

Chromosome-Level Reference Genome of the Beach False Foxglove, *Agalinis fasciculata* (Orobanchaceae)

Pedro H. Pezzi ¹, Maribeth Latvis ^{1,2,*}

¹Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

²University of Arkansas Herbarium, University of Arkansas, Fayetteville, AR 72701, USA

*Corresponding author: E-mail: mlatvis@uark.edu.

Accepted: April 30, 2026

Abstract

Orobanchaceae is the largest family of parasitic plants, encompassing a full spectrum of parasitic strategies, ranging from autotrophic to holoparasitic. *Agalinis* is a genus of facultative hemiparasites comprising about 70 species distributed throughout the Americas, including several endemic and rare taxa. *Agalinis fasciculata*, the beach false foxglove, is a widely distributed species across southeastern North America. Here, we use PacBio HiFi, Omni-C, and RNA-seq data to generate the first high-quality reference genome for the genus. The nuclear genome is 2.29 Gb in size, with most sequences anchored to 14 pseudochromosomes and an N50 of 162 Mb. BUSCO analyses indicate high completeness (98.4%). Structural genome annotation identified 34,133 protein-coding genes and 39,266 transcripts, most of which have at least one functional annotation. We have also assembled the plastid and mitochondrial genomes into complete circular molecules. We further examined genetic diversity and demographic history in *A. fasciculata*, revealing low genome-wide heterozygosity and evidence of inbreeding. This reference genome is an important resource for understanding the evolutionary history of the genus and the evolutionary patterns of parasitism within Orobanchaceae.

Key words: chromosome-level genome, Omni-C, PacBio HiFi, plastid genome, plant genomics.

Significance

This high-quality genome is the first chromosome-level assembly for *Agalinis*, a hemiparasitic genus in the plant family Orobanchaceae. It improves taxon sampling within Orobanchaceae, representing an important resource for investigating patterns of genome evolution in parasitic lineages. Furthermore, *Agalinis* has served as a focal genus for studies of the anatomy of haustorial development, and genome annotation incorporated RNA from multiple tissues, enabling the identification of genes expressed in different tissues, including roots and haustoria. This genome also serves as a reference for evolutionary studies of other *Agalinis* species, many of which are endemic and of conservation concern in North and South America. Overall, the beach false foxglove genome will support studies of the evolutionary history of *Agalinis* and genome evolution across Orobanchaceae.

Introduction

Orobanchaceae is the largest family of parasitic plants, comprising over 2,000 species across 90 genera, with a wide range of reliance on host plants and photosynthetic capability, including autotrophy, hemiparasitism (photosynthetic), and holoparasitism (nonphotosynthetic).

With relaxed selective pressure on photosynthesis, increased reliance on host plants is accompanied by a dramatic reduction in vegetative structures, as well as parallel patterns of gene family loss and expansion (Xu et al. 2022). Currently, fewer than 20 species have publicly available genomes, representing 11 genera.

© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Many Orobanchaceae are economically important as agricultural pests due to their parasitic lifestyle, and about half of the available genomes come from pest genera. However, genomic resources remain limited for species that do not parasitize crops. An improved taxon sampling is crucial for understanding genome evolution across the evolutionary breadth of the family.

Within Orobanchaceae, *Agalinis* (the false foxgloves) is a hemiparasitic genus in the tribe Pedicularideae distributed across the Americas, likely originating in North America and later dispersing into South America, accompanied by increased diversification and pollinator and life-history strategy shifts (Pezzi et al. 2026). Of the ~70 described species, which now includes the genus *Esterhazyia* (Latvis et al. 2024; Souza et al. 2025), many have highly restricted distribution, and in North America, several species are considered imperiled in at least one state. Moreover, *Agalinis* served as a focal genus in early anatomical studies on the induction of haustoria, the specialized roots of plant parasites that penetrate host plants, with implications for host recognition (Chang and Lynn 1986). A reference genome for *Agalinis* represents a valuable resource for conservation genomic studies of imperiled species in North and South America, as well as for understanding genome evolution and selective pressures in hemiparasitic taxa.

We present a chromosome-level genome of the beach false foxglove, *A. fasciculata* (Elliott) Raf. We assembled the nuclear genome and annotated it using RNA-seq data from flowers, roots, and leaves. We also estimated the genetic diversity and reconstructed the historical demographic history of the species. The addition of an Orobanchaceae reference genome will advance phylogenomic comparative analyses of parasitism and provide a valuable resource for studies of *Agalinis* evolution and conservation.

Results and Discussion

Sequencing and Nuclear Genome Assembly and Scaffolding

The newly assembled nuclear genome of *A. fasciculata* has a total length of 2.29 Gb, with an N50 of 162 Mb and an N90 of 131 Mb (Table 1; Fig. S1a). The genome was estimated to be ~2.3 Gb (Fig. S2), diploid (Fig. S3), and showed low heterozygosity (0.15%) and high repetitive content (80.4%). After decontamination (Fig. S1b), the assembly comprised 153 scaffolds. Of these, the 14 longest scaffolds corresponded to the 14 chromosomes, which agrees with previous estimates for the base chromosome number of this species (Canne 1984). We identified at least one telomeric repeat on each pseudochromosome (Fig. S4). Together, these 14

pseudochromosomes encompass 98.89% of the total genome length. The Omni-C contact map shows clear chromosome-scale scaffolds (Fig. S5).

Organelle Genome Assembly

We assembled the plastid genome as a circular molecule of 144,732 bp (Fig. S1c). Genome size was consistent with that reported for autotrophic and hemiparasitic Orobanchaceae (Zhang et al. 2020). In holoparasitic Orobanchaceae, plastid genomes undergo extensive restructuring associated with the loss of photosynthesis, likely due to relaxed selection (Wicke et al. 2013), leading to substantial variation in genome size and gene loss. However, studies have indicated that the loss of tRNAs and *ndh* genes precedes the loss of photosynthesis in some hemiparasitic groups (Zhang et al. 2020). Consistent with this pattern, the plastid genome of *A. fasciculata* contains 109 genes, including 76 protein-coding genes, 29 tRNAs, and 4 rRNAs (Table S1), lacking some *ndh* genes present in other hemiparasitic Orobanchaceae species (Zhang et al. 2020). However, a more detailed screening is necessary to understand the plastid genome evolution and selective pressures in the *Agalinis* plastome.

The mitochondrial genome was assembled into a circular molecule of 729,428 bp (Fig. S1d), containing 61 genes, including 35 protein-coding genes, 23 tRNAs, and 3 rRNAs (Table S1). Fan et al. (2016) reported limited

Table 1 Summary statistics of the *Agalinis fasciculata* genome

Metric	Value
Genome size (Mb)	2,286.85
Number of scaffolds	153
Scaffold N50 (Mb)	162.16
Scaffold L50	7
Longest scaffold (Mb)	223.48
GC content (%)	39.57
N's per 100 kbp	1.35
Gaps	808
Protein-coding genes	32,602
Repeats (%)	87.43
BUSCO genome completeness (eudicotyledons_odb12)	
Total BUSCO groups (n)	2,805
Complete BUSCOs (n, %)	2,761 (98.4)
Single-copy (n, %)	2,128 (75.9)
Duplicated (n, %)	633 (22.6)
Fragmented (n, %)	15 (0.5)
Missing (n, %)	29 (1.0)
BUSCO proteome completeness (eudicotyledons_odb12)	
Total BUSCO groups (n)	2,805
Complete BUSCOs (n, %)	2,723 (97.1)
Single-copy (n, %)	2,161 (77.0)
Duplicated (n, %)	562 (20.0)
Fragmented (n, %)	19 (0.7)
Missing (n, %)	63 (2.2)

gene loss in the mitogenomes of parasitic plants. However, Cai et al. (2026) showed that non-core genes are often pseudogenized. For example, *sdh3* has multiple copies in *A. fasciculata*, all of which lack start or stop codons. Studies have also demonstrated that plant mitogenomes experience dynamic selective pressures and exhibit an inverse relationship with plastid genome size in Orobanchaceae (Feng and Wicke 2025). The increasing availability of Orobanchaceae organelle genomes will enable a better understanding of how parasitism shapes organelle genome evolution, horizontal gene transfer, and interactions among the nuclear, plastid, and mitochondrial genomes, a promising area for future studies of parasitic lineages.

Nuclear Genome Annotation and Tissue Specificity

BUSCO analysis (eudicotyledons_odb12) (Table 1; Fig. S1a) indicated high genome completeness (98.4%), with 75.9% of single-copy and 22.6% of duplicated loci. The high proportion of duplicated loci is consistent with its Pedicularideae relatives *Castilleja foliosa* Hook. & Arn. (C:98.2%[S:74.3%; D23.9%]), but not with *Phtheirospermum japonicum* (Thunb.) Kanitz (C:97.5% [S:91.9%; D:5.6%]). The genome of *A. fasciculata* is larger than that of *C. foliosa* (751.2 Mb), indicating that transposable element expansion contributes to its large genome size and high repetitive content (87.43%), particularly retroelements (Table S2), which accounted for 73.31% of the repeats. Repetitive regions constitute most of the genomes in the 3 Pedicularideae species with available genomes, accounting for 72.22% in *C. foliosa* and 67.38% in *P. japonicum*. Both the genome size and repetitive content are consistent with the GenomeScope estimates (Fig. S2). Structural annotation predicted 34,133 protein-coding genes and 39,266 transcripts, which were reduced to 32,602 genes and 37,156 transcripts after functional annotation. Of these, 36,343 transcripts were assigned at least one functional annotation. BUSCO analysis of the longest isoform per locus in the proteome also indicated high completeness (97.1%), with 77.0% single-copy and 20.0% duplicated loci (Table 1). Combined RNA data from flowers and flower buds exhibited the highest number of enriched genes, with most genes showing strong tissue specificity (Fig. S6a, b). When analyzed separately, flower buds and roots + haustoria had the highest number of enriched genes (Fig. S6c), although overall Tau was lower (Fig. S6d).

Genetic Diversity, Runs of Homozygosity and Demographic History

Mean genome-wide heterozygosity in *A. fasciculata* was low (0.72 het./kb) and unevenly distributed across

chromosomes, with higher levels observed on chromosomes 8 and 14 (Fig. 1a). Scans for runs of homozygosity (ROH), i.e. continuous regions of reduced heterozygosity, identified 467 long segments (≥ 1 Mb) of up to 5.96 Mb in length (Fig. 1b, c). Using a mutation rate of 2.12×10^{-9} per generation, demographic inference indicated that *A. fasciculata* reached its peak effective population size (N_e) approximately 200,000 years ago, followed by a rapid decline until 20,000 years ago, when the population began a slow recovery (Fig. 1d). This recovery coincides with the Last Glacial Maximum, a period marked by the expansion of open vegetation habitats across southeastern North America (Hall and Valastro 1995), potentially facilitating population expansion of open vegetation species such as *Agalinis*. Although *A. fasciculata* is not currently considered a species of conservation concern, the low heterozygosity and N_e suggest high levels of inbreeding. While the reproductive biology of *A. fasciculata* remains unknown, *Agalinis* species are capable of self-fertilization, which may contribute to the low genetic diversity and high inbreeding levels.

Materials and Methods

DNA Extraction, PacBio HiFi and Omni-C Sequencing, and Quality Control

We collected several individuals of *A. fasciculata* at Cato Springs Prairie, University of Arkansas, Fayetteville (36°02'24.4"N, 94°10'59.9"W) in September of both 2024 and 2025, and snap-froze flowers, leaves, and roots + haustoria tissues in liquid nitrogen. A voucher specimen was deposited in the UARK Herbarium (accession no. UARK 095889). High-molecular-weight genomic DNA (gDNA) was extracted using a CTAB protocol and purified with a salt:chloroform wash. Library preparation was performed at the University of Oregon Genomics and Cell Characterization Core Facility (Eugene, OR), and sequencing was carried out on a PacBio Revio (Pacific Biosciences, Menlo Park, CA) using 2 Single Molecule, Real-Time (SMRT) cells. A different individual from the same population was used for Omni-C sequencing. Omni-C library preparation and sequencing were conducted at Dovetail Genomics (Scotts Valley, CA). Adapter sequences from HiFi reads were filtered using HiFiAdapterFilt (Sim et al. 2022). We used meryl (Rhie et al. 2020) to count 21-mers and used the output to estimate genome size, heterozygosity, and repeat content with GenomeScope 2.0 (Ranallo-Benavidez et al. 2020).

Genome Assembly, Scaffolding, and Quality Assessment

We assembled the genome using hifiasm v0.19.5 (Cheng et al. 2021). Given the high level of homozygosity,

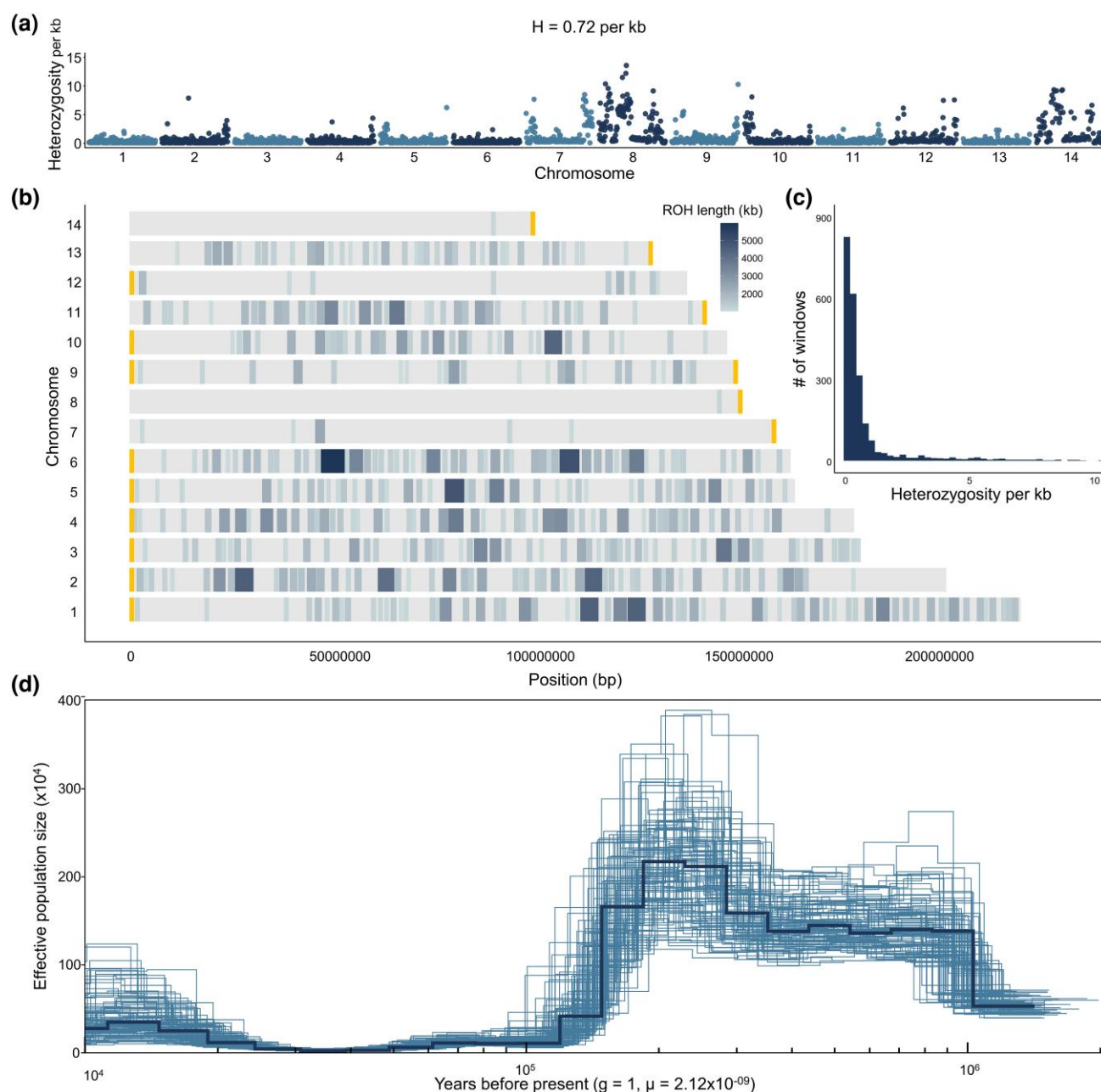


Fig. 1. Genetic diversity and demographic history of *Agalinis fasciculata*. a) Genome-wide heterozygosity was estimated in nonoverlapping 1-Mb windows across the 14 chromosomes. H indicates mean genome-wide heterozygosity. b) Genomic location and length of runs of homozygosity (ROH) across the 14 chromosomes, with darker colors indicating longer ROH. Yellow bars at chromosome ends indicate telomeric repeats. c) Frequency distribution and size of runs of homozygosity (ROH). d) Historical changes in effective population size (N_e) inferred for *A. fasciculata* using a mutation rate of 2.12×10^{-9} per generation. Light blue lines represent 100 bootstrap replicates, whereas the dark blue line represents the estimate based on the whole genome.

downstream analyses were conducted on the primary assembly. We used `purge_haplotigs` v1.1.3 (Roach et al. 2018) to remove allelic variants from the primary assembly, with read depth cutoffs of $-l$ 10, $-m$ 60, and $-h$ 200. Omni-C reads were mapped and filtered following the Dovetail Genomics reference guide.

We used YaHS v1.2.2 (Zhou et al. 2023) to scaffold our primary assembly. Plastid and mitochondrial scaffolds were removed by mapping the organelle genomes against the assembly using MUMmer v4.0.1 (Marçais et al. 2018), with a minimum identity of 95% and a minimum alignment length of 5,000 bp. To avoid removing

nuclear mitochondrial DNA (NUMTs) and nuclear plastid DNA (NUPTs), we only filtered out scaffolds shorter than the corresponding organelle genome and with $\geq 75\%$ coverage by that genome. We used Kraken v2.1.3 (Lu et al. 2022) to remove contaminant scaffolds, applying a confidence threshold of 0.3 with the PlusPFP database. We visualized contact maps using PretextView (<https://github.com/wtsi-hpag/PretextView>). We used tidk v0.2.65 (Brown et al. 2025) to identify telomeric repeat regions across scaffolds. As tidk detected telomeric repeats in the middle of the longest scaffold and karyotyping confirmed 14 chromosomes (Canne 1984), we manually split this scaffold to match the chromosome number. Assembly gaps were filled using YAGClosier (<https://github.com/merlyescalona/yagclosier>) with a flank size of 40 (-f 40) and a minimum of 2 reads (-mins 2) spanning each gap. We assessed genome quality and completeness using BUSCO v6.0.0 (Tegenfeldt et al. 2025) using the eudicotyledons_odb12 database. We assessed genome quality using BlobTools2 (Challis et al. 2020) and removed contaminant scaffolds that were kept by kraken2. We estimated ploidy level by analyzing heterozygous k-mer pairs with smudgeplot v0.4.0dev (Ranallo-Benavidez et al. 2020) and assessed genome quality with QUAST v5.3.0 (Gurevich et al. 2013). We used Oatk v1.0 (Zhou et al. 2025) to assemble the organelle genomes. The plastid genome was annotated using GeSeq (Tillich et al. 2017), and the mitogenome was annotated using PMGA (Li et al. 2025). Organelle genomes were visualized with OGDRAW v1.3.1 (Greiner et al. 2019).

RNA Sequencing, Genome Annotation, and Tissue Specificity

We generated a de novo, species-specific repetitive sequence and transposable element (TE) library using RepeatModeler2 v2.0.7 (Flynn et al. 2020). Sequences labeled as “unknown” were classified following Nachtigall et al. (2025). Briefly, we used DeepTE v1.0 (Yan et al. 2020) with the plant-specific model to classify those repetitive sequences and removed false positives using TERL v1.0 (Da Cruz et al. 2021). We used this dataset to soft-mask repetitive regions in the genome with RepeatMasker v4.2.1 (Smit et al. 2013; <https://www.repeatmasker.org>). We soft-masked NUMTs and NUPTs by using the assembled organelle genomes as queries in a BLASTn search against the nuclear genome.

Flower, flower bud, leaves, and root + haustoria tissues were sent to Novogene (Sacramento, CA) for RNA extraction, library preparation, and sequencing on an Illumina NovaSeq X Plus. We used fastp v1.0.1 (Chen et al. 2018) to process RNA-seq reads, followed by a correction step using Rcorrector v1.0.7

(Song and Florea 2015). We mapped the RNA-seq reads to the assembly using HISAT2 v2.2.1 (Kim et al. 2019) and used its output for structural genome annotation with BRAKER3 v3.0.8 (Gabriel et al. 2024). We also used the Viridiplantae protein dataset from OrthoDB v12 (Tegenfeldt et al. 2025) and the proteomes of nine Orobanchaceae species (Table S3) as protein evidence.

We used eggNOG-mapper v2.1.12 (Cantalapiedra et al. 2021; <http://eggno-mapper.embl.de/>) and InterProScan v5.59-91.0 + galaxy3 (Jones et al. 2014) via Galaxy Europe (Afgan et al. 2018; <https://usegalaxy.eu/>) for functional annotation. The outputs were then used as input for Funannotate v1.8.17 (Palmer and Stajich 2020). Functional annotation was supplemented with BLASTp searches for genes labeled as hypothetical proteins. We used three protein datasets (retrieved on 15 Oct 2025) hierarchically: (i) a reviewed and curated SwissProt protein dataset of eudicotyledons (28,390 sequences); (ii) an unreviewed TrEMBL dataset of lamiid proteins (2,376,208 sequences); and (iii) a curated SwissProt dataset across all taxa (573,661 sequences). We used STAR v2.7.11b (Dobin et al. 2013) to align the tissue RNA-seq reads to the novel genome. We compared expression across the tissues to understand transcript distribution and tissue specificity using the Tau (Yanai et al. 2005)

Genetic Diversity and Population Size History

As in Gonçalves et al. (2025), we evaluated genome-wide heterozygosity and ROH as proxies for genetic diversity in *A. fasciculata*. Filtered HiFi reads were mapped to the assembled genome using minimap2 (Li 2018). The resulting BAM file was sorted with SAMtools (Danecek et al. 2021), and a VCF file containing variants and invariants was created using BCFtools v1.22 (Danecek et al. 2021) with minimum base and mapping quality thresholds of 30 (-Q 30 and -q 30). We used pixy v2.0.0.beta14 (Korunes and Samuk 2021) to estimate nucleotide diversity across nonoverlapping 1 Mb sliding windows and PLINK v1.9.0-b.8 (Chang et al. 2015) to estimate ROH.

We used the Pairwise Sequentially Markovian Coalescent model (PSMC; <https://github.com/lh3/psmc>) to infer historical changes in N_e of *A. fasciculata*. PSMC was run with 100 bootstrap replicates using the following parameters: -N25 -t15 -r5 -p “1 + 1 + 1 + 1 + 25 × 2 + 4 + 6”. A mutation rate was estimated by calculating the number of 1:1 genomic matches between *A. fasciculata* and *C. foliosa* (Bürger et al. 2024). The masked genomes were aligned with nucmer from MUMmer and the output was filtered using delta-filter via the Galaxy Europe platform. The total number of SNPs was divided by the number of 1:1 matches to estimate overall genomic divergence,

which was converted into a mutation rate per generation using the formula described in Luo et al. (2023), assuming a mean generational time of 1 year and a divergence time between *Agalinis* and *Castilleja* of 32.93 mya (Fonseca 2021). These analyses were restricted to the 14 chromosomes.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

Computational analyses were supported by the Arkansas High Performance Computing Center, which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission. We also thank Jennifer Ogle, M.S., for assistance with sampling and species identification, Sarah Jacobs, Ph.D., for advice on DNA extraction, and Leonardo Tresoldi Gonçalves, Ph.D., for support with sampling and bioinformatics analyses.

Author Contributions

P.H.P.: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing; M.L.: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing—review & editing.

Funding

This research was supported by the University of Arkansas start-up research funds to M. L. and the Open Access Publishing Fund administered by the University of Arkansas Libraries.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Data Availability

Scripts with all the parameters and files used for the analyses are available on GitHub https://github.com/pedrohpuzzi/Genome_Agalinis_fasciculata and <https://doi.org/10.5281/zenodo.18393392>. Raw reads have been deposited in GenBank under the BioProject accession number PRJNA1345713. Primary and alternate genome assemblies are available under BioProject accession numbers PRJNA1454504 and PRJNA1454503, respectively. Chloroplast and mitochondrial genomes

are available under GenBank accession numbers PZ319439 and PZ330306 respectively.

Literature Cited

- Afgan E, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 2018;46:W537–W544. <https://doi.org/10.1093/nar/gky379>.
- Palmer JM, Stajich J. 2020. Funannotate v1.8.1: Eukaryotic genome annotation. [accessed 2025 Oct 22]. <https://zenodo.org/record/4054262>. <https://doi.org/10.5281/ZENODO.4054262>.
- Brown MR, Gonzalez De La Rosa PM, Blaxter M. Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics.* 2025;41:btaf049. <https://doi.org/10.1093/bioinformatics/btaf049>.
- Bürger M, Peterson D, Chory J. Strigolactones initiate the formation of haustorium-like structures in *Castilleja*. *iScience.* 2024;27:111491. <https://doi.org/10.1016/j.isci.2024.111491>.
- Cai L, Havird JC, Jansen RK. Recombination and retroprocessing in broomrapes reveal RNA-mediated gene transfer mechanism and a generalizable model for mitochondrial evolution in heterotrophic plants. *Genome Biol Evol.* 2026;18:evag025. <https://doi.org/10.1093/gbe/evag025>.
- Canne JM. Chromosome numbers and the taxonomy of North American *Agalinis* (Scrophulariaceae). *Can J Bot.* 1984;62:454–456. <https://doi.org/10.1139/b84-068>.
- Cantalapiedra CP, et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38:5825–5829. <https://doi.org/10.1093/molbev/msab293>.
- Challis R, et al. BlobToolKit—interactive quality assessment of genome assemblies. *G3 (Bethesda).* 2020;10:1361–1374. <https://doi.org/10.1534/g3.119.400908>.
- Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Chang M, Lynn DG. The haustorium and the chemistry of host recognition in parasitic angiosperms. *J Chem Ecol.* 1986;12:561–579. <https://doi.org/10.1007/BF01020572>.
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Cheng H, et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Da Cruz MHP, et al. TERL: classification of transposable elements by convolutional neural networks. *Brief Bioinform.* 2021;22:bbaa185. <https://doi.org/10.1093/bib/bbaa185>.
- Danecek P, et al. Twelve years of SAMtools and BCFtools. *GigaScience.* 2021;10:giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Fan W, et al. Limited mitogenomic degradation in response to a parasitic lifestyle in Orobanchaceae. *Sci Rep.* 2016;6:36285. <https://doi.org/10.1038/srep36285>.
- Feng Y, Wicke S. Systemic organellar genome reconfiguration along the parasitic continuum in the broomrape family (Orobanchaceae). *Plant Cell Physiol.* 2025;67:577–591. <https://doi.org/10.1093/pcp/pcaf131>.
- Flynn JM, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 2020;117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>.

- Fonseca LHM. Combining molecular and geographical data to infer the phylogeny of Lamiales and its dispersal patterns in and out of the tropics. *Mol Phylogenet Evol.* 2021;164:107287. <https://doi.org/10.1016/j.ympev.2021.107287>.
- Gabriel L, et al. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* 2024;34:769–777. <https://doi.org/10.1101/gr.278090.123>.
- Gonçalves LT, et al. The nuclear and mitochondrial genomes of the bellicose bumblebee (*Bombus bellicosus*, Hymenoptera: Apidae), a threatened pollinator in a changing South American landscape. *Genome.* 2025;68:1–10. <https://doi.org/10.1139/gen-2024-0166>.
- Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 2019;47:W59–W64. <https://doi.org/10.1093/nar/gkz238>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Hall SA, Valastro S. Grassland vegetation in the southern great plains during the last glacial Maximum. *Quat Res.* 1995;44:237–245. <https://doi.org/10.1006/qres.1995.1068>.
- Jones P, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Kim D, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
- Korunes KL, Samuk K. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour.* 2021;21:1359–1368. <https://doi.org/10.1111/1755-0998.13326>.
- Latvis M, et al. An evolutionary framework for *Agalinis* (Orobanchaceae; The False Foxgloves) reveals a rapid South American radiation that includes *Esterhazyia*. *Syst Bot.* 2024;49:455–481. <https://doi.org/10.1600/036364424X17095905880222>.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li J, et al. PMGA: a plant mitochondrial genome annotator. *Plant Commun.* 2025;6:101191. <https://doi.org/10.1016/j.xplc.2024.101191>.
- Lu J, et al. Metagenome analysis using the Kraken software suite. *Nat Protoc.* 2022;17:2815–2839. <https://doi.org/10.1038/s41596-022-00738-y>.
- Luo H, et al. A high-quality genome assembly highlights the evolutionary history of the great bustard (*Otis tarda*, Otidiformes). *Communications Biology.* 2023;6:197. <https://doi.org/10.1038/s42003-023-05137-x>.
- Marçais G, et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* 2018;14:e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>.
- Nachtigall PG, et al. A segregating structural variant defines novel venom phenotypes in the eastern diamondback rattlesnake. *Mol Biol Evol.* 2025;42:msaf058. <https://doi.org/10.1093/molbev/msaf058>.
- Pezzi PH, Soltis PS, Soltis DE, Latvis M. ‘Dispersification’ of *Agalinis* (Orobanchaceae) into South America is associated with hummingbird pollination and perennial life-history shifts. *J Biogeogr.* 2026;53:e70143. <https://doi.org/10.1111/jbi.70143>.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 2020;11:1432. <https://doi.org/10.1038/s41467-020-14998-3>.
- Rhie A, Walenz BP, Koren S, Phillippy AM. Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020;21:245. <https://doi.org/10.1186/s13059-020-02134-9>.
- Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics.* 2018;19:460. <https://doi.org/10.1186/s12859-018-2485-7>.
- Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics.* 2022;23:157. <https://doi.org/10.1186/s12864-022-08375-1>.
- Smit AFA, Hubley R, Green P. RepeatMasker. 2013. <http://www.repeatmasker.org>
- Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience.* 2015;4:48. <https://doi.org/10.1186/s13742-015-0089-y>.
- Souza VC, De Paula-Souza J, Latvis M. (3108) proposal to conserve *Agalinis*, nom. cons. against the additional name *Esterhazyia* (Orobanchaceae). *TAXON.* 2025;74:994–995. <https://doi.org/10.1002/tax.70025>.
- Tegenfeldt F, et al. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res.* 2025;53:D516–D522. <https://doi.org/10.1093/nar/gkac987>.
- Tillich M, et al. Geseq—versatile and accurate annotation of organellar genomes. *Nucleic Acids Res.* 2017;45:W6–W11. <https://doi.org/10.1093/nar/gkx391>.
- Wicke S, et al. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell.* 2013;25:3711–3725. <https://doi.org/10.1105/tpc.113.113373>.
- Xu Y, et al. Comparative genomics of orobanchaceous species with different parasitic lifestyles reveals the origin and stepwise evolution of plant parasitism. *Mol Plant.* 2022;15:1384–1399. <https://doi.org/10.1016/j.molp.2022.07.007>.
- Yan H, Bombarely A, Li S. DeepTE: a computational method for *de novo* classification of transposons with convolutional neural network. *Bioinformatics.* 2020;36:4269–4275. <https://doi.org/10.1093/bioinformatics/btaa519>.
- Yanai I, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics.* 2005;21:650–659. <https://doi.org/10.1093/bioinformatics/bti042>.
- Zhang R, et al. Transit from autotrophism to heterotrophism: sequence variation and evolution of chloroplast genomes in Orobanchaceae species. *Front Genet.* 2020;11:542017. <https://doi.org/10.3389/fgene.2020.542017>.
- Zhou C, et al. Oatk: a *de novo* assembly tool for complex plant organelle genomes. *Genome Biol.* 2025;26:235. <https://doi.org/10.1186/s13059-025-03676-6>.
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 2023;39:btac808. <https://doi.org/10.1093/bioinformatics/btac808>.

Associate editor: Chuan Ku