

UC Riverside

UC Riverside Previously Published Works

Title

Whole-genome sequence and annotation of *Penstemon davidsonii*.

Permalink

<https://escholarship.org/uc/item/5zb3k0q3>

Journal

G3: Genes, Genomes, Genetics, 14(3)

Authors

Alabady, Magdy

Zhang, Mengrui

Rausher, Mark

et al.

Publication Date

2024-03-06

DOI

10.1093/g3journal/jkad296

Peer reviewed

Whole-genome sequence and annotation of *Penstemon davidsonii*

Kate L. Ostevik,^{1,2,*} Magdy Alabady ,³ Mengrui Zhang,⁴ Mark D. Rausher²

¹Department of Evolution, Ecology, and Organismal Biology, University of California Riverside, Riverside, CA 92521, USA

²Department of Biology, Duke University, Durham, NC 27708, USA

³Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

⁴Department of Statistics, University of Georgia, Athens, GA 30602, USA

*Corresponding author: Department of Evolution, Ecology, and Organismal Biology, University of California Riverside, Duke University, EEOB, 900 University Ave., Riverside, CA 92521, USA. Email: kostevik@ucr.edu.

Penstemon is the most speciose flowering plant genus endemic to North America. *Penstemon* species' diverse morphology and adaptation to various environments have made them a valuable model system for studying evolution. Here, we report the first full reference genome assembly and annotation for *Penstemon davidsonii*. Using PacBio long-read sequencing and Hi-C scaffolding technology, we constructed a de novo reference genome of 437,568,744 bases, with a contig N50 of 40 Mb and L50 of 5. The annotation includes 18,199 gene models, and both the genome and transcriptome assembly contain over 95% complete eudicot BUSCOs. This genome assembly will serve as a valuable reference for studying the evolutionary history and genetic diversity of the *Penstemon* genus.

Keywords: *Penstemon davidsonii*; Davidson's beardtongue; PacBio Hifi; Hi-C; genome assembly; genome annotation

Introduction

Penstemon (Plantaginaceae) is a genus made up of many flowering plant species (~280) commonly known as the beardtongues. It is the largest plant genus endemic to North America. *Penstemon* species have diverse vegetative and floral morphology and are adapted to a wide range of environments (Nold 1999; Wolfe et al. 2006; Wilson et al. 2007; Thomson and Wilson 2008). This diversity has led to *Penstemon* emerging as a model system for the study of evolution, especially with regard to pollination syndromes (e.g. Wilson et al. 2004; Wessinger et al. 2014; Katzer et al. 2019), speciation (e.g. Straw 1955; Clark 1971; Wolfe et al. 1998; Stone et al. 2023), and repeated evolution (e.g. Wilson et al. 2007; Wessinger and Rausher 2014).

In addition to several chloroplast genomes (Ricks et al. 2017; Stettler et al. 2021), 2 nuclear genomes have recently become available for *Penstemon*, *P. barbatus* (Wessinger et al. 2023) and *P. kunthii* (Schlenk 2023). However, additional reference genomes for species in the genus would significantly enhance our understanding of *Penstemon*'s biology and open up new avenues of research, such as studies of chromosomal evolution and comparative genomics.

Penstemon davidsonii E. Greene is a perennial species that occurs in the alpine zone of the Sierra Nevada and Cascade Mountain ranges. It is a diploid with 8 chromosomes and a predicted genome size of 483 Mb (Broderick et al. 2011). *P. davidsonii* occurs within subgenus *Dasanthera* (Datwyler and Wolfe 2004; Wolfe et al. 2021), which is the outgroup to the rest of *Penstemon*. Although partial genome sequences have previously been generated for this species (Dockter et al. 2013), a full reference genome for

subgenus *Dasanthera* is needed. This will allow researchers to infer the ancestral genome for this clade and, therefore, better understand evolutionary changes that occurred within *Penstemon*.

Here, we report the first full genome assembly for *P. davidsonii*. We use a combination of PacBio and Hi-C sequencing to assemble a de novo reference genome for this species.

Materials and methods

Study organism and collection

We collected cuttings from multiple *P. davidsonii* plants along Piute Pass in the Sierra Nevada Mountain range in June 2016. These cuttings were allowed to root in water, transferred to a 1:1 mix of potting soil and pumice, and kept under controlled conditions in the Duke University greenhouse. Because *P. davidsonii* hybridizes with *P. newberryi* along Piute Pass (Clausen et al. 1940; Chabot and Billings 1972; Datwyler and Wolfe 2004; Kimball 2008), we used population genetic data from Garcia et al. (2023) and phenotypic characteristics to identify an individual with little *P. newberryi* ancestry. Of the plants collected, DNT005 appeared to be exclusively *P. davidsonii*. This individual was collected at 11,598 ft (latitude: 37.2422, longitude: -118.68265) and used for subsequent sequencing.

DNA isolation and PacBio sequencing

We finely ground flash-frozen leaf tissue from DNT005 using a mortar and pestle and extracted DNA using a Cetyltrimethyl Ammonium Bromide (CTAB) protocol (Clarke 2009) with the following modifications. After grinding tissue, we performed 3

washes with a 0.35 M sorbitol buffer to remove secondary metabolites before cell lysis (Inglis et al. 2018). We added 2% polyvinylpyrrolidone, 2% dithiothreitol, and 2% beta-mercaptoethanol to the CTAB buffer because each of these reagents has been shown to improve DNA extraction in difficult plant tissues (e.g. Porebski et al. 1997, Horne et al. 2004, Aboul-Maaty and Oraby 2019). Finally, we incubated samples with RNase A for 30 min at 37°C before precipitating DNA to remove RNA from the samples.

We purified the DNA using AMPure beads (Beckmann Coulter, USA) and assessed quality using a Large Fragment Analysis Kit (Agilent, USA). The library was prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, USA) and Sequel II Binding Kit 2.0 with a peak insert size of ~18 kbp (mean = 12,986 bp). The sequencing was performed on the Sequel II system with a movie time of 30 h. This yielded 3,829,617 reads with a mean length of 30,922. The SMRTbell library and sequencing were done at the Georgia Genomics and Bioinformatics Core.

Hi-C library preparation and sequencing

We used intact leaf tissue from DNT005C and the Proximo TM Hi-C kit (Phase Genomics, USA) to generate libraries for Hi-C sequencing. In brief, nuclear DNA was crosslinked before endonuclease treatment with DpnII, and the fragmented dsDNA was biotinylated to create junctions as per Phase genomics protocols (van Berkum et al. 2010). These biotinylated fragments were purified and subjected to paired-end Illumina 2 × 150 sequencing yielding 141.6 million read pairs.

Genome assembly

The genome was assembled using HiCanu v2.1 (Nurk et al. 2020) consensus sequences and duplicate purged using “purge_dups” (Guan et al. 2020) to reduce the assembly to haplotype as much as possible. Then, we used the 3D-DNA pipeline (Dudchenko et al. 2017) with the default values for scaffolding. Specifically, we calculated the enzyme restriction sites for DpnII with respect to the draft assembly from HiCanu using the script generate_site_positions.py provided by the Juicer (Durand et al. 2016) toolkit. This restriction site map was used by the Juicer toolkit when aligning the Hi-C sequences against the draft assembly using bwa (Li 2013) to generate a duplicate-free list of paired alignments that were used by the 3D-DNA pipeline during scaffolding.

NextPolish (Hu et al. 2020) is a high-accuracy genome polishing tool that can incorporate both short reads and Pacbio HiFi reads. We prepared our reads for use by NextPolish as follows. The Hi-C reads were trimmed using cutadapt (Martin 2011) to remove low-quality bases and Ns at the start of reads. The HiFi reads were generated from the Pacbio subreads.bam using the Pacbio CCS14 tool. Then, we instructed the NextPolish pipeline to use bwa for the short read mapping and minimap2 in the mode designed for Pacbio CCS genomic reads for the HiFi read mapping. The specific modifications we made to the NextPolish configuration file are as follows: (1) the lgs_option section was removed as the raw Pacbio subreads were not being used, (2) a hifi_option section was added and set to “- min_read_len 1k -max_depth 100,” and (3) the hifi_minimap2_options section was set to “-x asm20.” The completeness of the assembly was assessed using BUSCO v4.0.6. (Simão et al. 2015) based on evolutionarily informed expectations of gene content from the eudicot ODB10 database. Finally, we used the Foreign Contamination Screening tool from the National Center for Biotechnology Information (Sayers et al. 2023) to identify contaminating sequences in the final assembly. This revealed no sequences from unintended organisms and only 4 small

regions (20–30 bp) of adaptor contamination, which we subsequently replaced with Ns.

RNA-seq and transcriptome assembly

We extracted RNA from frozen roots, leaves, and whole seedlings of the DNT005C plant using a Spectrum Plant Total RNA Kit (Sigma-Aldrich, USA). Libraries were made using the KAPA Stranded mRNA-Seq Kit (Roche, Switzerland) with NEBNext Multiplex Oligos (96 Unique Dual Index Primer Pairs) for Illumina Barcodes (New England Biolabs, USA) and sequenced on an Illumina NovaSeq 6000 S2 150 bp PE flow cell at the Duke Center for Genomic and Computational Biology Sequencing and Genomic Technologies Core.

The RNA-seq short reads were trimmed and quality filtered using trim_galore (Krueger 2015). All cleaned paired and unpaired reads were de novo assembled using Trinity v2.6.6 (Grabherr et al. 2011) with default parameters (Haas et al. 2013) to generate the transcriptome. We used BUSCO in the transcriptome mode to estimate transcriptome completeness.

Genome annotation and masking

We identified and masked sequence repeats through the following steps. First, we identified tandem repeats within the genome using Tandem Repeat Finder v4.09 with specific parameter settings, including matching weight = 2, mismatching penalty = 5, indel penalty = 7, match probability = 0.8, indel probability = 0.1, minimum alignment score = 50, and maximum period size = 2,000 (Benson 1999). Second, we identified interspersed repeats and LTR retrotransposons using RepeatModeler v2.0.2 (Flynn et al. 2020). For the LTR transposons, we used the LTRStruct flag in RepeatModeler. Next, the identified repeat families were clustered using CD-HIT-EST v4.8.1 with 90% sequence identity and a seed size of 8 bp (Fu et al. 2012). Finally, the generated repeat library was applied to mask interspersed repeat elements in the assembly and develop repeat annotation files using RepeatMaskerv4.1.0 (Smit et al. 2015).

To prepare for gene annotation using the MAKER pipeline (Campbell et al. 2014), we took the following steps. The cleaned paired RNA-seq reads were mapped to the genome assembly using Tophat v2.1.2 followed by gene annotation using Cufflinks v2.2.1. We predicted tRNAs in the final assembly using tRNAscan-SE v2.0.7 with the default parameters (Chan and Lowe 2019). To identify miRNA precursor genes, we first retrieved 38,589 miRNAs precursors from the miRBase database (Griffiths-Jones et al. 2008) and clustered them separately to remove redundancies using CD-HIT-EST v4.8.1 (Fu et al. 2012), resulting in a total of 25,760 sequences. Then, these nonredundant sequences were used to identify the miRNA precursors in the *Penstemon* genome using homology-based search by BLASTN alignment tool with the default thresholds (Altschul et al. 1990).

We used the soft-masked genome generated using RepeatMasker v4.1.0 with the Repbase repeat library and all the previous annotations (i.e. TE, RNA-seq, miRNA, and tRNA) for gene prediction in the MAKER pipeline. Evidence from ab initio and homology-based methods were integrated to perform the final gene predictions. Specifically, the EST evidence from the RNA-seq assembly and ab initio gene predictions of the genome were used to construct the gene set using the MAKER pipeline. AUGUSTUS v3.2.3 (Stanke and Morgenstern 2005) was used for the ab initio gene prediction, and the BLAST alignment tool was used for homology-based gene prediction using the EST evidence in the MAKER pipeline. Exonerate v2.4.0 (Slater and Birney 2005)

Table 1. Summary statistics for the genome assembly of *Penstemon davidsonii*.

Total bases	437,568,744
Contig count	2,688
Largest contig	55,463,854
N50	40,949,504
L50	5
N90	50,886
L90	390
Ns	125
GC content	34.7%
BUSCO scores	C: 95.3% (S: 85.3%, D: 10%), F: 1.2%, M: 3.5%, n: 2326

BUSCO parameters are as follows: C, complete BUSCOs; S, complete and single-copy BUSCOs; D, complete and duplicated BUSCOs; F, fragmented BUSCOs; M, missing BUSCOs; and n, total BUSCO groups searched.

Table 2. Summary statistics for the *Penstemon davidsonii* annotation.

Genes	18,199
Protein-coding genes	17,299
Exons	109,386
CDSs	108,061
Expressed sequence matches	41,594
Protein matches	417,305
miRNA	4,331
tRNAs	901

was used to polish and curate the BLAST alignment results. And, we manually removed partial CDSs for 35 genes.

Results and discussion

Our PacBio sequencing produced 117.5 Gb (3.8 million reads with a mean length of 30,922), which represents 268x average coverage of our 438 Mb final genome (Table 1). This genome size is 91% of the 483 Mb estimate based on flow cytometry (Broderick et al. 2011).

Our initial assembly was 613 Mb, made up of 9,237 contigs. Duplicate purging reduced this to 438 Mb across 4,811 contigs. Then, scaffolding using 141.6 million Hi-C read pairs further reduced the number of contigs to 2,688 (Table 1). In the final assembly, 8 contigs >20 mb make up >60% of the reference and likely represent the 8 expected chromosomes. Completeness, as assessed by our BUSCO analysis, was ~95% (Table 1).

The assembled transcriptome consists of a total of 247,266,183 bases across 274,807 isoform “transcripts” with a 99% BUSCO score [C: 99.2% (S: 31.4%, D: 67.8%), F: 0%, M: 0.8%, n: 255; see Table 1 for abbreviations]. These transcripts clustered into 180,352 putative genes, of which 152,610 had a single isoform and 2,604 had 10 or more isoforms. Genome annotation using this transcriptome yielded slightly >18,000 gene models, of which slightly >17,000 were protein-coding genes (Table 2). We also identified 4,331 miRNAs and 901 tRNAs (Table 2).

In our final genome assembly, we identified and masked 257,613,589 bases of repeat content, which represents 58.87% of the total sequence. This included 135,653 transposable elements, made up of mostly Ty1/Copia (14.93% of genome sequence) and Gypsy/DIRS1 (8.96% of genome sequence) LTR retroelements (Table 3).

This assembled and annotated genome of *P. davidsonii* provides a needed resource for the *Penstemon* evolutionary biology community. To date, most evolutionary analysis of *Penstemon* species has

Table 3. Summary of repetitive element content found in the *Penstemon davidsonii* genome assembly.

Element type	Number of elements	Percent of sequence
SINES	0	0
LINEs	6,831	0.83
LTR retroelements	110,753	29.15
DNA transposons	18,069	1.74
Rolling circles	5,414	0.40
Small RNA	1,225	0.31
Satellites	2,380	0.12
Simple repeats	91,183	1.28
Low complexity	13,875	0.15
Unclassified	391,416	24.90

relied primarily on phenotypic analysis (e.g. Castellanos et al. 2006), examination of patterns exhibited by a small number of genes (e.g. Wessinger and Rausher 2015), or examination of patterns discernible from reduced representation sequencing (e.g. Stone and Wolfe 2021). As more *Penstemon* genomes come online, however, these analyses can be extended to a genomic scale, and investigators will be able to examine the evolutionary characteristics of entire *Penstemon* genomes.

Nevertheless, there is still room for improvement of the *P. davidsonii* genome. In particular, constructing a chromosomal-level assembly from these scaffolds will be necessary for determining properties such as recombination rate variation across the genome (e.g. Smukowski and Noor 2011) and addressing issues such as the evolution of structural variants (e.g. Ostevik et al. 2020) and genomic patterns of introgression (e.g. Martin and Jiggins 2017).

Data availability

All raw genomic data (PacBio WGS, Hi-C, and RNA-seq) and the annotated reference genome produced for this study have been deposited into the NCBI Sequence Read Archive under the accession number PRJNA1010203. The version of the genome and annotation described in this paper is accession JAYDYQ010000000. The assemblies, annotation files, and scripts for analysis can also be found on Dryad (doi:10.5061/dryad.4f4qrfjrr).

Acknowledgments

The authors would like to thank Kari Ostevik and Kimmy Stanton for their help in the field and Ben Stone for suggestions about this project. We also thank the staff at the Duke Greenhouses for plant care and maintenance, the staff at IgenBio for Hi-C sequencing and help with scaffolding, and the Duke Center for Genomic and Computational Biology Sequencing and Genomic Technologies Core for sequencing and guidance. The authors also thank the members of the Georgia Genomics and Bioinformatics Core (GGBC) as well as the Georgia Advanced Computing Resource Center (GACRC) for their help.

Funding

This work was supported by National Science Foundation grants DEB-1542387 and IOS-1555434 to MDR.

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

- Aboul-Maaty NAF, Oraby HAS. 2019. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bull Natl Res Cent.* 43(1):25. doi:[10.1186/s42269-019-0066-1](https://doi.org/10.1186/s42269-019-0066-1).
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410. doi:[10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2).
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27(2):573–580. doi:[10.1093/nar/27.2.573](https://doi.org/10.1093/nar/27.2.573).
- Broderick SR, Stevens MR, Geary B, Love SL, Jellen EN, Dockter RB, Daley SL, Lindgren DT. 2011. A survey of *Penstemon*'s genome size. *Genome.* 54(2):160–173. doi:[10.1139/g10-106](https://doi.org/10.1139/g10-106).
- Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinform.* 48(1):4.11.1–4.11.39. doi:[10.1002/0471250953.bi0411s48](https://doi.org/10.1002/0471250953.bi0411s48).
- Castellanos MC, Wilson P, Keller SJ, Wolfe AD, Thomson JD. 2006. Anther evolution: pollen presentation strategies when pollinators differ. *Am Nat.* 167(2):288–296. doi:[10.1086/498854](https://doi.org/10.1086/498854).
- Chabot BF, Billings WD. 1972. Origins and ecology of the Sierran alpine flora and vegetation. *Ecol Monogr.* 42(2):163–199. doi:[10.2307/1942262](https://doi.org/10.2307/1942262).
- Chan PP, Lowe TM. 2019. Gene prediction, methods and protocols. *Methods Mol Biol.* 1962:1–14. doi:[10.1007/978-1-4939-9173-0_1](https://doi.org/10.1007/978-1-4939-9173-0_1).
- Clark DV. 1971. Speciation in *Penstemon* (Scrophulariaceae). Missoula (MT): University of Montana.
- Clarke JD. 2009. Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harb Protoc.* 2009(3):pdb.prot5177. doi:[10.1101/pdb.prot5177](https://doi.org/10.1101/pdb.prot5177).
- Clausen J, Keck DD, Hiesey WM. 1940. *Experimental Studies on the Nature of Species. I. Effect of Varied Environments on Western North American Plants.* Washington (DC): Carnegie Institution of Washington.
- Datwyler SL, Wolfe AD. 2004. Phylogenetic relationships and morphological evolution in *Penstemon* subg. *Dasanthera* (Veronicaceae). *Syst Bot.* 29(1):165–176. doi:[10.1600/036364404772974077](https://doi.org/10.1600/036364404772974077).
- Dockter RB, Elzinga DB, Geary B, Maughan PJ, Johnson LA, Tumbleson D, Franke J, Dockter K, Stevens MR. 2013. Developing molecular tools and insights into the *Penstemon* genome using genomic reduction and next-generation sequencing. *BMC Genet.* 14(1):66. doi:[10.1186/1471-2156-14-66](https://doi.org/10.1186/1471-2156-14-66).
- Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* 356(6333):92–95. doi:[10.1126/science.aal3327](https://doi.org/10.1126/science.aal3327).
- Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3(1):95–98. doi:[10.1016/j.cels.2016.07.002](https://doi.org/10.1016/j.cels.2016.07.002).
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci.* 117(17):9451–9457. doi:[10.1073/pnas.1921046117](https://doi.org/10.1073/pnas.1921046117).
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 28(23):3150–3152. doi:[10.1093/bioinformatics/bts565](https://doi.org/10.1093/bioinformatics/bts565).
- García Y, Ostevik KL, Anderson J, Rausher MD, Parachnowitsch AL. 2023. Floral scent divergence across an elevational hybrid zone with varying pollinators. *Oecologia.* 201(1):45–57. doi:[10.1007/s00442-022-05289-3](https://doi.org/10.1007/s00442-022-05289-3).
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652. doi:[10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883).
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36(Database):D154–D158. doi:[10.1093/nar/gkm952](https://doi.org/10.1093/nar/gkm952).
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 36(9):2896–2898. doi:[10.1093/bioinformatics/btaa025](https://doi.org/10.1093/bioinformatics/btaa025).
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8(8):1494–1512. doi:[10.1038/nprot.2013.084](https://doi.org/10.1038/nprot.2013.084).
- Horne EC, Kumpatla SP, Patterson KA, Gupta M, Thompson SA. 2004. Improved high-throughput sunflower and cotton genomic DNA extraction and PCR fidelity. *Plant Mol Biol Report.* 22(1):83–84. doi:[10.1007/BF02773352](https://doi.org/10.1007/BF02773352).
- Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics.* 36(7):2253–2255. doi:[10.1093/bioinformatics/btz891](https://doi.org/10.1093/bioinformatics/btz891).
- Inglis PW, Pappas MCR, Resende LV, Grattapaglia D. 2018. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One.* 13(10):e0206085. doi:[10.1371/journal.pone.0206085](https://doi.org/10.1371/journal.pone.0206085).
- Katzer AM, Wessinger CA, Hileman LC. 2019. Nectary size is a pollination syndrome trait in *Penstemon*. *N Phytol.* 223(1):377–384. doi:[10.1111/nph.15769](https://doi.org/10.1111/nph.15769).
- Kimball S. 2008. Links between floral morphology and floral visitors along an elevational gradient in a *Penstemon* hybrid zone. *Oikos.* 117(7):1064–1074. doi:[10.1111/j.0030-1299.2008.16573.x](https://doi.org/10.1111/j.0030-1299.2008.16573.x).
- Krueger F. 2015. Trim Galore!: a wrapper around cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files. [accessed 2020 Aug 19]. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 3997. <https://doi.org/10.48550/arxiv.1303.3997>, preprint: not peer reviewed.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJ.* 17(1):10–12. doi:[10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200).
- Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. *Curr Opin Genet Dev.* 47:69–74. doi:[10.1016/j.gde.2017.08.007](https://doi.org/10.1016/j.gde.2017.08.007).
- Nold R. 1999. *Penstemons.* Portland (OR): Timber Press.
- Nurk S, Walenz BP, Rhie A, Vollger MR, Logsdon GA, Grothe R, Miga KH, Eichler EE, Phillippy AM, Koren S. 2020. Hicanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* 30(9):1291–1305. doi:[10.1101/gr.263566.120](https://doi.org/10.1101/gr.263566.120).
- Ostevik KL, Samuk K, Rieseberg LH. 2020. Ancestral reconstruction of karyotypes reveals an exceptional rate of nonrandom chromosomal evolution in sunflower. *Genetics.* 214(4):1031–1045. doi:[10.1534/genetics.120.303026](https://doi.org/10.1534/genetics.120.303026).
- Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Report.* 15(1):8–15. doi:[10.1007/BF02772108](https://doi.org/10.1007/BF02772108).
- Ricks NJ, Stettler JM, Stevens MR. 2017. The complete plastome sequence of *Penstemon fruticosus* (Pursh) Greene (Plantaginaceae).

- Mitochondrial DNA B Resour. 2(2):768–769. doi:10.1080/23802359.2017.1398620.
- Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Sherry ST, Yankie L, Karsch-Mizrachi I. 2023. GenBank 2023 update. Nucleic Acids Res. 51(D1):141–144. doi:10.1093/nar/gkac1012.
- Schlenk N. 2023. The *Penstemon kunthii* draft genome: integrating a genetic map with assembled sequence data [doctoral dissertation]. Lawrence (KS): University of Kansas.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210–3212. doi:10.1093/bioinformatics/btv351.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinform. 6(1):31. doi:10.1186/1471-2105-6-31.
- Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. [accessed 2020 Aug 19]. <http://www.repeatmasker.org/>
- Smukowski CS, Noor MAF. 2011. Recombination rate variation in closely related species. Heredity (Edinb). 107(6):496–508. doi:10.1038/hdy.2011.44.
- Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33(Web Server):W465–W467. doi:10.1093/nar/gki458.
- Stettler JM, Stevens MR, Meservey LM, Crump WW, Grow JD, Porter SJ, Love LS, Maughan PJ, Jellen EN. 2021. Improving phylogenetic resolution of the Lamiales using the complete plastome sequences of six *Penstemon* species. PLoS One. 16(12):e0261143. doi:10.1371/journal.pone.0261143.
- Stone BW, Rodríguez-Peña RA, Wolfe AD. 2023. Testing hypotheses of hybrid taxon formation in the shrubby beardtongues (*Penstemon* subgenus *Dasanthera*). Am J Bot. 110(1):e16118. doi:10.1002/ajb2.16118.
- Stone BW, Wolfe AD. 2021. Phylogeographic analysis of shrubby beardtongues reveals range expansions during the last glacial Maximum and implicates the Klamath Mountains as a hotspot for hybridization. Mol Ecol. 30(15):3826–3839. doi:10.1111/mec.15992.
- Straw RM. 1955. Hybridization, homogamy, and sympatric speciation. Evolution. 9(4):441–444. doi:10.1111/j.1558-5646.1955.tb01553.x.
- Thomson JD, Wilson P. 2008. Explaining evolutionary shifts between bee and hummingbird pollination: convergence, divergence, and directionality. Int J Plant Sci. 169(1):23–38. doi:10.1086/523361.
- van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES. 2010. Hi-C: a method to study the three-dimensional architecture of genomes. J Vis Exp (39):1869. doi:10.3791/1869.
- Wessinger CA, Hileman LC, Rausher MD. 2014. Identification of major quantitative trait loci underlying floral pollination syndrome divergence in *Penstemon*. Philos Trans R Soc Lond B Biol Sci. 369(1648):20130349. doi:10.1126/science.1148428.
- Wessinger CA, Katzer AM, Hime PM, Rausher MD, Kelly JK, Hileman LC. 2023. A few essential genetic loci distinguish *Penstemon* species with flowers adapted to pollination by bees or hummingbirds. PLoS Biol. 21(9):e3002294. doi:10.1371/journal.pbio.3002294.
- Wessinger CA, Rausher MD. 2014. Predictability and irreversibility of genetic changes associated with flower color evolution in *Penstemon barbatus*. Evolution. 68(4):1058–1070. doi:10.1111/evo.12340.
- Wessinger CA, Rausher MD. 2015. Ecological transition predictably associated with gene degeneration. Mol Biol Evol. 32(2):347–354. doi:10.1093/molbev/msu298.
- Wilson P, Castellanos MC, Hogue JN, Thomson JD, Armbruster WS. 2004. A multivariate search for pollination syndromes among penstemons. Oikos. 104(2):345–361. doi:10.1111/j.0030-1299.2004.12819.x.
- Wilson P, Wolfe AD, Armbruster WS, Thomson JD. 2007. Constrained lability in floral evolution: counting convergent origins of hummingbird pollination in *Penstemon* and *Keckiella*. New Phytol. 176(4):883–890. doi:10.1111/j.1469-8137.2007.02219.x.
- Wolfe AD, Blischak PD, Kubatko LS. 2021. Phylogenetics of a rapid, continental radiation: diversification, biogeography, and circumscription of the beardtongues (*Penstemon*; Plantaginaceae). bioRxiv 440652. <https://doi.org/10.1101/2021.04.20.440652>, preprint: not peer reviewed.
- Wolfe AD, Randle CP, Datwyler SL, Morawetz JJ, Arguedas N, Diaz J. 2006. Phylogeny, taxonomic affinities, and biogeography of *Penstemon* (Plantaginaceae) based on ITS and cpDNA sequence data. Am J Bot. 93(11):1699–1713. doi:10.3732/ajb.93.11.1699.
- Wolfe AD, Xiang Q-Y, Kephart SR. 1998. Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). Proc National Acad Sci. 95(9):5112–5115. doi:10.1073/pnas.95.9.5112.