Genome sequence of the ornamental plant *Aquilegia vulgaris* reveals the flavonoid biosynthesis gene repertoire

Ronja Friedhoff[1+], Katharina Wolff[1+], Boas Pucker[1*]

[1] Plant Biotechnology and Bioinformatics, Institute of Plant Biology & BRICS, TU Braunschweig, Mendelssohnstrasse 4, 38106 Braunschweig, Germany

+ these authors contributed equally
* corresponding author: b.pucker@tu-braunschweig.de

RF: 0009-0002-0270-4661
KW: 0000-0002-3616-0482
BP: 0000-0002-3321-7471

Abstract

Aquilegia vulgaris is a widespread ornamental plant. The species is well known for a huge variation of different flower colors and can even be considered as a model species for the evolution of flower morphology. Anthocyanins are a major pigment group responsible for pigmentation of flowers in *A. vulgaris* and many plant species. Here, we report a highly continuous genome sequence of a European *Aquilegia vulgaris* plant displaying purple flowers. The corresponding annotation facilitates research on flower color and morphology evolution. Candidate genes for all steps in the core anthocyanidin biosynthesis were identified and investigated with respect to their expression in different seedling parts. The discovery of a flavonoid 3'5' hydroxylase (*F3'5'H*), a gene required for the biosynthesis bluish delphinidin derivatives, matches previous reports about metabolites and transcripts in *A. vulgaris*.

Introduction

Aquilegia vulgaris, also known as Columbine or Granny's Bonnet, is cultivated as an ornamental plant in many gardens. This plant species is perennial and easy to grow. *A. vulgaris* is well known for a huge variation of different flower colors and can even be considered as a model species for the evolution of flower morphology (Kramer & Hodges, 2010). Flower color is often caused by anthocyanins which are produced by one branch of the flavonoid biosynthesis. Independently produced carotenoids can also contribute to the flower color phenotype. PCR-based searches with degenerated primers and RACE have resulted in the identification of 34 genes associated with the flavonoid biosynthesis (Hodges & Derieg, 2009). Many of these genes appear to be single copy thus providing a potential explanation for the phenotypic diversity that could result from loss-of-function mutations (Hodges & Derieg, 2009). Multiple cases of independent floral pigmentation loss have been studied in *Aquilegia* discovering low expression of the central anthocyanin biosynthesis genes dihydroflavonol 4-reductase (*DFR*)

and anthocyanidin synthase (*ANS*) (Whittall *et al.*, 2006). Anthocyanins have a wide range of functions in plants including scavenging reactive oxygen species, protecting the plant against intense light, and attracting pollinators and seed dispersers (Grünig *et al.*, 2024). In *Aquilegia*, anthocyanins have been investigated in the context of coadaptation between plants and pollinators (Taylor, 1984). Besides anthocyanins, numerous other floral traits have been investigated in the context of pollinator attraction (Edwards *et al.*, 2021). A genome sequencing approach involving multiple *Aquilegia* plants from different geological regions and producing a genome sequence for *A. coerulea* revealed a high degree of allele sharing between different species (Filiault *et al.*, 2018). Another study sequenced the genome of *A. oxysepala* var. *kansuensis* (Xie *et al.*, 2020).

Fig. 1: Purple flowering *Aquilegia vulgaris* plant. Photo credit: Jakob Horz.

While previous studies were focused on lines from North America (Whittall *et al.*, 2006; Hodges & Derieg, 2009), this study investigates a European *Aquilegia* plant. Seeds were collected from a purple flowering and a white flowering plant in a German garden to grow the individuals for the sequencing project. The reference sequence described in here, was generated based on the sequencing data of a purple flowering plant.

Results and Discussion

Genome sequence and annotation of *Aquilegia vulgaris*

We report a genome sequence of the ornamental plant *Aquilegia vulgaris*. Based on ONT long read sequencing, a highly continuous assembly was generated (Table 1). The assembly size of 324 Mbp aligns well with the results of a flow cytometry study that reported a DNA amount of 0.39 pg for 1C (Pustahija *et al.*, 2013). The N50 of 3 Mbp indicates a high continuity that is suitable for comparative genomics. The assembly completeness was indicated to be about 96% based on BUSCO genes.

Table1: Statistics of the *Aquilegia vulgaris* genome sequence produced in this study and comparison against publicly available genome sequences of other *Aquilegia* species.

	Aquilegia vulgaris (this study)	*A. coerulea* 'Goldsmith' (Filiault et al., 2018)	*A. oxysepala* var. *kansuensis* (Xie et al., 2020)
Total assembly size	323.7 Mbp	291.7 Mbp	297.8 Mbp
Number of sequences	324 (contigs)	2529 (scaffolds)	852 (contigs)
Maximal contig length	16.8 Mbp	-	7.8 Mbp
N50	3.1 Mbp	3.1 Mbp	2.2 Mbp
BUSCO v5.6.1	C:95.7% [S:83.9%,D:11.8%], F:1.8%,M:2.5%, n:1614	-	-

The structural annotation of all protein encoding genes facilitates studies of the genetic basis of flower color formation and other floral traits. Different approaches to identify protein encoding genes have been attempted (**Table 2**). GeMoMa generated the best structural annotation based on hints derived from various related plant species. A completeness check of the annotation completeness with BUSCO revealed about over 95% of all expected genes. The total number of genes is also close to the average number of genes in a plant species (Pucker & Brockington, 2018).

Table 2: Comparison of different structural annotation approaches. BRAKER3 was run with RNA-seq hints. GeMoMa was supplied with RNA-seq hints and data sets of *Coptis chinensis* (GCA_015680905.1) (Liu *et al.*, 2021), *Aquilegia coerulea* v3.1 (Acoerulea_322, GCA_002738505.1) (Filiault *et al.*, 2018), *Thalictrum thalictroides* (GCA_013358455.1) (Arias *et al.*, 2021), and *Ranunculus cassubicifolius* (Karbstein *et al.*, 2024).

	BRAKER3	GeMoMa (filtered)
Number of genes	68419	27302
Number of transcripts	73913	59612
BUSCO v5.6.1	C:91.8%[S:67.8%,D:24.0%], F:4.4%,M:3.8%,n:1614	C:94.1%[S:48.3%,D:45.8%], F:1.4%,M:4.5%,n:1614

Flavonoid biosynthesis in *Aquilegia vulgaris*

The flavonoid biosynthesis associated genes have been identified (**Table 3**). As expected for an anthocyanin-pigmented plant, genes for all steps in the anthocyanin biosynthesis were identified in the *A. vulgaris* gene set. This aligns with a previous study that reported 34 genes associated with the anthocyanin biosynthesis (Hodges & Derieg, 2009). The presence of F3'5'H is expected in this plant as delphinidin derivatives are expected to contribute to the purple coloration. In a previous study, *F3'5'H* of the closely related *A. buergeriana* was utilized in an attempt to engineer pigmentation of *Petunia hybrida* (Lee *et al.*, 2023). The presence of F3'H and F3'5'H in *A. vulgaris* enables future investigations of the evolution of this cytochrome P450 gene family. It is already established that F3'5'H evolved multiple times independently from F3'H (Seitz *et al.*, 2006, 2015), but the number of independent events remains unknown.

Table 3: Annotation of the flavonoid biosynthesis genes in *Aquilegia vulgaris*. Candidates lacking a residue considered crucial for their predicted function are marked with an asterisk.

Function	**GeneID**
CHS	AV22884
CHI	AV02347
F3H	AV00525
F3'H	AV20460,AV12434,AV20461,AV12303
F3'5'H	AV10359
DFR	AV17555
ANS	AV13293*
arGST (TT19)	AV15502
FLS	AV18520
LAR	AV11162
ANR	AV09695, AV24619
Anthocyanin MYB	AV17364
Flavonol MYB	AV04352,AV18794
Proanthocyanidin MYB	AV08268
bHLH42/*TT8*	AV03879,AV06712,AV09933,AV15110
TTG1	AV10074

An analysis of flavonoid biosynthesis gene expression was performed based on RNA-seq data of different seedling parts (**Fig. 2**, AdditionalFile 1).

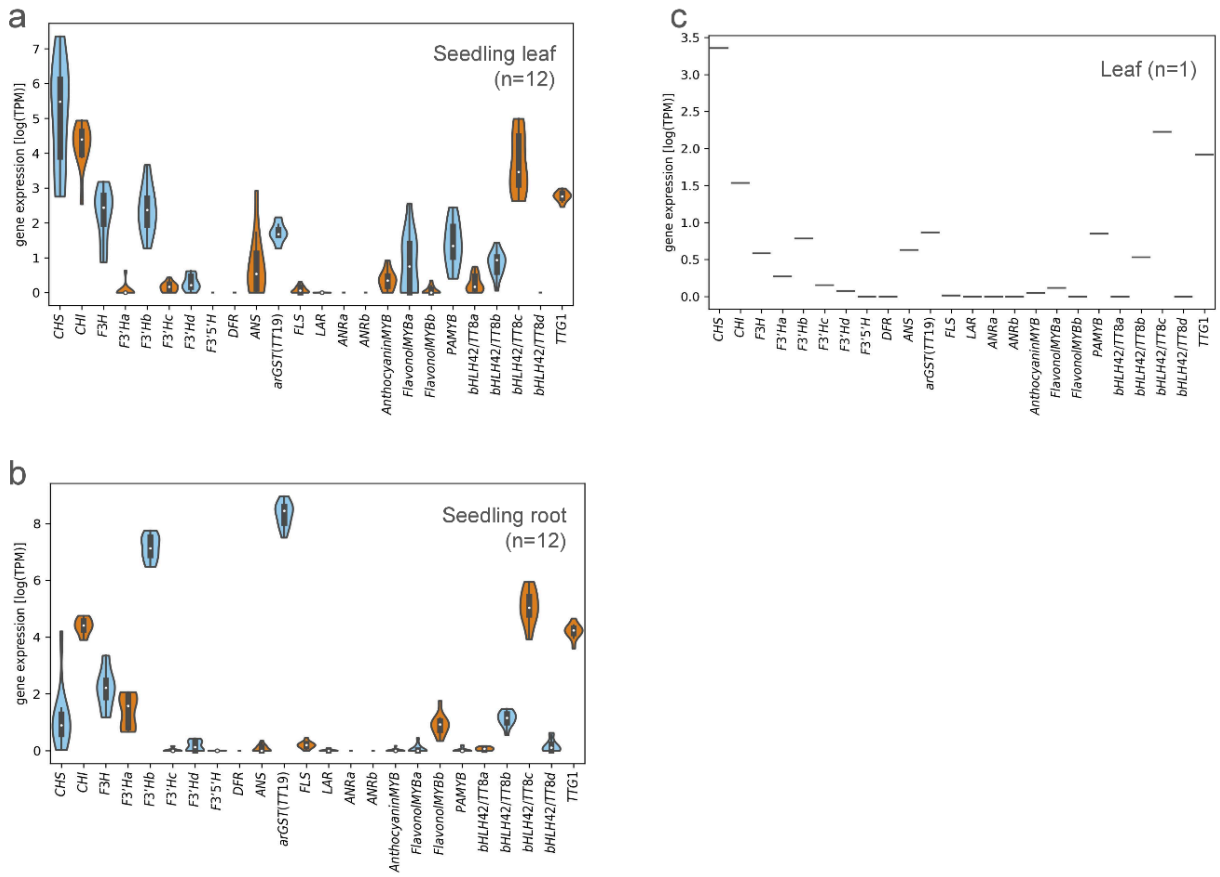

Fig. 2: Expression analysis of flavonoid biosynthesis candidate genes in the leaves of seedlings (a), the roots of seedlings (b), and adult leaves (c). Displayed expression values are log-transformed transcripts per million (TPMs).

The identified anthocyanin biosynthesis genes and corresponding transcription factors (**Fig. 3**) can serve as basis for future studies exploring the genetic basis of flower color evolution. The huge diversity of flower colors in *Aquilegia* provides an excellent opportunity to conduct in-depth investigations of molecular mechanisms underlying flower color transitions (Kramer & Hodges, 2010). Besides wholesome loss events caused often by inactivation of transcription factors (Marin-Recinos & Pucker, 2024), more subtle changes can be explored. The contribution of flower colors to the adaptive radiation of plants can be investigated through comparison of the anthocyanin biosynthesis in different *Aquilegia* species. Interestingly, no candidate for an UDP-dependent anthocyanidin 3-O-glucosyltransferase was identified by KIPEs. A previous study reported the discovery of a flavonoid glucosyltransferase belonging to the UGT75B2 lineage (homolog of AT1G05530), but does not report a dedicated anthocyanin 3-O-glucosyltransferase of the UGT78D2 type (Hodges & Derieg, 2009). While the absence of such a gene in the *A. vulgaris* gene set could be a technical artifact caused by the assembly or

annotation process, it could also indicate a different glycosylation of anthocyanins. Several other glycosyltransferases were detected that might also accept anthocyanidins as substrate.

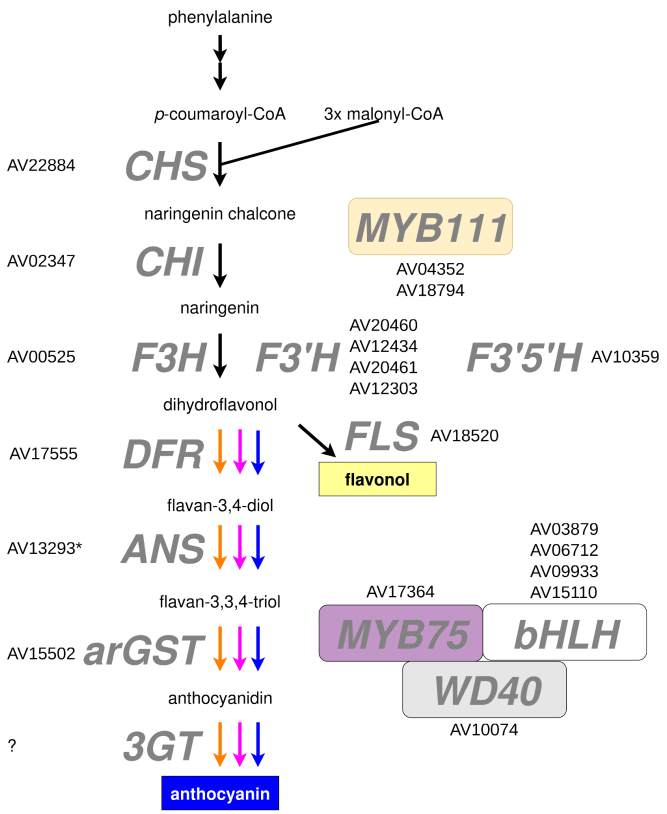

Fig. 3: Candidate genes for steps in the flavonoid biosynthesis. The formation of different anthocyanins is indicated by arrows in the respective color: pelargonidin (orange), cyanidin (magenta), and delphinidin (blue). The layout is based on (Horz *et al.*, 2024).

Material and Methods

DNA extraction and quality control

Plants were grown under long day conditions (16h light, 8 h darkness) in a plant cultivation room at approximately 20 °C. Prior to harvesting samples, the plants were incubated for one day in darkness to reduce starch content. Young leaves were harvested and homogenized by grinding in liquid nitrogen. High molecular DNA was extracted with a modified CTAB-based protocol as previously described (Siadjeu *et al.*, 2020). Quality control of the DNA via NanoDrop measurement, agarose gel electrophoresis, and Qubit measurement was performed as previously described (Horz *et al.*, 2024). Short DNA fragments were depleted with the Short Read Eliminator kit (Pacific Biosciences) following the suppliers instructions.

Nanopore sequencing

Libraries for the nanopore sequencing were prepared with the SQK-LSK109 ligation-based kit (Oxford Nanopore Technologies) using 1µg of DNA and following the suppliers' instructions. Sequencing was conducted on a MinION-Mk1B with R9.4.1 flow cells. Upon blockage of a large proportion of nanopores, a wash step was performed followed by the loading of a fresh library to achieve optimal performance of the flow cell. Basecalling of the raw sequencing data was performed with Guppy v6.4.6 (ONT) on a GPU in the de.NBI cloud.

Genome sequence assembly

The genome sequence was assembled with NextDenovo v2.5.2 (Hu *et al.*, 2024) using the following settings genome_size = 370m and read_cutoff = 1k. Assembly statistics were calculated with a customized Python script (contig_stats3.py, (Meckoni *et al.*, 2023)). Contig names were cleaned to avoid technical issues in the following steps using the Python script clean_genomic_fasta.py (https://github.com/bpucker/GenomeAssembly). A screen for contamination with bacterial or fungal sequences was conducted with assembly_wb_screen.py (https://github.com/bpucker/GenomeAssembly). BUSCO v5.6.1 (Manni *et al.*, 2021) was run in the genome mode to assess the completeness of the assembly based on the embroyphyta_odb10 reference data set.

Genome sequence annotation

To generate hints for the structural annotation, RNA-seq data sets were retrieved from the Sequence Read Archive (Leinonen *et al.*, 2011; Katz *et al.*, 2022) with fastq-dump (NCBI, 2020). Reads of the following data sets were included in this process: SRR26667430, SRR26667431, SRR26667432, SRR26667433, SRR26667434, SRR26667435, SRR26667436, SRR26667437, SRR26667438, SRR26667439, SRR26667440, and SRR26667441. Read mappings were performed with STAR v2.7.4a (Dobin *et al.*, 2013; Dobin & Gingeras, 2015). A BAM file containing the RNA-seq read mapping was passed to BRAKER3 (Gabriel *et al.*, 2023) to enable the inference of hints for the gene prediction. GeMoMa v1.9 (Keilwagen *et al.*, 2016, 2019) was run on the assembled genome sequence with the following data sets following data sets as hints: *Coptis chinensis* (GCA_015680905.1) (Liu *et al.*, 2021), *Aquilegia coerulea* v3.1 (Acoerulea_322, GCA_002738505.1) (Filiault *et al.*, 2018), *Thalictrum thalictroides* (GCA_013358455.1) (Arias *et al.*, 2021), and *Ranunculus cassubicifolius* (Karbstein *et al.*, 2024). The initial annotation was filtered with GeMoMa using the following criteria f="start=='M' and stop=='*' and (isNaN(tie) or tie>0) and tpc>0 and aa>50" atf="tie>0 and tpc>0". An analysis of all predicted polypeptide sequences with BUSCO v3.0.2 (Simão *et al.*, 2015) in protein mode was conducted to assess the annotation completeness. Functional annotation terms were assigned to the predicted polypeptide sequences with the Python script construct_anno.py (Pucker & Iorizzo, 2023) based on the Araport11 annotation of *Arabidopsis thaliana* (Lamesch *et al.*, 2012; Cheng *et al.*, 2017). A detailed annotation of the flavonoid biosynthesis genes was

performed with KIPEs v3 (Rempel *et al.*, 2023) based on the flavonoid biosynthesis bait set v3.3.7. The anthocyanin biosynthesis regulating MYBs were identified with the MYB_annotator v1.0.1 (Pucker, 2022) and the anthocyanins biosynthesis regulating bHLHs were identified with the bHLH_annotator v1.0.4 (Thoben & Pucker, 2023). A TTG1 candidate was identified based on high sequence similarity in an initial search via BLASTp (Altschul *et al.*, 1990, 1997) with TTG1 sequences from numerous plant species (Pucker *et al.*, 2020). The candidate was validated through construction of a phylogenetic tree with FastTree v2.1.10 (Price *et al.*, 2010) based on an alignment with MAFFT v7.453 (Katoh & Standley, 2013).

Gene expression analysis

All available RNA-seq data sets were processed with kallisto v0.44 (Bray *et al.*, 2016) based on the coding sequences predicted for *A. vulgaris*. Customized Python scripts were deployed to merge all individual count tables into one final combined table (Pucker & Iorizzo, 2023). Violin plots visualizing the expression of flavonoid biosynthesis genes were generated with a Python script using the modules matplotlib, pandas, scipy, numpy, and seaborn (Hunter, 2007; Virtanen *et al.*, 2020; Harris *et al.*, 2020; Waskom, 2021; Pucker & Iorizzo, 2023; The pandas development team, 2024).

Data availability

All data sets underlying this study are publicly available. Sequencing data have been deposited at the European Nucleotide Archive (PRJEB63451, AdditionalFile 2). The assembled genome sequence and corresponding annotation have been published via LeoPARD (https://leopard.tu-braunschweig.de/receive/dbbs_mods_00078110).

Acknowledgements

This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A). We also thank all students participating in our 'Data Literacy in Genome Research' course and members of the research group Plant Biotechnology and Bioinformatics for discussion and support. We thank Jakob Horz for providing a picture of our *Aquilegia vulgaris* plant.

Author contributions

KW and BP designed the experiment. KW did the sequencing. RF and BP conducted the bioinformatic analyses. BP wrote the manuscript. All authors reviewed the final version of the manuscript and consented to its submission.

Additional Files

AdditionalFile 1: Gene expression information inferred from RNA-seq data.
AdditionalFile 2: Description of *Aquilegia vulgaris* nanopore sequencing data sets.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. **1990**. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410. doi: 10.1016/S0022-2836(05)80360-2.
Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. **1997**. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.
Arias T, Riaño-Pachón DM, Stilio VSD. **2021**. Genomic and transcriptomic resources for candidate gene discovery in the Ranunculids. *Applications in Plant Sciences* **9**: e11407. doi: 10.1002/aps3.11407.
Bray NL, Pimentel H, Melsted P, Pachter L. **2016**. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**: 525–527. doi: 10.1038/nbt.3519.
Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. **2017**. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *The Plant Journal* **89**: 789–804. doi: 10.1111/tpj.13415.
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. **2013**. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi: 10.1093/bioinformatics/bts635.
Dobin A, Gingeras TR. **2015**. Mapping RNA-seq Reads with STAR. *Current Protocols in Bioinformatics* **51**: 11.14.1-11.14.19. doi: 10.1002/0471250953.bi1114s51.
Edwards MB, Choi GPT, Derieg NJ, Min Y, Diana AC, Hodges SA, Mahadevan L, Kramer EM, Ballerini ES. **2021**. Genetic architecture of floral traits in bee- and hummingbird-pollinated sister species of Aquilegia (columbine). *Evolution* **75**: 2197–2216. doi: 10.1111/evo.14313.
Filiault DL, Ballerini ES, Mandáková T, Aköz G, Derieg NJ, Schmutz J, Jenkins J, Grimwood J, Shu S, Hayes RD, *et al.*2018. The Aquilegia genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *eLife* **7**: e36426. doi: 10.7554/eLife.36426.
Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. **2023**. BRAKER3: Fully Automated Genome Annotation Using RNA-Seq and Protein Evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. : 2023.06.10.544449. doi: 10.1101/2023.06.10.544449.
Grünig N, Horz JM, Pucker B. **2024**. Diversity and Ecological Functions of Anthocyanins. doi: 10.20944/preprints202408.2272.v1.
Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, *et al.*2020. Array programming with NumPy. *Nature* **585**: 357–362. doi: 10.1038/s41586-020-2649-2.
Hodges SA, Derieg NJ. **2009**. Adaptive radiations: from field to genomic studies. *Proceedings of the National Academy of Sciences of the United States of America* **106 Suppl 1**: 9947–9954. doi: 10.1073/pnas.0901594106.
Horz JM, Wolff K, Friedhoff R, Pucker B. **2024**. Genome sequence of the ornamental plant Digitalis purpurea reveals the molecular basis of flower color and morphology variation. : 2024.02.14.580303. doi: 10.1101/2024.02.14.580303.
Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K, *et al.*2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biology* **25**: 107. doi: 10.1186/s13059-024-03252-4.
Hunter JD. **2007**. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**: 90–95. doi: 10.1109/MCSE.2007.55.
Karbstein K, Choudhary N, Xie T, Tomasello S, Wagner ND, Barke BH, Paetzold C, Bradican JP, Preick M, Himmelbach A, *et al.*2024. Efficient assembly of plant genomes: A

case study with evolutionary implications in Ranunculus (Ranunculaceae). : 2023.08.08.552429. doi: 10.1101/2023.08.08.552429.
Katoh K, Standley DM. **2013**. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**: 772–780. doi: 10.1093/molbev/mst010.
Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. **2022**. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Research* **50**: D387–D390. doi: 10.1093/nar/gkab1053.
Keilwagen J, Hartung F, Grau J. **2019**. GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. *Methods in Molecular Biology (Clifton, N.J.)* **1962**: 161–177. doi: 10.1007/978-1-4939-9173-0_9.
Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. **2016**. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research* **44**: e89. doi: 10.1093/nar/gkw092.
Kramer EM, Hodges SA. **2010**. Aquilegia as a model system for the evolution and ecology of petals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 477. doi: 10.1098/rstb.2009.0230.
Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, *et al.*2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**: D1202–D1210. doi: 10.1093/nar/gkr1090.
Lee YA, Cheon KS, Shin JY, Kim JH, Song B, Kim SJ, Park PM, An HR, Kim YJ, Lee J, *et al.*2023. Flower color modification through expression of Aquilegia buergeriana F3′5′H in Petunia hybrida. *Horticulture, Environment, and Biotechnology* **64**: 683–694. doi: 10.1007/s13580-022-00505-8.
Leinonen R, Sugawara H, Shumway M, on behalf of the International Nucleotide Sequence Database Collaboration. **2011**. The Sequence Read Archive. *Nucleic Acids Research* **39**: D19–D21. doi: 10.1093/nar/gkq1019.
Liu Y, Wang B, Shu S, Li Z, Song C, Liu D, Niu Y, Liu J, Zhang J, Liu H, *et al.*2021. Analysis of the Coptis chinensis genome reveals the diversification of protoberberine-type alkaloids. *Nature Communications* **12**: 3276. doi: 10.1038/s41467-021-23611-0.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. **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. doi: 10.1093/molbev/msab199.
Marin-Recinos MF, Pucker B. **2024**. Genetic factors explaining anthocyanin pigmentation differences. *BMC Plant Biology* **24**: 627. doi: 10.1186/s12870-024-05316-w.
Meckoni SN, Nass B, Pucker B. **2023**. Phylogenetic placement of Ceratophyllum submersum based on a complete plastome sequence derived from nanopore long read sequencing data. *BMC Research Notes* **16**: 187. doi: 10.1186/s13104-023-06459-z.
NCBI. **2020**. sra-tools.
Price MN, Dehal PS, Arkin AP. **2010**. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**: e9490. doi: 10.1371/journal.pone.0009490.
Pucker B. **2022**. Automatic identification and annotation of MYB gene family members in plants. *BMC Genomics* **23**: 220. doi: 10.1186/s12864-022-08452-5.
Pucker B, Brockington SF. **2018**. Genome-wide analyses supported by RNA-Seq reveal non-canonical splice sites in plant genomes. *BMC Genomics* **19**: 980. doi: 10.1186/s12864-018-5360-z.
Pucker B, Iorizzo M. **2023**. Apiaceae FNS I originated from F3H through tandem gene duplication. *PLOS ONE* **18**: e0280155. doi: 10.1371/journal.pone.0280155.
Pucker B, Reiher F, Schilbert HM. **2020**. Automatic Identification of Players in the Flavonoid

Biosynthesis with Application on the Biomedicinal Plant Croton tiglium. *Plants* **9**: 1103. doi: 10.3390/plants9091103.
Pustahija F, Brown SC, Bogunić F, Bašić N, Muratović E, Ollier S, Hidalgo O, Bourge M, Stevanović V, Siljak-Yakovlev S. **2013**. Small genomes dominate in plants growing on serpentine soils in West Balkans, an exhaustive study of 8 habitats covering 308 taxa. *Plant and Soil* **373**: 427–453.
Rempel A, Choudhary N, Pucker B. **2023**. KIPEs3: Automatic annotation of biosynthesis pathways. *PLOS ONE* **18**: e0294342. doi: 10.1371/journal.pone.0294342.
Seitz C, Ameres S, Schlangen K, Forkmann G, Halbwirth H. **2015**. Multiple evolution of flavonoid 3',5'-hydroylase. *Planta* **242**: 561–573.
Seitz C, Eder C, Deiml B, Kellner S, Martens S, Forkmann G. **2006**. Cloning, functional identification and sequence analysis of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase cDNAs reveals independent evolution of flavonoid 3',5'-hydroxylase in the Asteraceae family. *Plant Molecular Biology* **61**: 365–381. doi: 10.1007/s11103-006-0012-0.
Siadjeu C, Pucker B, Viehöver P, Albach DC, Weisshaar B. **2020**. High Contiguity de novo Genome Sequence Assembly of Trifoliate Yam (Dioscorea dumetorum) Using Long Read Sequencing. *Genes* **11**: 274. doi: 10.3390/genes11030274.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. **2015**. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212. doi: 10.1093/bioinformatics/btv351.
Taylor RJ. **1984**. Floral Anthocyanins of Aquilegia and Their Relationship to Distribution and Pollination Biology of the Species. *Bulletin of the Torrey Botanical Club* **111**: 462–468. doi: 10.2307/2995896.
The pandas development team. **2024**. Pandas. doi: 10.5281/zenodo.13819579.
Thoben C, Pucker B. **2023**. Automatic annotation of the bHLH gene family in plants. *BMC Genomics* **24**: 780. doi: 10.1186/s12864-023-09877-2.
Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, *et al.*2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**: 261–272. doi: 10.1038/s41592-019-0686-2.
Waskom M. **2021**. seaborn: statistical data visualization. *Journal of Open Source Software* **6**: 3021. doi: 10.21105/joss.03021.
Whittall JB, Voelckel C, Kliebenstein DJ, Hodges SA. **2006**. Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in Aquilegia. *Molecular Ecology* **15**: 4645–4657. doi: 10.1111/j.1365-294X.2006.03114.x.
Xie J, Zhao H, Li K, Zhang R, Jiang Y, Wang M, Guo X, Yu B, Kong H, Jiao Y, *et al.*2020. A chromosome-scale reference genome of Aquilegia oxysepala var. kansuensis. *Horticulture Research* **7**: 1–13. doi: 10.1038/s41438-020-0328-y.