White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/218529/

Version: Published Version

Article:
Smit, Samuel J, Whitehead, Caragh, James, Sally R et al. (2024) Pseudomolecule-scale genome assemblies of Drepanocaryum sewerzowii and Marmoritis complanata. G3: Genes, Genomes, Genetics. jkae172. ISSN: 2160-1836

https://doi.org/10.1093/g3journal/jkae172

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

G3, 2024, 14(10), jkae172
https://doi.org/10.1093/g3journal/jkae172
Advance Access Publication Date: 24 July 2024
Genome Report

Pseudomolecule-scale genome assemblies of *Drepanocaryum sewerzowii* and *Marmoritis complanata*

Samuel J. Smit [1,*] Caragh Whitehead,[1] Sally R. James,[2] Daniel C. Jeffares [3] Grant Godden [4] Deli Peng,[5,6] Hang Sun [7] Benjamin R. Lichman[1,*]

[1]Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK
[2]Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, UK
[3]York Biomedical Research Institute, Department of Biology, University of York, York YO10 5DD, UK
[4]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
[5]School of Life Science, Yunnan Normal University, Kunming 650092, Yunnan, China
[6]Key Laboratory of Yunnan for Biomass Energy and Biotechnology of Environment, Yunnan Normal University, Kunming 650500, China
[7]Key Laboratory for Plant Diversity and Biogeography of East Asia/Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

*Corresponding author: Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK. Email: cobus.smit@york.ac.uk;
*Corresponding author: Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK. Email: benjamin.lichman@york.ac.uk

The Nepetoideae, a subfamily of Lamiaceae (mint family), is rich in aromatic plants, many of which are sought after for their use as flavors and fragrances or for their medicinal properties. Here, we present genome assemblies for two species in Nepetiodeae: *Drepanocaruym sewerzowii* and *Marmoritis complanata*. Both assemblies were generated using Oxford Nanopore Q20 + reads with contigs anchored to nine pseudomolecules that resulted in 335 Mb and 305 Mb assemblies, respectively, and BUSCO scores above 95% for both the assembly and annotation. We furthermore provide a species tree for the Lamiaceae using only genome-derived gene models, complementing existing transcriptome and marker-based phylogenies.

Keywords: Lamiaceae; Nepetinae; chromosome-level assembly; nanopore sequencing; Hi-C sequencing

Introduction

The mint family (Lamiaceae) is the sixth largest plant family with a number of species regarded as important for medicinal, aromatic, and ornamental properties (Harley *et al.* 2004; Zhao *et al.* 2021; Rose *et al.* 2022). Within the Lamiaceae, species from the Nepetoideae are renowned for the accumulation of terpenoids, with tissues used for the extraction of essential oils or as traditional herbal medicines (Wink 2003; Frezza *et al.* 2019). The clade includes widely recognized aromatic species such as mint, lavender, lemon balm, and catnip; the volatile terpenoids produced by these plants are responsible for their characteristic fragrances. The ethnobotanical and commercial relevance of this plant family has resulted in considerable scientific interest, including genome assemblies for 36 species at the time of writing (Published Plant Genomes).

Here, we present the genome assemblies for two Nepetoideae species, namely *Drepanocaryum sewerzowii* (Regel) Pojark. and *Marmoritis complanata* (Dunn) A.L.Budantzev. *M. complanata* is endemic to the subnival band of the Himalaya–Hengduan Mountains, a unique arctic–alpine region recognized as a biodiversity hotspot (Myers *et al.* 2000; Sun *et al.* 2017). This unique habitat necessitates careful control of seed germination to ensure survival (Peng *et al.* 2018). *M. complanata* and other species of the genus are also used as traditional herbal medicines to treat a variety of ailments that include digestive, reproductive, musculoskeletal, and skin disorders (Zaman *et al.* 2022). *D. sewerzowii* is native to a region that ranges from Iran to Central Asia and Pakistan and is the sole representative of this genus (Serpooshan *et al.* 2018).

These two species are part of the Nepetinae, a subtribe of the mint family (Lamiaceae, subfamily Nepetoideae, tribe Mentheae) that consists of 375 species and 9–12 genera of which *Nepeta* L. is considered the type genus encompassing 200–300 species. Other genera in this subfamily include *Dracocephalum* L., *Hymenocrater* Fisch. and C.A. Mey., *Lophanthus* Adans., *Agastache* Clayton ex Gronov., and *Schizonepeta* (Benth.) Briq. (Serpooshan *et al.* 2018; Rose *et al.* 2023). The phylogenetic relationship of *M. complanata* and *D. sewerzowii* relative to *Nepeta cataria* L., *N. racemosa* Lam., *Agastache rugosa* (Fisch. and C.A. Mey.) Kuntze and *Schizonepeta tenuifolia* (Benth.) Briq. is what prompted our efforts to assemble these genomes. We have been exploring the evolutionary, genomic, and enzymatic innovations of monoterpenoid biosynthesis in these species (Lichman *et al.* 2019, 2020; Hernández Lozada *et al.* 2022; Liu *et al.* 2023). However, the available genomic resources provide limited taxonomic coverage. The genome assemblies presented here will allow us to further explore the evolutionary innovations that have impacted terpenoid biosynthesis in the mint family.

Methods and Materials

Plant growth conditions

D. sewerzowii seeds were obtained from the Millennium Seed Bank at the Royal Botanic Gardens, Kew (serial no. 0694027).

Received on 02 May 2024; accepted on 16 July 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of The Genetics Society of America.
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.

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

M. complanata seeds were collected from Puyong Pass Shangri-la County, Yunnan Province, SW China (99°55′E, 28°24′N), 4,620 m a.s.l. (Peng *et al.* 2018). Seeds were germinated on 1% water agar in a growth room set to 16 h day length, temperature of 20 (±2) °C, relative humidity of 60% (±10%), and a NS12 light spectrum at 120 µmol m^{-2} s^{-1} PPFD using Valoya L28 LED lights (Helsinki, Finland). Once a radical emerged, the seedlings were transferred to 7 cm square pots containing Levington Advance Seed and Modular FS2 (ICL Professional Horticulture) seedling soil that was pretreated with Calypso (Bayer). Once established, a single individual was selected and maintained as a clonal population by propagation using cuttings.

Genome size and heterozygosity estimation

Genome size estimations were determined through flow cytometry (FCM) using the method of Dolezel *et al.* (2007). Briefly, the LB01 buffer was used together with *N. cataria* and *N. racemosa* tissues to prepare a reference standard with previously reported genome sizes (Mint Evolutionary Genomics Consortium 2018). A CytoFLEX LX (Beckman Coulter) FCM with a 561 nm excitation laser, 610/20 emission filter, and a flow rate of 30 µL/min was used. The threshold was set to 488 nm forward scatter to exclude instrument noise and background signal from the buffer. Fluorescence intensities of stained nuclei were used to analyze the nuclear DNA content and determine the genome size relative to the aforementioned *Nepeta* spp. reference samples.

Paired-end 150 Illumina short reads were used to estimate genome size and heterozygosity by determining k-mer frequencies with KMC v3.1.2 (Kokot *et al.* 2017) using a k-mer of 21 and transformed into a histogram using kmc_tools with a count cutoff of 15,000. Genomescope v2.0 (Ranallo-Benavidez *et al.* 2020) with ploidy set to 2 was used to estimate the genome size and level of heterozygosity.

Nucleic acid isolation

High molecular weight DNA isolation and sequencing

High molecular weight (HMW) DNA was extracted in duplicate from ~1 g of young leaf tissue using the Nucleobond HMW DNA Extraction kit (Macherey-Nagel, Germany). HMW DNA purity and concentration were assessed by Nanodrop and Qubit, whereafter the extractions were combined. Small fragment DNA elimination was performed with the Circulomics short-read eliminator kit (PacBio). Briefly, an equal volume of SRE reagent was added to the sample, and this was centrifuged for 1 h at 12,000×*g*. The pellet was washed with 70% ethanol before resuspending in TE buffer with low EDTA. DNA quality and quantity were assessed with a nanodrop spectrophotometer (Thermo Fischer Scientific), Agilent Tapestation (running genomic DNA screentape), and Qubit fluorimeter (Invitrogen). Sequencing was performed with the ligation sequencing kit SQK-LSK114 (Oxford Nanopore Technologies), as per the manufacturer's guidelines, with limited modifications; namely extending the reaction times for end preparation to 30 min at each temperature, and extending adapter ligation steps to an hour). Sequencing was performed on a single promethION FLO-PRO114 flowcell (Oxford Nanopore Technologies) per species, with nuclease flush and sample reload steps performed every 24 h through the run time. For *M. complanata* two additional runs using the SQK-LSK112 ligation sequencing kit (Oxford Nanopore Technologies) and FLO-MIN112 minION flowcells (Oxford Nanopore Technologies) were performed.

Base calling of the ONT reads was performed using guppy (Oxford Nanopore Technologies) version 6.1.5 for *D. sewerzowii* and version 6.3.9 for *M. complanata* with the super accuracy (sup) model. Read length and quality were assessed using Nanoplot (De Coster and Rademakers 2023). *D. sewerzowii* reads were filtered for a 10 kb minimum length using Nanofilt (De Coster *et al.* 2018). For *M. complanata* we combined all reads from the promethION and minION runs and then filtered using Nanofilt (De Coster *et al.* 2018) with a 3 kb length and Q15 quality cutoff.

Genomic DNA isolation and Illumina sequencing

Genomic DNA (gDNA) was extracted from 100 mg of young leaf tissue, in duplicate, using a CTAB extraction method (Doyle and Doyle 1990) and treated with RNAse A. Removal of RNA was confirmed through gel electrophoresis followed by gDNA quality and quantity assessment with a nanodrop spectrophotometer and a Qubit fluorometer (Invitrogen). A total of 508 ng and 752 ng of gDNA for *D. sewerzowii* and *M. complanata*, respectively, was sent for library preparation and paired-end Illumina sequencing with Novogene (Cambridge, UK).

RNA isolation and sequencing

RNA was extracted from 80 to 100 mg of tissue with the Direct-Zol RNA extraction kit (Zymo Research, CA, USA) as per the manufacturer guidelines. For *D. sewerzowii*, young and mature leaves, closed and open flowers, and stems were used. For *M. complanata* root, young and mature leaf and stem tissues were used. RNA quality was assessed with an Agilent bioanalyzer. Library preparation and paired-end Illumina sequencing were performed by Novogene (Cambridge, UK).

Hi-C sequencing

Freshly harvested young leaf tissue was fixed in 1% formaldehyde and washed as per the Phase Genomics (Seattle, WA, USA) sample preparation protocol. Following fixation, the tissue was flash frozen in liquid nitrogen and homogenized using a tissue lyser. The Hi-C libraries were prepared and sequenced by Phase Genomics.

Genome assembly

Filtered nanopore reads for the respective genomes were used for assembly and error correction. Both species were first assembled using Flye v2.9 (Lin *et al.* 2016; Kolmogorov *et al.* 2019; --iterations 0 and --nano-hq flags). *M. complanata* was also assembled with NECAT v0.0.1 (Chen *et al.* 2021) using the default configuration file settings. Our error correction pipeline entailed polishing with long reads by two rounds of RACON v1.5 (Vaser *et al.* 2017), with reads mapped using minimap2 v2.24 (Li 2018), followed by two rounds of MEDAKA v1.6 (Medaka: Sequence Correction Provided by ONT Research 2018) polishing. Short reads were mapped using bwa-mem v0.7.17 (Li 2013) and duplicate reads were marked using Picard v2.25.5 ("Picard Toolkit" 2019) prior to two iterative rounds of polishing with Pilon v1.23 (Walker *et al.* 2014). Short-read alignment and pairing rates were determined using flagstat command from samtools v.1.17 (Danecek *et al.* 2021).

The *M. complanata* Flye and NECAT assemblies were merged with Quickmerge v0.3 (Chakraborty *et al.* 2016; Solares *et al.* 2018) due to the low N50 scores. The overlap cutoff (-c flag) was 5 and the length cutoff (-l) was 100,000 with the NECAT assembly used as the query. The NECAT-Flye merged assembly underwent another two rounds of short read error correction using Pilon. For *M. complanata*, we purged the merged assembly of haplotigs prior to HiC scaffolding while *D. sewerzowii* was purged after HiC scaffolding. Haplotig purging was performed using the purge haplotigs pipeline (Roach *et al.* 2018). Contigs were scaffolded into pseudomolecules by Phase Genomics using the Proximo Genome

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

Scaffolding Platform. Contiguity and completeness were assessed throughout the assembly pipeline using BUSCO (Benchmarking for University Single Copy Orthologs) v5.4.2 with the embryophyta_odb10 dataset (Manni *et al.* 2021).

Genome annotation

Repeats and transposable elements were annotated using the Earl Grey v3.2 (Baril *et al.* 2023, 2024) pipeline with default settings followed by softmasking of the repeats using the maskfasta function of bedtools. The BRAKER3 pipeline (v3.0.6; Stanke *et al.* 2006, 2008; Gotoh 2008; Iwata and Gotoh 2012; Buchfink *et al.* 2015; Hoff *et al.* 2016, 2019; Kovaka *et al.* 2019; Pertea and Pertea 2020; Brůna *et al.* 2021; Bruna *et al.* 2024) was used to predict gene models using mRNA and protein evidence. For protein evidence, we generated a representative database from 52 Mint species (48 Lamiaceae and 4 from Lamiales families) using the transcriptomes from (Mint Evolutionary Genomics Consortium 2018). MMseqs2 (Steinegger and Söding 2017) was used to remove identical sequences from the database. For mRNA evidence we aligned RNAseq reads from the different tissues using STAR v2.7 (Dobin *et al.* 2013) with default settings and the "--outSAMstrandField intronMotif" flag. The respective bam outputs were merged using samtools (Danecek *et al.* 2021) and used as input for BRAKER3. The BRAKER annotation output was reformatted to GFF3 using AGAT v1.1 (Dainat *et al.* 2020) followed by extraction and translation of the longest open-reading for each predicted coding sequence. Annotation completeness was assessed using BUSCO v5.4.2 (Manni *et al.* 2021) in protein mode with the embryophyta_odb10 dataset.

Functional annotations were assigned by searching predicted proteins against the Arabidopsis TAIR10 proteome (Lamesch *et al.* 2012) and Magnoliopsida (taxon ID 3398) Swiss-Prot database (2024_03 Release; UniProt Consortium 2023). Searches were performed using DIAMOND v2.1.0 (Buchfink *et al.* 2021) with an *e*-value of 0.001 and retaining the top hit only. Functional domains were assigned with hmmsearch through HMMER v3.4 (http://hmmer.org/) against the PFam v37.0 database (Mistry *et al.* 2021; UniProt Consortium 2023) with an *e*-value cutoff of 0.001 with the highest scoring domain hit retained.

Species tree and macrosynteny analysis

Markerminer v1.0 (Chamala *et al.* 2015) was used to identify single-copy genes using predicted coding genes from representative Lamiaceae genomes (Supplementary Table 1) and *Paulownia fortunei* (Seem.) Hemsl. as an outgroup. Genes present in 26 of the 27 species were included. The MAFFT alignments generated as part of the Markerminer pipeline were trimmed for gaps using the gappyout algorithm of trimAl v1.4.1 (Capella-Gutiérrez *et al.* 2009) and concatenated into a supermatrix with partitions using the catfasta2phyml script (https://github.com/nylander/catfasta2phyml). A species-tree was inferred by maximum likelihood with partition models (Chernomor *et al.* 2016) using IQ-TREE 2 (Minh *et al.* 2020) with ModelFinder (Kalyaanamoorthy *et al.* 2017), ultrafast bootstraps (UFBoot2, ×1000; Hoang *et al.* 2018), and SH-aLRT supports (×1000; Guindon *et al.* 2010). In addition, a species tree using protein sequences was inferred using the STAG (Species Tree inference from All Genes) method of Orthofinder v2.5.4 (Emms and Kelly 2015, 2017, 2018, 2019). Pairwise macrosynteny analyses were performed against *A. rugosa* (Park *et al.* 2023) and *S. tenuifolia* (Liu *et al.* 2023) using the JCVI v.1.2.7 (Tang *et al.* 2024) implementation of MCScan (Tang *et al.* 2008). MCScan orthologs were identified in full mode with predicted protein sequences and default settings.

Expression analysis

RNAseq read alignments were evaluated with STAR v2.7 (Dobin *et al.* 2013) and assed with qualimap v2.2.1 (García-Alcalde *et al.* 2012; Okonechnikov *et al.* 2016). Qualimap reports were aggregated with MultiQC v1.13 (Ewels *et al.* 2016). Expression counts as transcripts per million (TPM) were generated using Salmon v1.10.0 (Patro *et al.* 2017). The transcript index for Salmon was generated using the full set of predicted coding sequences from BRAKER3.

Results and discussion

Genome size and heterozygosity estimation

The genome size was estimated by fluorescence of stained nuclei (Dolezel *et al.* 2007) relative to that of *N. cataria* and its previously reported genome size (Mint Evolutionary Genomics Consortium 2018). *D. sewerzowii* was estimated to be 330 Mb and *M. complanata* to be 337 Mb in size. The k-mer estimation of genome size (Supplementary Fig. 1) was similar to that of FCM with a genome size of 334 Mb and 328 Mb for *D. sewerzowii* and *M. complanata*, respectively. K-mer analysis estimated the heterozygosity at 0.08% and 0.6% for *D. sewerzowii* and *M. complanata*, respectively.

Chromosome level assemblies

We sequenced the genomes for *D. sewerzowii* and *M. complanata* using Oxford Nanopore long reads and Proximo HiC scaffolding (Phase Genomics) resulting in two chromosome-level assemblies (Fig. 1). A total of 99.24 Gb of super accurate nanopore reads were generated for *D. sewerzowii* with 80 Gb of reads being greater than 10 kb at a mean read quality (Q-score) of 16.6. The size filtered reads provided 242× coverage when using the genome size determined through FCM (Supplementary Fig. 1). The initial Flye assembly resulted in 472 contigs, an N50 of 17 Mb, a total assembly length of 333.75 Mb and a BUSCO score of 98.7%. Polishing with long and short reads reduced the number of contigs to 134 and the assembly size to 332.85 Mb while maintaining an N50 of 17 Mb. The BUSCO score increased slightly to 98.8% after polishing. HiC scaffolding orientated the assembly to 9 pseudomolecules (Supplementary Fig. 2a), which is in agreement with the chromosome counts reported by Bordbar (2023). The 9 pseudomolecules contained 97.6% of the contigs, representing 324.87 Mb of the total assembly at an N50 of 35.2 Mb and L50 of 5 (Table 1).

We obtained 59.3 Gb of reads after length and quality filtering, providing 176× coverage when using the 337 Mb FCM genome size estimation with a mean Q-score of 18. We tried various different read filtering cutoffs for both length and quality with all attempts using Flye failing to reach an N50 greater than ~335 kb. After polishing the best Flye assembly was 420 Mb in size with an N50 of 335 kb, 3,001 contigs and a BUSCO score of 98.5%, of which 17.1% were duplicated. NECAT resulted in a more contiguous genome assembly of 457 Mb with an N50 of 1 Mb, 869 contigs, and 98.6% BUSCO, of which 36.7% were duplicated. The inflated genome size and high number of duplicate BUSCO genes suggested that the fragmented assemblies contained a high number of haplotigs (contigs of a single haplotype), that would artificially inflate genome size.

In an attempt to increase the continuity of the assembly (N50 score) we merged the Flye and NECAT assemblies. The NECAT assembly had fewer contigs and greater N50 and was, therefore, selected to be the query genome with the Flye assembly used to improve the query genome. We evaluated the impact of haplotig

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

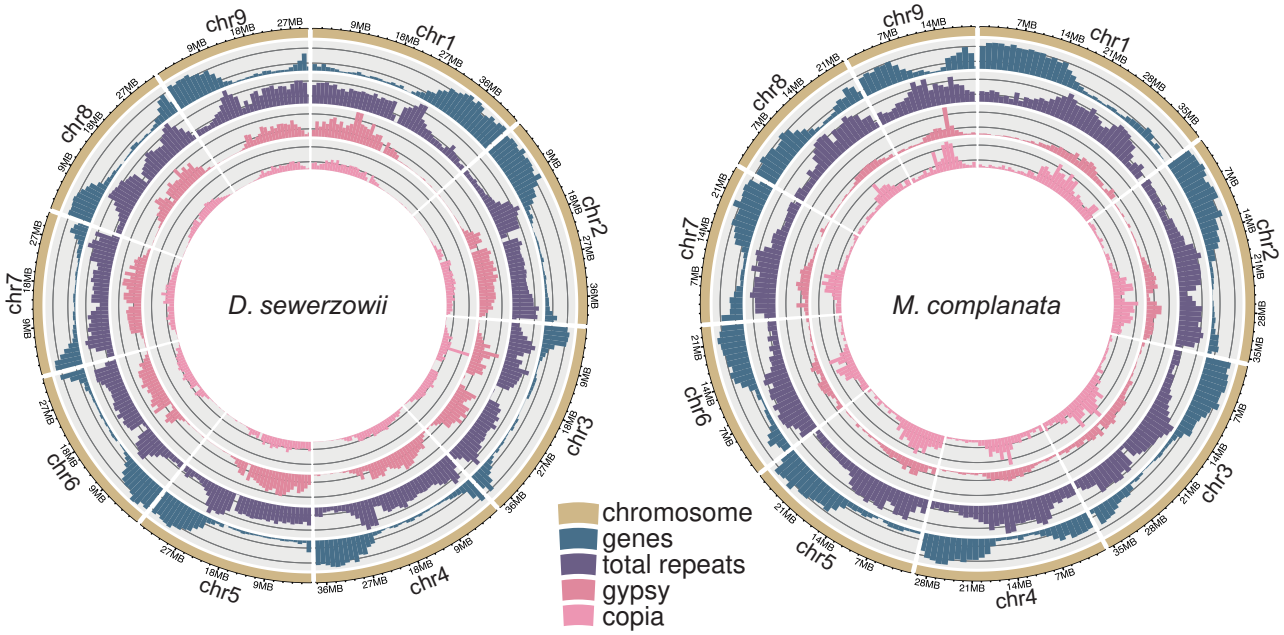


Fig. 1. Circos plots for the genome assemblies of *D. sewerzowii* and *M. complanata* depicting density (1 Mb bins) of genes, total repeats, gypsy, and copia elements along the 9 pseudomolecules.

Table 1. Assembly and annotation metrics.

	D. sewerzowii	*M. complanata*
Assembly statistics		
Assembly size (Mb)	332.85	305.55
Number of pseudomolecules	9	9
N50 (Mb)	35.13	27.69
L50	5	5
L90	9	28
GC%	38.56	37.45
Number of Ns	2,600	18,800
Mapped short-reads	95%	92.2%
Properly paired short-reads	93.8%	84.1%
Annotation statistics		
Assembly BUSCO[a]	C: 99.0%	C: 95.7%
$n = 1{,}440$	S: 96.3%	S: 85.7%
	D: 2.7%	D: 10.0%
Annotation BUSCO[a]	C: 95.0%	C: 95.3%
$n = 1{,}440$	S: 92.1%	S: 86.1%
	D: 2.9%	D: 9.2%
Predicted coding genes	24,221	25,080
Predicted proteins	26,989	28,384
Percentage repeats	Total: 62%	Total: 53%
	DNA: 2.38%	DNA: 4.43%
	LINE: 0.53%	LINE: 2.92%
	LTR: 40.61%	LTR: 25.83%

[a] Complete (C), single (S), duplicated (D).

purging before and after merging. Each assembly was purged of haplotigs prior to merging and compared to a merged assembly that was purged as the final step. In each iteration, we polished twice with short reads (18 Gb of PE150 reads) after merging. The merging increased the N50 to 3 Mb regardless of when we purged the haplotigs. The timing of the purging step had a large impact on the number of contigs together with a minor impact on the duplicated BUSCOs. Merging, polishing, and then purging the haplotigs resulted in the most contiguous assembly (305.6 Mb) with the fewest number of contigs (338) and a BUSCO score of 95.6%. HiC scaffolding assembled the contigs into 9 pseudomolecules (Supplementary Fig. 2b), which is in agreement with karyotype information (Sun 2016), totaling 258 Mb (85% of the total assembly). The pseudomolecules had a BUSCO score of 91.3% with the total assembly having a BUSCO of 95.7% (Table 1).

Pseudomolecule termini were manually inspected for the presence of the TTTAGGG telomeric repeat. Seven of the *D. sewerzowii* pseudomolecules contained this repeat on at least one end with Chr. 4 and 7 having it on both ends. For *M. complanata* we found this repeat on 6 pseudomolecules with Chr. 7 and 8 having it on both ends. The presence of this repeat on both ends indicates a telomere-to-telomere assembly for these chromosomes.

A limited number of genome assemblies are available for species within the Nepetinae with the nearest sequenced species, *N. cataria* and *N. racemosa*, having highly fragmented assemblies. We, therefore, compared our assemblies to the closest relatives with pseudomolecule assemblies, namely *A. rugosa* (Park *et al.* 2023) and *S. tenuifolia* (Liu *et al.* 2023). *A. rugosa* has a 9 chromosome assembly with macrosynteny revealing that Chr. 3, 4, and 6 have remarkably similar genomic structures for both assemblies reported here (Fig. 2). Pairwise comparisons (Supplementary Fig. 3) highlight the extent of genomic rearrangements within the Nepetinae, for example, fusion events resulted in the 6 chromosome structure of *S. tenuifolia*. *D. sewerzowii*, and *M. complanata* have very similar chromosome structures to each other (Supplementary Fig. 3); however, large relative inversions are present on *M. complanata* Chr. 3 and *D. sewerzowii* Chr. 2, with a translocation on Chr. 7. The high proportion of properly mapped and paired short-reads (Table 1) suggests that *D. sewerzowii* is a nearly complete assembly. The mapping and pairing rates for *M. complanata* short reads (Table 1) indicate a less accurate assembly, likely due to the higher level of heterozygosity (0.6%). Misassembly due to the merging of *M. complanata* NECAT and Flye assemblies may also be a reason for the lower read pairings. Higher accuracy reads and phasing will be required to fully resolve the problematic regions. Nevertheless, the *M. complanata* and *D. sewerzowii* genome assemblies provide a valuable genomic resource for intergeneric analyses.

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

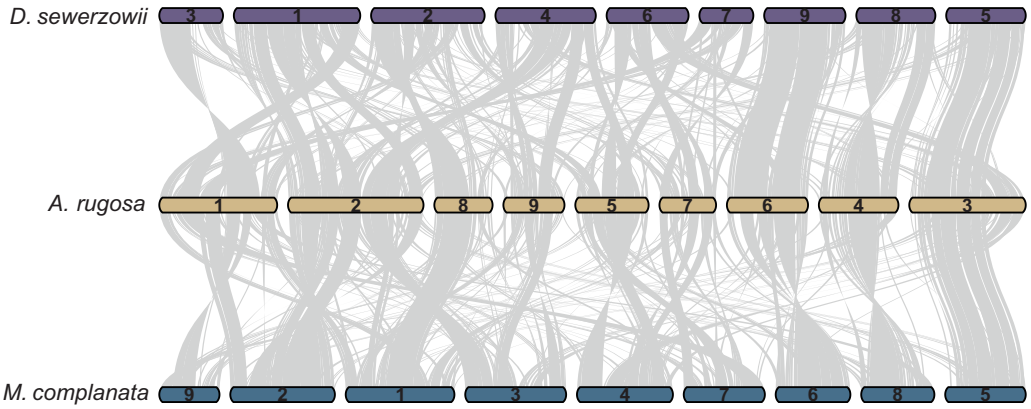


Fig. 2. Pairwise macrosynteny analysis of the assembled genomes relative to the chromosome level assembly of *A. rugosa*. Conserved collinear blocks are linked by the gray lines.

Repeat and genome annotations

Repeat annotation revealed that 62% of the *D. sewerzowii* genome and 53% of the *M. complanata* genome are repeats (Table 1). The largest portion of the repeats were long terminal repeats (LTRs), occupying 40.6% and 25.8% of the respective genomes. Subsequent to repeat masking our gene annotation, using ab initio, protein and mRNA predictions, resulted in 24,221 and 25,080 gene regions that encode for 26,989 and 28,384 proteins for the respective genomes. BUSCO analysis of the primary isoforms was 95% for both genomes. Gene and repeat density showed an inverse relationship along the chromosomes (Fig. 1).

The RNAseq data we produced found evidence for the expression of the majority of genes. RNAseq reads mapped to gene models showed that 85% (22,794/26,815) of the genes were expressed in at least 1 tissue type for *D. sewerzowii* and 88% (24,919/28,384) of the genes in *M. complanata*. Expression matrices as TPM are available in Supplementary Tables 2 and 3 with functional annotations available in Supplementary Tables 4 and 5.

Phylogenies using genome-derived gene models

The phylogenetic relationships of the Lamiaceae have been reported using plastid, nuclear, and transcriptome approaches. We used genome-derived gene models to construct a species tree via phylogenomic inference, complementing existing species trees (Fig. 3; Mint Evolutionary Genomics Consortium 2018; Serpooshan *et al.* 2018; Rose *et al.* 2022, 2023). The STAG species-tree used multicopy gene families (i.e. orthogroups) predicted by Orthofinder using protein sequences. The consensus tree (Fig. 3a) shows internal bipartition support for 5,296 orthogroups in which all species are present. The ML tree (Fig. 3b) was inferred from a single-copy gene supermatrix totaling 340,706 nucleotide sites with all but two branches showing above 98% support for both ultrafast bootstraps and SH-aLRT. The two branches indicated by the asterisk were not well supported, bootstrap and SH-aLRT <85%.

In both the ML and STAG topologies, *D. sewerzowii* is recovered as a sister to a clade that includes *M. complanata* and *Nepeta*, which are sister to each other. While our phylogenomic results corroborate existing hypotheses regarding the close relationships among these genera, our trees are incongruent with previously reported topologies (Supplementary Fig. 4). For example, nuclear phylogenetic results by Rose et al. (2023) report *D. sewerzowii* as sister to *Nepeta*, which together are sister to the sister lineages *Hymenocrater* and (*Lophanthus* + *Marmoritis*). This contrasts with plastid-based phylogenetic results reported in the same study, which recover *Nepeta* as sister to a clade comprising the sister taxa, *Drepanocaryum* and *Hymenocrater*, and their sister, (*Lophanthus* + *Marmoritis*), and with results by Serpooshan *et al.* (2018), which recover *D. sewerzowii* as sister to a mixed and partially unresolved clade of *Hymenocrater*, *Lophanthus*, *Marmoritis*, and *Nepeta*. Topological discordances among trees reported in this and previous studies likely reflect differences in taxonomic and molecular sampling, but they also highlight the complexity of resolving intergeneric relationships within Nepetinae.

Understanding these intergeneric relationships is especially relevant in the context of specialized metabolism. Evolutionary innovations in the *Mentha longifolia* and *S. tenuifolia* lineages, for example, have resulted in the convergent evolution of biosynthetic genes for monoterpenoid biosynthesis (e.g. pulegone and menthol; Srividya *et al.* 2020, 2022; Vining *et al.* 2022; Liu *et al.* 2023). Similarly, iridoid biosynthesis in *N. cataria* and *N. racemosa* evolved through the concomitant emergence of iridoid biosynthetic genes (ISY, NEPS, and MLPL) that also colocalize to form a biosynthetic gene cluster (Sherden *et al.* 2018; Lichman *et al.* 2019, 2020). Comparative genomics in these species has suggested that genomic rearrangements and gene movement through ectopic recombination or transposition are likely involved in organizing pathway genes into BGCs (Smit and Lichman 2022). The species-tree presented here (Fig. 3) provide necessary context for further comparative genomics, although interpretations should be considered alongside available transcriptome- and marker-based phylogenies until additional Nepetinae genomes and phylogenomic results become available. Nevertheless, the genomes presented here provide a valuable resource to explore the evolutionary trajectories underpinning the remarkable innovations in specialized metabolism within the Lamiaceae.

Conclusion

Plant genome assemblies are being generated at a remarkable rate, with two-thirds of available plant genome assemblies generated within the last three years (Xie *et al.* 2024). Here, we present the chromosome-level genome assemblies of *D. sewerzowii* and *M. complanata*, representing the first assemblies from these genera. The gene and repeat annotations, along with expression matrices, present a comprehensive resource for comparative genomics. The species-tree using gene models from available Lamiaceae genome assemblies provides a reference point that celebrates the number of sequenced species. These genome assemblies will allow us to decipher the evolutionary innovations that resulted in the

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

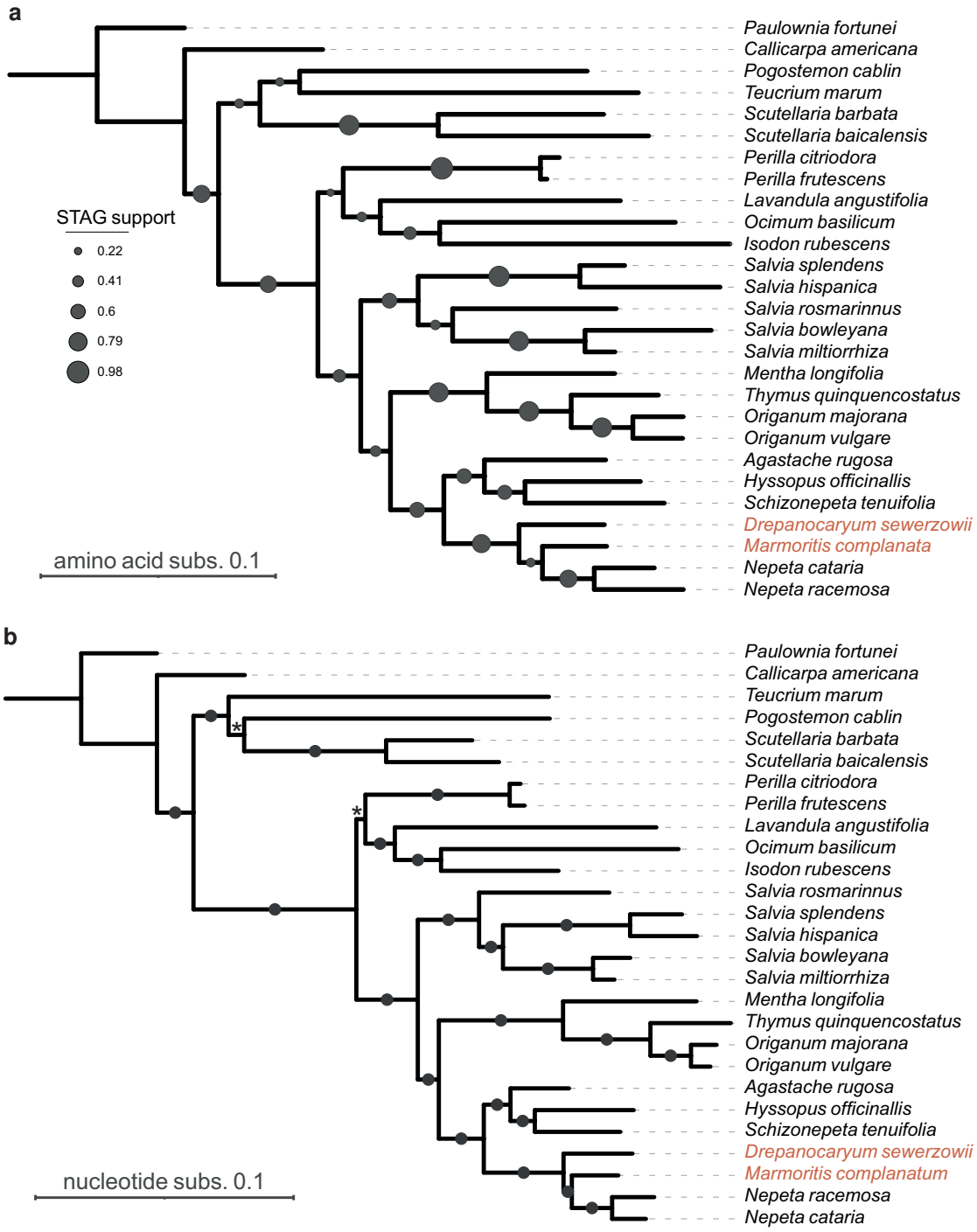


Fig. 3. Species-trees inferred with Lamiaceae genome-derived gene models. a) STAG species-tree inferred with Orthofinder protein orthogroups. Support values show the proportion of trees at which the internal bipartitions occur for all species. b) Maximum-likelihood species tree using single copy nucleotide sequences. Branches with circles are fully supported (>98%) as judged by ultrafast bootstraps and SH-aLRT. Branches indicated by the asterisk are less well supported (<85%).

remarkably diverse number of specialized metabolites found in the Lamiaceae.

Data availability

Genome assemblies are available through Genbank under accession numbers JBCJKZ000000000 and JBCLUX000000000. The raw reads for whole genome and transcriptome sequencing are available in the National Center for Biotechnology Information Sequence Read Archive BioProject PRJNA1097548 and PRJNA1095452. The genome assembly and annotation files are available through figshare as Supplementary data: https://doi.org/10.25387/g3.25671948.

Supplemental material available at G3 online.

Acknowledgments

We would like to thank Prof. C. Robin Buell for her advice and on assembling mint genomes. The Viking cluster was used during

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

this project, which is a high-performance compute facility provided by the University of York. We are grateful for computational support from the University of York, IT Services, and the Research IT team. We are grateful to the University of York Horticulture Team for the propagation and care of our plant material. We would like to thank Karen Hobb from the University of York Imaging and Cytometry Laboratory for assistance with FCM.

Funding

This work was financially supported by the Biotechnology and Biological Sciences Research Council (BB/V006452/1) and UK Research and Innovation (MR/S01862X/1).

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Baril T, Galbraith J, Hayward A. 2023. Earl Grey. Zenodo. doi:10.5281/zenodo.5654615.

Baril T, Galbraith J, Hayward A. 2024. Earl Grey: a fully automated user-friendly transposable element annotation and analysis pipeline. Mol Biol Evol. 41(4):msae068. doi:10.1093/molbev/msae068.

Bordbar F. 2023. New chromosome counts in Lamiaceae from flora of Iran—II. J Appl Biol Sci. 17(2):298–305. doi:10.5281/zenodo.8018929.

Bornowski N, Hamilton JP, Liao P, Wood JC, Dudareva N, Buell CR. 2020. Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. DNA Res. 27(3):dsaa016. doi:10.1093/dnares/dsaa016.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 3(1):lqaa108. doi:10.1093/nargab/lqaa108.

Bruna T, Lomsadze A, Borodovsky M. 2024. GeneMark-ETP: automatic gene finding in eukaryotic genomes in consistency with extrinsic data. bioRxiv 2023.01.13.524024. https://doi.org/10.1101/2023.01.13.524024, preprint: not peer reviewed.

Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 18(4):366–368. doi:10.1038/s41592-021-01101-x.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12(1):59–60. doi:10.1038/nmeth.3176.

Cao Y, Sun G, Zhai X, Xu P, Ma L, Deng M, Zhao Z, Yang H, Dong Y, Shang Z, *et al.* 2021. Genomic insights into the fast growth of paulownias and the formation of Paulownia witches' broom. Mol Plant. 14(10):1668–1682. doi:10.1016/j.molp.2021.06.021.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25(15):1972–1973. doi:10.1093/bioinformatics/btp348.

Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 44(19):e147. doi:10.1093/nar/gkw654.

Chamala S, García N, Godden GT, Krishnakumar V, Jordon-Thaden IE, De Smet R, Barbazuk WB, Soltis DE, Soltis PS. 2015. MarkerMiner 1.0: a new application for phylogenetic marker development using angiosperm transcriptomes. Appl Plant Sci. 3(4):apps.1400115. doi:10.3732/apps.1400115.

Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, Wang Y-X, Xing J-F, Huang Z-J, Wang D-P, *et al.* 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat Commun. 12(1):60. doi:10.1038/s41467-020-20236-7.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 65(6):997–1008. doi:10.1093/sysbio/syw037.

Dainat J, Hereñú D, Pucholt P. 2020. AGAT: another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. Zenodo. doi:10.5281/zenodo.3552717.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10(2):giab008. doi:10.1093/gigascience/giab008.

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 34(15):2666–2669. doi:10.1093/bioinformatics/bty149.

De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 39(5):btad311. doi:10.1093/bioinformatics/btad311.

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(1):15–21. doi:10.1093/bioinformatics/bts635.

Dolezel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc. 2(9):2233–2244. doi:10.1038/nprot.2007.310.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. Focus. 12(1):13–15.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16(1):157. doi:10.1186/s13059-015-0721-2.

Emms DM, Kelly S. 2017. STRIDE: species tree root inference from gene duplication events. Mol Biol Evol. 34(12):3267–3278. doi:10.1093/molbev/msx259.

Emms DM, Kelly S. 2018. STAG: species tree inference from all genes. bioRxiv 267914. https://doi.org/10.1101/267914, preprint: not peer reviewed.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238. doi:10.1186/s13059-019-1832-y.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32(19):3047–3048. doi:10.1093/bioinformatics/btw354.

Frezza C, Venditti A, Serafini M, Bianco A. 2019. Phytochemistry, chemotaxonomy, ethnopharmacology, and nutraceutics of Lamiaceae. Stud Nat Prod Chem. 62:125–178. doi:10.1016/B978-0-444-64185-4.00004-6.

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. 28(20):2678–2679. doi:10.1093/bioinformatics/bts503.

Gotoh O. 2008. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. Nucleic Acids Res. 36(8):2630–2638. doi:10.1093/nar/gkn105.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321. doi:10.1093/sysbio/syq010.

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

Hamilton JP, Godden GT, Lanier E, Bhat WW, Kinser TJ, Vaillancourt B, Wang H, Wood JC, Jiang J, Soltis PS, *et al.* 2020. Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing Lamiaceae species, *Callicarpa americana*. Gigascience. 9(9):giaa093. doi:10.1093/gigascience/giaa093.

Hamilton JP, Vaillancourt B, Wood JC, Wang H, Jiang J, Soltis DE, Buell CR, Soltis PS. 2023. Chromosome-scale genome assembly of the "Munstead" cultivar of *Lavandula angustifolia*. BMC Genom Data. 24(1):75. doi:10.1186/s12863-023-01181-y.

Han D, Li W, Hou Z, Lin C, Xie Y, Zhou X, Gao Y, Huang J, Lai J, Wang L, *et al.* 2023. The chromosome-scale assembly of the *Salvia rosmarinus* genome provides insight into carnosic acid biosynthesis. Plant J. 113(4):819–832. doi:10.1111/tpj.16087.

Harley R, Atkins MS, Budantsev AL, Cantino PD, Conn BJ, Grayer R, Harley MM, Kok R, Krestovskaja T, Morales R, *et al.* 2004. Labiateae. In: Kadereit JW, editor. The Families and Genera of Vascular Plants. Berlin, Heidelberg: Springer. p. 167–275.

Hernández Lozada NJ, Hong B, Wood JC, Caputi L, Basquin J, Chuang L, Kunert M, Rodríguez López CE, Langley C, Zhao D, *et al.* 2022. Biocatalytic routes to stereo-divergent iridoids. Nat Commun. 13(1):4718. doi:10.1038/s41467-022-32414-w.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522. doi:10.1093/molbev/msx281.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32(5):767–769. doi:10.1093/bioinformatics/btv661.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. Methods Mol Biol. 1962:65–95. doi:10.1007/978-1-4939-9173-0_5.

Iwata H, Gotoh O. 2012. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. 40(20):e161. doi:10.1093/nar/gks708.

Jia K-H, Liu H, Zhang R-G, Xu J, Zhou S-S, Jiao S-Q, Yan X-M, Tian X-C, Shi T-L, Luo H, *et al.* 2021. Chromosome-scale assembly and evolution of the tetraploid *Salvia splendens* (Lamiaceae) genome. Hortic Res. 8(1):177. doi:10.1038/s41438-021-00614-y.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14(6):587–589. doi:10.1038/nmeth.4285.

Kokot M, Dlugosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 33(17):2759–2761. doi:10.1093/bioinformatics/btx304.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37(5):540–546. doi:10.1038/s41587-019-0072-8.

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20(1):278. doi:10.1186/s13059-019-1910-1.

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 Res. 40(Database issue):D1202–D1210. doi:10.1093/nar/gkr1090.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio.GN]. https://doi.org/10.48550/arXiv.1303.3997, preprint: not peer reviewed.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094–3100. doi:10.1093/bioinformatics/bty191.

Lichman BR, Godden GT, Hamilton JP, Palmer L, Kamileen MO, Zhao D, Vaillancourt B, Wood JC, Sun M, Kinser TJ, *et al.* 2020. The evolutionary origins of the cat attractant nepetalactone in catnip. Sci Adv. 6(20):eaba0721. doi:10.1126/sciadv.aba0721.

Lichman BR, Kamileen MO, Titchiner GR, Saalbach G, Stevenson CEM, Lawson DM, O'Connor SE. 2019. Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis. Nat Chem Biol. 15(1):71–79. doi:10.1038/s41589-018-0185-2.

Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. 2016. Assembly of long error-prone reads using de Bruijn graphs. Proc Natl Acad Sci U S A. 113(52):E8396–E8405. doi:10.1073/pnas.1604560113.

Liu C, Smit SJ, Dang J, Zhou P, Godden GT, Jiang Z, Liu W, Liu L, Lin W, Duan J, *et al.* 2023. A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*. Mol Plant. 16(3):533–548. doi:10.1016/j.molp.2023.01.004.

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. Mol Biol Evol. 38(10):4647–4654. doi:10.1093/molbev/msab199.

Medaka: Sequence Correction Provided by ONT Research. 2018. https://github.com/nanoporetech/medaka

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37(5):1530–1534. doi:10.1093/molbev/msaa015.

Mint Evolutionary Genomics Consortium. 2018. Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. Mol Plant. 11(8):1084–1096. doi:10.1016/j.molp.2018.06.002.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, *et al.* 2021. Pfam: the protein families database in 2021. Nucleic Acids Res. 49(D1):D412–D419. doi:10.1093/nar/gkaa913.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature. 403(6772):853–858. doi:10.1038/35002501.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 32(2):292–294. doi:10.1093/bioinformatics/btv566.

Pan X, Chang Y, Li C, Qiu X, Cui X, Meng F, Zhang S, Li X, Lu S. 2023. Chromosome-level genome assembly of *Salvia miltiorrhiza* with orange roots uncovers the role of Sm2OGD3 in catalyzing 15,16-dehydrogenation of tanshinones. Hortic Res. 10(6):uhad069. doi:10.1093/hr/uhad069.

Park H-S, Jo IH, Raveendar S, Kim N-H, Gil J, Shim D, Kim C, Yu J-K, So Y-S, Chung J-W. 2023. A chromosome-level genome assembly of Korean mint (*Agastache rugosa*). Sci Data. 10(1):792. doi:10.1038/s41597-023-02714-x.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 14(4):417–419. doi:10.1038/nmeth.4197.

Peng D-L, Hu X-J, Yang J, Sun H. 2018. Seed dormancy, germination and soil seed bank of *Lamiophlomis rotata* and *Marmoritis complanatum* (Labiatae), two endemic species from Himalaya–Hengduan

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

Mountains. Plant Biosyst. 152(4):642–648. doi:10.1080/11263504.2017.1311959.
Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. F1000Res. 9:ISCB Comm J-304. doi:10.12688/f1000research.23297.2.
"Picard Toolkit." 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/
Published Plant Genomes. c2016–2024. [accessed 2024 May]. https://www.plabipd.de/plant_genomes_pa.ep
Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11(1):1432. doi:10.1038/s41467-020-14998-3.
Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 19(1):460. doi:10.1186/s12859-018-2485-7.
Rose JP, Wiese J, Pauley N, Dirmenci T, Celep F, Xiang C-L, Drew BT. 2023. East Asian-North American disjunctions and phylogenetic relationships within subtribe Nepetinae (Lamiaceae). Mol Phylogenet Evol. 187:107873. doi:10.1016/j.ympev.2023.107873.
Rose JP, Xiang C-L, Sytsma KJ, Drew BT. 2022. A timeframe for mint evolution: towards a better understanding of trait evolution and historical biogeography in Lamiaceae. Bot J Linn Soc. 200(1):15–38. doi:10.1093/botlinnean/boab104.
Serpooshan F, Jamzad Z, Nejadsattari T, Mehregan I. 2018. Molecular phylogenetics of *Hymenocrater* and allies (Lamiaceae): new insights from nrITS, plastid *trnL* intron and *trnL-F* intergenic spacer DNA sequences. Nord J Bot. 36(1_2):njb–01600. doi:10.1111/njb.01600.
Shen Y, Li W, Zeng Y, Li Z, Chen Y, Zhang J, Zhao H, Feng L, Ma D, Mo X, *et al.* 2022. Chromosome-level and haplotype-resolved genome provides insight into the tetraploid hybrid origin of patchouli. Nat Commun. 13(1):3511. doi:10.1038/s41467-022-31121-w.
Sherden NH, Lichman B, Caputi L, Zhao D, Kamileen MO, Buell CR, O'Connor SE. 2018. Identification of iridoid synthases from Nepeta species: iridoid cyclization does not determine nepetalactone stereochemistry. Phytochemistry. 145:48–56. doi:10.1016/j.phytochem.2017.10.004.
Smit SJ, Ayten S, Radzikowska BA, Hamilton JP, Langer S, Unsworth WP, Larson TR, Buell CR, Lichman BR. 2024. The genomic and enzymatic basis for iridoid biosynthesis in cat thyme (*Teucrium marum*). Plant J. 118(5):1589–1602. doi:10.1111/tpj.16698.
Smit SJ, Lichman BR. 2022. Plant biosynthetic gene clusters in the context of metabolic evolution. Nat Prod Rep. 39(7):1465–1482. doi:10.1039/D2NP00005A.
Solares EA, Chakraborty M, Miller DE, Kalsow S, Hall K, Perera AG, Emerson JJ, Hawley RS. 2018. Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. G3 (Bethesda). 3(8):3143–3154. doi:10.1534/g3.118.200162.
Srividya N, Lange I, Lange BM. 2020. Determinants of enantiospecificity in limonene synthases. Biochemistry. 59(17):1661–1664. doi:10.1021/acs.biochem.0c00206.
Srividya N, Lange I, Richter JK, Wüst M, Lange BM. 2022. Selectivity of enzymes involved in the formation of opposite enantiomeric series of p-menthane monoterpenoids in peppermint and Japanese catnip. Plant Sci. 314:111119. doi:10.1016/j.plantsci.2021.111119.
Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 24(5):637–644. doi:10.1093/bioinformatics/btn013.
Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7(1):62. doi:10.1186/1471-2105-7-62.
Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 35(11):1026–1028. doi:10.1038/nbt.3988.
Sun W-G. 2016. Karyotype of nine endemic species from alpine subnival belt in the Hengduan mountains, SW China. J Jpn Bot. 91:242–249. https://www.researchgate.net/publication/308582198
Sun Y, Shao J, Liu H, Wang H, Wang G, Li J, Mao Y, Chen Z, Ma K, Xu L, *et al.* 2023. A chromosome-level genome assembly reveals that tandem-duplicated CYP706V oxidase genes control oridonin biosynthesis in the shoot apex of *Isodon rubescens*. Mol Plant. 16(3):517–532. doi:10.1016/j.molp.2022.12.007.
Sun H, Zhang J, Deng T, Boufford DE. 2017. Origins and evolution of plant diversity in the Hengduan Mountains, China. Plant Divers. 39(4):161–166. doi:10.1016/j.pld.2017.09.004.
Sun M, Zhang Y, Zhu L, Liu N, Bai H, Sun G, Zhang J, Shi L. 2022. Chromosome-level assembly and analysis of the Thymus genome provide insights into glandular secretory trichome formation and monoterpenoid biosynthesis in thyme. Plant Commun. 3(6):100413. doi:10.1016/j.xplc.2022.100413.
Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. Science. 320(5875):486–488. doi:10.1126/science.1153917.
Tang H, Krishnakumar V, Zheng X, Xu Z, Taranto A, Lomas JS, Zhang Y, Huang Y, Wang Y, Yim WC, *et al.* 2024. JCVI: A versatile toolkit for comparative genomics analysis. iMeta. doi:10.1002/imt2.211.
UniProt Consortium. 2023. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 51(D1):D523–D531. doi:10.1093/nar/gkac1052.
Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27(5):737–746. doi:10.1101/gr.214270.116.
Vining KJ, Pandelova I, Lange I, Parrish AN, Lefors A, Kronmiller B, Liachko I, Kronenberg Z, Srividya N, Lange BM. 2022. Chromosome-level genome assembly of *Mentha longifolia* L. reveals gene organization underlying disease resistance and essential oil traits. G3 (Bethesda). 12(8):jkac112. doi:10.1093/g3journal/jkac112.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9(11):e112963. doi:10.1371/journal.pone.0112963.
Wang L, Lee M, Sun F, Song Z, Yang Z, Yue GH. 2022. A chromosome-level genome assembly of chia provides insights into high omega-3 content and coat color variation of its seeds. Plant Commun. 3(4):100326. doi:10.1016/j.xplc.2022.100326.
Wink M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry. 64(1):3–19. doi:10.1016/S0031-9422(03)00300-5.
Xie L, Gong X, Yang K, Huang Y, Zhang S, Shen L, Sun Y, Wu D, Ye C, Zhu QH, *et al.* 2024. Technology-enabled great leap in deciphering plant genomes. Nat Plants. 10(4):551–566. doi:10.1038/s41477-024-01655-6.
Xu Z, Gao R, Pu X, Xu R, Wang J, Zheng S, Zeng Y, Chen J, He C, Song J. 2020. Comparative genome analysis of *Scutellaria baicalensis* and *Scutellaria barbata* reveals the evolution of active flavonoid biosynthesis. Genomics Proteomics Bioinformatics. 18(3):230–240. doi:10.1016/j.gpb.2020.06.002.
Zaman W, Ye J, Ahmad M, Saqib S, Shinwari ZK, Chen Z. 2022. Phylogenetic exploration of traditional Chinese medicinal plants: a case study on Lamiaceae. Pak J Bot. 54(3):1033–1040. doi:10.30848/PJB2022-3(19).

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024

Zhang Y, Shen Q, Leng L, Zhang D, Chen S, Shi Y, Ning Z, Chen S. 2021. Incipient diploidization of the medicinal plant Perilla within 10,000 years. Nat Commun. 12(1):5508. doi:10.1038/s41467-021-25681-6.

Zhao F, Chen Y-P, Salmaki Y, Drew BT, Wilson TC, Scheen A-C, Celep F, Bräuchler C, Bendiksby M, Wang Q, *et al.* 2021. An updated tribal classification of Lamiaceae based on plastome phylogenomics. BMC Biol. 19(1):2. doi:10.1186/s12915-020-00931-z.

Zheng X, Chen D, Chen B, Liang L, Huang Z, Fan W, Chen J, He W, Chen H, Huang L, *et al.* 2021. Insights into salvianolic acid B biosynthesis from chromosome-scale assembly of the *Salvia bowleyana* genome. J Integr Plant Biol. 63(7):1309–1323. doi:10.1111/jipb.13085.

Editor: T. Pyhäjärvi

Downloaded from https://academic.oup.com/g3journal/article/14/10/jkae172/7719366 by guest on 17 October 2024