: Linking trade-offs within and among species. *Evolution* 68: 3149–3165.
- Assour, H. R., T.-L. Ashman, and M. M. Turcotte. 2024. Neopolyploidy-induced changes in giant duckweed (*Spirodela polyrhiza*) alter herbivore preference and performance and plant population performance. *American Journal of Botany* 111: e16301.
- Astashyn, A., E. S. Tvedte, D. Sweeney, V. Sapojnikov, N. Bouk, V. Joukov, E. Mozes, et al. 2024. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biology* 25: 60.
- Auber, R. P., T. Suttiyut, R. M. McCoy, M. Ghaste, J. W. Crook, A. L. Pendleton, J. R. Widhalm, and J. H. Wisecaver. 2020. Hybrid de novo genome assembly of red gromwell (*Lithospermum erythrorhizon*) reveals evolutionary insight into shikonin biosynthesis. *Horticulture Research* 7: 82.

- Bao, Z., and S. R. Eddy. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research* 12: 1269–1276.
- Barker, M. S., N. Arrigo, A. E. Baniaga, Z. Li, and D. A. Levin. 2016. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* 210: 391–398.
- Barker, M. S., Y. Jiao, and K. L. Glennon. 2024. Doubling down on polyploid discoveries: Global advances in genomics and ecological impacts of polyploidy. *American Journal of Botany* 111: e16395.
- Barron-Gafford, G. A., A. L. Angert, D. L. Venable, A. P. Tyler, K. L. Gerst, and T. E. Huxman. 2013. Photosynthetic temperature responses of co-occurring desert winter annuals with contrasting resource-use efficiencies and different temporal patterns of resource utilization may allow for species coexistence. *Journal of Arid Environments* 91: 95–103.
- Bilinski, P., P. S. Albert, J. J. Berg, J. A. Birchler, M. N. Grote, A. Lorant, J. Quezada, et al. 2018. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genetics* 14: e1007162.
- Bremer, B., K. Bremer, N. Heidari, P. Erixon, R. G. Olmstead, A. A. Anderberg, M. Källersjö, and E. Barkhordarian. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Molecular Phylogenetics and Evolution* 24: 274–301.
- Brúna, T., A. Lomsadze, and M. Borodovsky. 2020. GeneMark-EP+: Eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics* 2: lqaa026.
- Brúna, T., K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky. 2021. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* 3: lqaa108.
- Bureš, P., T. L. Elliott, P. Veselý, P. Šmarda, F. Forest, I. J. Leitch, E. Nic Lughadha, et al. 2024. The global distribution of angiosperm genome size is shaped by climate. *New Phytologist* 242: 744–759.
- Cacho, N. I., P. J. McIntyre, D. J. Kliebenstein, and S. Y. Strauss. 2021. Genome size evolution is associated with climate seasonality and glucosinolates, but not life history, soil nutrients or range size, across a clade of mustards. *Annals of Botany* 127: 887–902.
- Cang, F. A., S. R. Welles, J. Wong, M. Ziaee, and K. M. Dlugosch. 2024. Genome size variation and evolution during invasive range expansion in an introduced plant. *Evolutionary Applications* 17: e13624.
- Carvalho Leonardo, I., M. T. Barreto Crespo, J. Capelo, and F. Bustos Gaspar. 2022. The complete plastome of *Echium plantagineum* L. (Boraginaceae), the first chloroplast genome belonging to the *Echium* genus. *Mitochondrial DNA Part B, Resources* 7: 1154–1156.
- Chacón, J., F. Luebert, H. H. Hilger, S. Ovchinnikova, F. Selvi, L. Cecchi, C. M. Guilliams, et al. 2016. The borage family (Boraginaceae s.str.): A revised infrafamilial classification based on new phylogenetic evidence, with emphasis on the placement of some enigmatic genera. *Taxon* 65: 523–546.
- Chacón, J., F. Luebert, and M. Weigend. 2017. Biogeographic events are not correlated with diaspore dispersal modes in Boraginaceae. *Frontiers in Ecology and Evolution* 5: 26.
- Challis, R., E. Richards, J. Rajan, G. Cochrane, and M. Blaxter. 2020. BlobToolKit: Interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics* 10: 1361–1374.
- Chen, H., A. Zwaenepoel, and Y. Van de Peer. 2024. wgd v2: A suite of tools to uncover and date ancient polyploidy and whole-genome duplication. *Bioinformatics (Oxford, England)* 40: btac272.
- Chen, Y., Y. Zhang, A. Y. Wang, M. Gao, and Z. Chong. 2021. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biology* 22: 312.
- Chen, Y.-L., Z.-F. Wang, S.-G. Jian, H.-M. Liao, and D.-M. Liu. 2023. Genome assembly of *Cordia subcordata*, a coastal protection species in tropical coral islands. *International Journal of Molecular Sciences* 24: 16273.
- Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18: 170–175.
- Chesson, P. 2000. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* 31: 343–366.
- Chesson, P., R. L. E. Gebauer, S. Schwinning, N. Huntly, K. Wiegand, M. S. K. Ernest, A. Sher, et al. 2004. Resource pulses, species interactions, and diversity maintenance in arid and semi-arid environments. *Oecologia* 141: 236–253.
- Contreras-Moreira, B., C. V. Filippi, G. Naamati, C. García Girón, J. E. Allen, and P. Flicek. 2021. K-mer counting and curated libraries drive efficient annotation of repeats in plant genomes. *Plant Genome* 14: e20143.
- Dainat, J. 2022. AGAT: Another Gff analysis toolkit to handle annotations in any GTF/GFF format (Version v0.0.0). Available at Zenodo repository: <https://doi.org/10.5281/zenodo.3552717> [accessed 18 April 2025].
- Darwin Tree of Life Project Consortium. 2022. Sequence locally, think globally: The Darwin Tree of Life Project. *Proceedings of the National Academy of Sciences, USA* 119: e2115642118.
- Doležel, J., and J. Bartoš. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* 95: 99–110.
- Doležel, J., J. Greilhuber, and J. Suda. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2: 2233–2244.
- Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Drake, P. L., R. H. Froend, and P. J. Franks. 2013. Smaller, faster stomata: Scaling of stomatal size, rate of response, and stomatal conductance. *Journal of Experimental Botany* 64: 495–505.
- El Baidouri, M., and O. Panaud. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biology and Evolution* 5: 954–965.
- Emms, D. M., and S. Kelly. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 238.
- Flynn, J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences, USA* 117: 9451–9457.
- Gabriel, L., K. J. Hoff, T. Brúna, M. Borodovsky, and M. Stanke. 2021. TSEBRA: Transcript selector for BRAKER. *BMC Bioinformatics* 22: 566.
- Gabriel, L., T. Brúna, K. J. Hoff, M. Ebel, A. Lomsadze, M. Borodovsky, and M. Stanke. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *Genome Research* 34: 769–777.
- Ge, X.-Y. M., J. P. Scholl, U. Basinger, T. E. Huxman, and D. L. Venable. 2019. Functional trait trade-off and species abundance: Insights from a multi-decadal study. *Ecology Letters* 22: 583–592.
- Gottschling, M., N. Diane, H. H. Hilger, and M. Weigend. 2004. Testing hypotheses on disjunctions present in the primarily woody Boraginales: Ehretiaceae, Cordiaceae, and Heliotropiaceae, inferred from ITS1 sequence data. *International Journal of Plant Sciences* 165: S123–S135.
- Gregory, T. R. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews of the Cambridge Philosophical Society* 76: 65–101.
- Greiner, S., P. Lehwark, and R. Bock. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47: W59–W64.
- Grime, J. P., and M. A. Mowforth. 1982. Variation in genome size—An ecological interpretation. *Nature* 299: 151–153.
- Guilliams, C. M., K. E. Hasenstab-Lehman, M. E. Mabry, and M. G. Simpson. 2017. Memoirs of a frequent flier: Phylogenomics reveals 18 long-distance dispersals between North America and South America in the popcorn flowers (Asteraceae). *American Journal of Botany* 104: 1717–1728.
- Guilliams, C. M., and R. B. Kelley. 2021. *Pectocarya recurvata*. In Jepson Flora Project [eds.], Jepson eFlora, Revision 9. Available at: [https://ucjeps.berkeley.edu/eflora/eflora\\_display.php?tid=36533/](https://ucjeps.berkeley.edu/eflora/eflora_display.php?tid=36533/) [accessed 26 August 2024].

- Harris, A. J., S. Ickert-Bond, and A. Rodríguez. 2018. Long distance dispersal in the assembly of floras: A review of progress and prospects in North America: Long distance dispersal of plants in North America. *Journal of Systematics and Evolution* 56: 430–448.
- Haudry, A., A. E. Platts, E. Vello, D. R. Hoen, M. Leclercq, R. J. Williamson, E. Forczek, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics* 45: 891–898.
- Hoff, K. J., S. Lange, A. Lomsadze, M. Borodovsky, and M. Stanke. 2016. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics (Oxford, England)* 32: 767–769.
- Hoff, K. J., A. Lomsadze, M. Borodovsky, and M. Stanke. 2019. Whole-genome annotation with BRAKER. *Methods in Molecular Biology* 1962: 65–95.
- Hori, K., Y. Yamada, R. Purwanto, Y. Minakuchi, A. Toyoda, H. Hirakawa, and F. Sato. 2018. Mining of the uncharacterized cytochrome P450 genes involved in alkaloid biosynthesis in California poppy using a draft genome sequence. *Plant & Cell Physiology* 59: 222–233.
- Huxman, T. E., K. A. Snyder, D. Tissue, A. J. Leffler, K. Ogle, W. T. Pockman, D. R. Sandquist, et al. 2004. Precipitation pulses and carbon fluxes in semiarid and arid ecosystems. *Oecologia* 141: 254–268.
- Huxman, T. E., G. Barron-Gafford, K. L. Gerst, A. L. Angert, A. P. Tyler, and D. L. Venable. 2008. Photosynthetic resource-use efficiency and demographic variability in desert winter annual plants. *Ecology* 89: 1554–1563.
- Huxman, T. E., S. Kimball, A. L. Angert, J. R. Gremer, G. A. Barron-Gafford, and D. L. Venable. 2013. Understanding past, contemporary, and future dynamics of plants, populations, and communities using Sonoran Desert winter annuals. *American Journal of Botany* 100: 1369–1380.
- Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, et al. 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics (Oxford, England)* 30: 1236–1240.
- Kielbasa, S. M., R. Wan, K. Sato, P. Horton, and M. C. Frith. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Research* 21: 487–493.
- Kimball, S., A. L. Angert, T. E. Huxman, and D. L. Venable. 2010. Contemporary climate change in the Sonoran Desert favors cold-adapted species. *Global Change Biology* 16: 1555–1565.
- Kimball, S., J. R. Gremer, T. E. Huxman, D. L. Venable, and A. L. Angert. 2013. Phenotypic selection favors missing trait combinations in co-existing annual plants. *American Naturalist* 182: 191–207.
- Knight, C. A., N. A. Molinari, and D. A. Petrov. 2005. The large genome constraint hypothesis: Evolution, ecology and phenotype. *Annals of Botany* 95: 177–190.
- Knight, C. A., and J. M. Beaulieu. 2008. Genome size scaling through phenotype space. *Annals of Botany* 101: 759–766.
- Kobrlóvá, L., and M. Hroneš. 2019. First insights into the evolution of genome size in the borage family: A complete data set for Boraginaceae from the Czech Republic. *Botanical Journal of the Linnean Society* 189: 115–131.
- Kokot, M., M. Długosz, and S. Deorowicz. 2017. KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics* 33: 2759–2761.
- Krzywinski, M. I., J. E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* 19: 1639–1645.
- Li, H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094–3100.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Li, Z., M. T. W. McKibben, G. S. Finch, P. D. Blischak, B. L. Sutherland, and M. S. Barker. 2021. Patterns and processes of diploidization in land plants. *Annual Review of Plant Biology* 72: 387–410.
- Lin, Y., C. Ye, X. Li, Q. Chen, Y. Wu, F. Zhang, R. Pan, et al. 2023. quarTeT: A telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture Research* 10: uhad127.
- Lomsadze, A., V. Ter-Hovhannisyan, Y. O. Chernoff, and M. Borodovsky. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research* 33: 6494–6506.
- Lovell, J. T., A. Sreedasyam, M. E. Schranz, M. Wilson, J. W. Carlson, A. Harkess, D. Emms, et al. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* 11: e78526.
- Luebert, F., L. Cecchi, M. W. Frohlich, M. Gottschling, C. M. Williams, K. E. Hasenstab-Lehman, H. H. Hilger, et al. 2016. Familial classification of the Boraginales. *Taxon* 65: 502–522.
- Lv, Z., C. Addo Nyarko, V. Ramtekey, H. Behn, and A. S. Mason. 2024. Defining autopolyploidy: Cytology, genetics, and taxonomy. *American Journal of Botany* 111: e16292.
- Mabry, M. E., and M. G. Simpson. 2018. Evaluating the monophyly and biogeography of *Cryptantha* (Boraginaceae). *Systematic Botany*. 43: 53–76.
- Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov. 2021. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* 38: 4647–4654.
- McKibben, M. T. W., G. Finch, and M. S. Barker. 2024. Species-tree topology impacts the inference of ancient whole-genome duplications across the angiosperm phylogeny. *American Journal of Botany* 111: e16378.
- Moore, M. J., A. Tye, and R. K. Jansen. 2006. Patterns of long-distance dispersal in *Tiquilia* subg. *Tiquilia* (Boraginaceae): Implications for the origins of amphitropical disjuncts and Galápagos Islands endemics. *American Journal of Botany* 93: 1163–1177.
- Mulroy, T. W., and P. W. Rundel. 1977. Annual plants: Adaptations to desert environments. *Bioscience* 27: 109–114.
- Nathan, R. 2006. Long-distance dispersal of plants. *Science* 313: 786–788.
- Northing, P. C. 2025. *Pectocarya recurvata* genome assembly and annotation. Available at: Zenodo repository: <https://doi.org/10.5281/zenodo.13851912> [posted 21 January 2025; accessed 22 April 2025].
- Noy-Meir, I. 1973. Desert ecosystems: Environment and producers. *Annual Review of Ecology, Evolution, and Systematics* 4: 25–51.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.
- Pake, C. E., and D. L. Venable. 1995. Is coexistence of Sonoran desert annuals mediated by temporal variability reproductive success. *Ecology* 76: 246–261.
- Pake, C. E., and D. L. Venable. 1996. Seed banks in desert annuals: Implications for persistence and coexistence in variable environments. *Ecology* 77: 1427–1435.
- Peska, V., and S. Garcia. 2020. Origin, diversity, and evolution of telomere sequences in plants. *Frontiers in Plant Science* 11: 117.
- Price, A. L., N. C. Jones, and P. A. Pevzner. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1): i351–i358.
- Proost, S., J. Fostier, D. De Witte, B. Dhoedt, P. Demeester, Y. Van de Peer, and K. Vandepoele. 2012. i-ADHoRe 3.0: Fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Research* 40: e11.
- Ranallo-Benavidez, T. R., K. S. Jaron, and M. C. Schatz. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11: 1432.
- Raven, P. H. 1963. Amphitropical relationships in the floras of North and South America. *The Quarterly Review of Biology* 38: 151–177.
- Ren, R., H. Wang, C. Guo, N. Zhang, L. Zeng, Y. Chen, H. Ma, and J. Qi. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Molecular Plant* 11: 414–428.
- Scheffers, B. R., L. De Meester, T. C. L. Bridge, A. A. Hoffmann, J. M. Pandolfi, R. T. Corlett, S. H. M. Butchart, et al. 2016. The broad

- footprint of climate change from genes to biomes to people. *Science* 354: aaf7671.
- Servick, S., C. J. Visger, M. A. Gitzendanner, P. S. Soltis, and D. E. Soltis. 2015. Population genetic variation, geographic structure, and multiple origins of autopolyploidy in *Galax urceolata*. *American Journal of Botany* 102: 973–982.
- Shen, W., S. Le, Y. Li, and F. Hu. 2016. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* 11: e0163962.
- Simpson, M. G., C. M. Williams, K. E. Hasenstab-Lehman, M. E. Mabry, and L. Ripma. 2017a. Phylogeny of the popcorn flowers: Use of genome skimming to evaluate monophyly and interrelationships in subtribe Amsinckiinae (Boraginaceae). *Taxon* 66: 1406–1420.
- Simpson, M. G., L. A. Johnson, T. Villaverde, and C. M. Williams. 2017b. American amphitropical disjuncts: Perspectives from vascular plant analyses and prospects for future research. *American Journal of Botany* 104: 1600–1650.
- Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56: 13–30.
- Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.
- Soltis, D. E., C. J. Visger, and P. S. Soltis. 2014. The polyploidy revolution then... and now: Stebbins revisited. *American Journal of Botany* 101: 1057–1078.
- Soltis, P. S., and D. E. Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences, USA* 97: 7051–7057.
- Soto Gomez, M., M. J. M. Brown, S. Pironon, P. Bureš, L. D. Verde Arregoitia, P. Veselý, T. L. Elliott, et al. 2024. Genome size is positively correlated with extinction risk in herbaceous angiosperms. *New Phytologist* 243: 2470–2485.
- Stanke, M., O. Schöffmann, B. Morgenstern, and S. Waack. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7: 62.
- Stebbins, Jr., G. L. 1940. The significance of polyploidy in plant evolution. *American Naturalist* 74: 54–66.
- Tang, C.-Y., S. Li, Y.-T. Wang, and X. Wang. 2020. Comparative genome/transcriptome analysis probes Boraginales' phylogenetic position, WGDs in Boraginales, and key enzyme genes in the alkannin/shikonin core pathway. *Molecular Ecology Resources* 20: 228–241.
- Tang, H., X. Wang, J. E. Bowers, R. Ming, M. Alam, and A. H. Paterson. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research* 18: 1944–1954.
- Tillich, M., P. Lehwark, T. Pellizzer, E. S. Ulbricht-Jones, A. Fischer, R. Bock, and S. Greiner. 2017. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45: W6–W11.
- Van de Peer, Y., E. Mizrachi, and K. Marchal. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18: 411–424.
- Venable, D. L. 2007. Bet hedging in a guild of desert annuals. *Ecology* 88: 1086–1090.
- Veno, B. 1979. A revision of the genus *Pectocarya* (Boraginaceae) including reduction to synonymy of the genus *Harpagonella* (Boraginaceae). Doctoral dissertation, University of California, Los Angeles, California, USA.
- Veselý, P., P. Bureš, P. Šmarda, and T. Pavlíček. 2012. Genome size and DNA base composition of geophytes: The mirror of phenology and ecology? *Annals of Botany* 109: 65–75.
- Veselý, P., P. Šmarda, P. Bureš, C. Stirton, A. M. Muasya, L. Mucina, L. Horová, et al. 2020. Environmental pressures on stomatal size may drive plant genome size evolution: Evidence from a natural experiment with Cape geophytes. *Annals of Botany* 126: 323–330.
- Visger, C. J., C. C. Germain-Aubrey, M. Patel, E. B. Sessa, P. S. Soltis, and D. E. Soltis. 2016. Niche divergence between diploid and autotetraploid *Tolmiea*. *American Journal of Botany* 103: 1396–1406.
- Vitte, C., and O. Panaud. 2005. LTR retrotransposons and flowering plant genome size: Emergence of the increase/decrease model. *Cytogenetic and Genome Research* 110: 91–107.
- Vurture, G. W., F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, and M. C. Schatz. 2017. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* 33: 2202–2204.
- Ward, D. E. 1983. Chromosome counts from New Mexico and southern Colorado. *Phytologia* 54: 302–309.
- Wen, J., and S. M. Ickert-Bond. 2009. Evolution of the Madrean-Tethyan disjunctions and the North and South American amphitropical disjunctions in plants. *Journal of Systematics and Evolution* 47: 331–348.
- Whiting, J. R., T. R. Booker, C. Rougeux, B. M. Lind, P. Singh, M. Lu, K. Huang, et al. 2024. The genetic architecture of repeated local adaptation to climate in distantly related plants. *Nature Ecology & Evolution*: 8: 1933–1947.
- Wick, R. R., M. B. Schultz, J. Zobel, and K. E. Holt. 2015. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* 31: 3350–3352.
- Zhang, C., T. Zhang, F. Luebert, Y. Xiang, C.-H. Huang, Y. Hu, M. Rees, et al. 2020. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Molecular Biology and Evolution* 37: 3188–3210.
- Zhong, X. 2020. Assembly, annotation and analysis of chloroplast genomes. Doctoral dissertation, The University of Western Australia, Crawley, Western Australia, Australia. <https://doi.org/10.26182/5f333d9ac2bee>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Appendix S1.** Supplemental tables for “Chromosome-scale reference genome of *Pectocarya recurvata*, the species with the smallest reported genome size in Boraginaceae.”

**Appendix S2.** Supplemental figures for “Chromosome-scale reference genome of *Pectocarya recurvata*, the species with the smallest reported genome size in Boraginaceae.”

**How to cite this article:** Northing, P. C., J. A. Pelosi, D. L. Venable, and K. M. Dlugosch. 2025. Chromosome-scale reference genome of *Pectocarya recurvata*, the species with the smallest reported genome size in Boraginaceae. *Applications in Plant Sciences* 13(3): e70008. <https://doi.org/10.1002/aps3.70008>