Title: A *de novo* long-read genome assembly of the sacred datura plant (*Datura wrightii*) reveals a role of tandem gene duplications in the evolution of herbivore-defense response

Authors: Jay K Goldberg[1]*, Aaron Olcerst[2], Michael McKibben[1], J Daniel Hare[2], Michael S Barker[1], and Judith Bronstein[1]

***Author for correspondence:** jaykgold@arizona.edu

Running Title: Datura genome paper

Author affiliations: [1] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA; [2] Department of Entomology, University of California Riverside, Riverside, CA, USA

Abstract: The sacred datura plant (Solanales: Solanaceae: *Datura wrightii*) has been used to study plant-herbivore interactions for decades. The wealth of information that has resulted leads it to have potential as a model system for studying the ecological and evolutionary genomics of these interactions. We present a *de novo Datura wrightii* genome assembled using PacBio HiFi long-reads. Our assembly is highly complete and contiguous (N50 = 179Mb, BUSCO Complete = 97.6%). We successfully detected a previously documented ancient whole genome duplication using our assembly and have classified the gene duplication history that generated its coding sequence content. We use it as the basis for a genome-guided differential expression analysis to identify the induced responses of this plant to one of its specialized herbivores (Coleoptera: Chrysomelidae: *Lema daturaphila*). We find over 3000 differentially expressed genes associated with herbivory and that elevated expression levels of over 200 genes last for several days. We also combined our analyses to determine the role that different gene duplication categories have played in the evolution of *Datura*-herbivore interactions. We find that tandem duplications have expanded multiple functional groups of herbivore responsive genes with defensive functions, including UGT-glycosyltranserases, oxidoreductase enzymes, and peptidase

23 inhibitors. Overall, our results expand our knowledge of herbivore-induced plant transcriptional

24 responses and the evolutionary history of the underlying herbivore-response genes.

25

Introduction

Datura wrightii (Solanales: Solanaceae) has begun to serve as a model system for research into the ecology and evolution of various plant traits, including both physical and chemical defenses [1,2], tolerance to herbivory [3], floral phenotypes [4,5], life histories [6,7], and ontogenetic changes in defense production throughout an individual plant's lifetime [8,9]. This is due to the wealth of ecological knowledge already gathered regarding this common plant and the specialist insects with which it interacts [10,11]. Furthermore, field-based studies have already provided some insights into the evolutionary processes maintaining a trichome dimorphism in naturally occurring populations [2,12]. As such, the publication of a reference genome for this species would accelerate research into both the molecular mechanisms governing the expression of plant traits and their evolutionary trajectories in a rapidly changing environment.

Here, we present a *de novo* genome assembly for *D. wrightii*, generated using highly accurate long-reads (PacBio HiFi) [13]. Our assembly is highly complete and contiguous when compared to recent assemblies of closely related species [14,15,16]. To demonstrate the utility of our genome to future research, we conducted two downstream analyses with it: 1) an assessment of the duplication history of coding regions with our assembly and 2) a differential expression study examining herbivore-induced transcriptional changes. We further use the results of these analyses to determine the role that different categories of gene duplications have had on the herbivore-induced responses of this species.

Gene and whole genome duplications (WGDs) have broad implications for evolutionary processes including the generation of novel traits [17,18]. WGDs have occurred numerous times throughout the diversification of angiosperms [19]; and these 'ancient' WGDs have been shown to underlie the arms-race dynamic occurring between plants and herbivores [20]. There is widespread evidence of an ancient hexaploidy event that occurred early in the evolution of the Solanaceae [19]. As

49 such we used two different analyses to infer and characterize this ancient WGD in our genome to both
50 detect its presence and assess the ancestral ploidy level of the lineage.

51 We further use our assembly to conduct a differential expression study into the induced
52 responses of *D. wrightii* to one of its specialist herbivores, the three-lined potato beetle (*Lema*
53 *daturaphila*) [10,11]. This analysis utilized older data, generated roughly a decade ago [21]. Using both
54 pairwise and time course analyses, we identified thousands of genes that are significantly differentially
55 expressed when plants are under attack by one of their closest natural enemies, and further show that
56 many of them remain upregulated for multiple days. We conclude by using an over-representation
57 analysis approach to determine the molecular functions that various gene duplication categories have
58 expanded.

59 **Methods**

60 *Initial sample collection and preparation, and sequencing*

61 *Datura wrightii* seeds were originally collected in June 2016 from a wild plant growing at the
62 intersection of Portal Rd. and Foothills Rd. in Portal, Arizona, USA (see
63 https://www.inaturalist.org/observations/156947070 for more details). A cohort of seeds from this
64 collection was germinated in May 2021. Leaf tissue collected for genome sequencing originated from a
65 single individual. Tissue samples were flash frozen and ground under liquid nitrogen immediately before
66 storage at -80C. Samples for differential expression analysis were collected from two experiments
67 conducted at UC Riverside in 2012/13 [21]. Both studies were conducted with the MVV6 line, a
68 backcrossed line originating from seeds originally collected from Moreno Valley, CA that induces
69 production of defenses in response to herbivore attack [36].

70 *Nucleic acid extractions and sequencing strategies*

71 DNA extraction, SMRT bell library preparation, and sequencing (PacBio HiFi, Pacific Bioscience,
72 San Francisco, CA, USA) were performed by the Arizona Genomics Institute (University of Arizona,
73 Tucson, AZ, USA). High molecular weight DNA was extracted from young leaves using the protocol
74 of Doyle and Doyle [22] with minor modifications. Flash-frozen young leaves were ground to a fine
75 powder in a frozen mortar with liquid nitrogen followed by very gentle extraction in 2% CTAB buffer
76 (that included proteinase K, PVP-40 and beta-mercaptoethanol) for 30min to 1h at 50 °C. After
77 centrifugation, the supernatant was gently extracted twice with 24:1 chloroform:isoamyl alcohol. The
78 upper (aqueous) phase was then removed and 1/10th volume 3 M NaAc was added, gently mixed, and
79 then had DNA precipitated with iso-propanol. DNA was collected by centrifugation, washed with 70%
80 ethanol, air dried for 20 minutes and dissolved thoroughly in elution buffer at room temperature
81 followed by RNAse treatment. DNA purity was measured with Nanodrop, DNA concentration measured
82 with Qubit HS kit (Invitrogen) and DNA size was validated by Femto Pulse System (Agilent).

83 DNA was sheared to an appropriate size range (10-20 kb) using Megaruptor 3 (Diagenode)
84 followed by Ampure bead purification. The sequencing library was constructed following manufacturers
85 protocols using SMRTbell Prep kit 3.0. The final library was size selected on a Pippin HT (Sage Science)
86 using S1 marker with a 10-25 kb size selection. The recovered final library was quantified with Qubit HS
87 kit (Invitrogen) and size checked on Femto Pulse System (Agilent). The final library was prepared for
88 sequencing with PacBio Sequel II Sequencing kit 2.0 for HiFi library, loaded on 8M SMRT cells, and
89 sequenced in CCS mode for 30 hours.

90 RNA was extracted from root and bud tissue using a ZYMO (Irvine, CA, USA) direct-zol miniprep
91 kit (Cat. # R2050) and sequenced using NovaSeq (Illumina, San Diego, CA, USA) paired-end (150bp)
92 sequencing performed by Novogene (Sacramento, CA, USA). RNA for differential expression was
93 extracted and sequenced as described in Olcerst (2017) [21].

Genome assembly/annotation

CCS output (ie: HiFi reads; 3952061 reads; 65.26Gb total; mean length = 16524) were assembled using hifiasm-0.16.0 [23] with default settings. Jellyfish v2.2.10 [31] was used for kmer counting (kmer size = 101bp) before using the GenomeScope2.0 web portal [41] to estimate genome size (Figure S1). Genome quality was examined using Bandage v0.8.1 [27], BUSCO v5.1.3 (odb10_solanales; Figure 1) [28], the blobtools v1.1 pipeline (Figure S2) [29] employing minimap v2-2.24 [30] for alignment and the nt sequence database for taxonomic identification (BLAST 2.13.0) [42], and Inspector [43]. Inspector was also used for error correction/polishing (Table S1). Repetitive elements were identified using RepeatModeler v2.0.1 and RepeatMasker v4.1.0 [24] (Table S2). Structural gene annotation was carried out using the Helixer v0.3.1 algorithm pipeline [44,45] using the pre-made land plant training dataset. Functional annotation was done using InterProScan v5.45-80.0 [46] and blastp (using blast v2.13.0) [42] comparisons to the UniProt-Swissprot database [47]. Annotation files were combined into a single gff using the manage_functional_annotation.pl script in the AGAT v 1.2.0 toolkit [48]. Detailed annotation statistics are shown in Table S3.

RNA-seq read alignments

Raw RNA-seq reads were aligned to the annotated reference genome and counted using STAR [37] with the default parameters. STAR was selected for its flexible alignment parameters (up to 10 mismatches per read, so long as total mismatches do not exceed 30% of their length), and automatic read counting functionality when given an annotation (a .gff3 file in our case). Reads aligned to our reference genome at an average rate of 80% per library, with a multi-mapping rate of less than 10% for all libraries. An average of 0.3% of reads from each library was removed for mapping to over 10 loci. Alignment details can be found in Table S4. Untransformed read counts were passed to DESeq2 [38] in R [39] for statistical analysis of differential gene expression.

Gene duplication categorization

We used two programs, MCScanX [32] and frackify [49], to investigate the history of gene duplication events that generated our observed *D. wrightii* gene content and test for signatures of ancient WGDs. Using MCScanX we made inter- and intraspecific syntenic comparisons of the *D. wrightii* and *Ipomoea purpurea* reference genomes [32]. We visualized the syntenic depth ratio of each collinear loci in the interspecific comparison using matplotlib [35]. Additionally we appended Ka/Ks values for each collinear gene pain in both the inter- and intraspecific syntenic comparison using the add_kaks_to_synteny.pl script available through MCScanX [32]. The distribution of Ks values for each comparison were transformed into density functions with the nparam_density and gaussian_kde functions from the Numpy and Scipy python libraries [59]. Maxima of each WGD and ortholog divergence peak in each density distribution were found with the find_peaks function from the Scipy python library [59]. We then used the syntenic and Ks inferences as input for Frackify to identify paleologs in the D. wrightii reference genome [49]. Finally, we classified other paralogs in the genome as tandem, dispersed, proximal, and segmental duplicates using duplicate_gene_classifier available through MCScanX [32].

Differential expression experiment 1: pairwise comparison & gene set enrichment analysis

Adults and larvae of *L. daturaphila both* feed on the leaves of *D. wrightii.* Both life stages remove leaf tissue in irregular holes between the veins (Hare & Elle 2002). The overall lifecycle of the beetle is about a month, and the duration of its larval period is about a week, depending upon temperature. Many generations of *L. daturaphila* are produced over the nine-month growing season of *D. wrightii.* Our pairwise experiment consisted of three replicates of control/ *L. daturaphila*-induced leaf samples (N = 6 total RNAseq libraries) collected from greenhouse grown plants in May 2012. Samples were collected after *L. daturaphila* larvae had been allowed to feed for 24h. Additional details regarding

140 growing conditions and sample collection can be found in Olcerst [21]. We used the Wald test in DESeq2
141 [38] to test for pairwise differences between control and *L. daturaphila*-challenged samples while
142 accounting for individual differences between sample pairs. We used a significance cutoff of P_{adj}=0.05
143 for all gene-wise analyses without any fold-change cutoff for differential expression. We then performed
144 a gene set enrichment analysis (GSEA) on the results of our pairwise experiment using the
145 ClusterProfiler 4.0 package [50]. Each GO ontology (biological processes, molecular functions, and
146 cellular component) was analyzed separately. We further separated each ontology into separate up- and
147 down-regulated gene lists as prior studies have found this approach to be more robust than grouping all
148 DEGs [51].

149 *Differential expression experiment 2: time course*

150 The quantities and composition of volatile compounds induced in *D. wrightii* by *L. daturaphila*
151 both vary with the time after herbivore damage [57]. This time course study asked how the pattern of
152 gene induction might also vary over time since induction. Our experiment consisted of samples taken at
153 5 timepoints during larval herbivory (0h [before treatment], 12h, 24h, 48h, 96h; 6 per timepoint with a
154 balanced design; except 24h where N = 5, 3 control, 2 *L. daturaphila*-induced). We used a likelihood
155 ratio test (LRT) approach to analyze these data. Our full model consisted of timepoint, treatment type,
156 and the interaction term. The reduced model lacked the interaction term; thus, this analysis tested the
157 significance of the interaction term specifically (i.e. genes for which expression over time differed by
158 treatment). We then used the DEGreport package [52] to identify co-expressed gene groups within the
159 time course dataset and visualize their expression levels over time.

160 *Duplicated gene over-representation analyses*

161 To determine the role of gene duplication categories in expanding functional gene groups, we
162 used an over-representation analysis (ORA) approach via the ClusterProfiler package in R [50]. This tests

163 for enriched GO terms within a subset of genes compared to a background set of genes. We only used
164 the molecular function ontology for this analysis. We analyzed each gene duplication category as
165 separate subsets against the total set of genes with assigned GO terms (N = 24939). We also conducted
166 this analysis using only the set of significantly differentially expressed genes with assigned GO-terms (N
167 = 1954).

168 **Results**

169 *Genome assembly and annotation*

170 Our *de novo* assembly is highly complete and contiguous (N50 = 179Mb; Longest contig =
171 202Mb; 1144 contigs; BUSCO odb10_Solanales Complete=97.7% [Single-copy=92.9%, Duplicated=4.8%],
172 Fragmented=0.2%, Missing=2.1%; Figure 1; Table 1). The total length of our assembly is 2.2Gb, larger
173 than other *Datura* assemblies [14,15] and close to the prediction obtained from GenomeScope (2.086
174 Gb; Figure S1). When compared to previous *Datura* assemblies, we find that our assembly contains far
175 fewer contigs that are orders of magnitude longer (see Table 1). This improvement is due entirely to our
176 use of PacBio HiFi reads, rather than previous generation sequencing technologies. Inspector analysis
177 found several structural errors (N=50) and small-scale errors (N = 37211; 16.63 per Mb) in our assembly.
178 Polishing our assembly did not substantially reduce the number of structural errors ($N_{polished}$ = 49), but
179 did reduce the number of small-scale errors in our assembly ($N_{polished}$ = 2808; 1.255 per Mb). Detailed
180 output from inspector, before and after polishing, are found in Table S1. Blobtools analysis determined
181 that no contamination was present in our assembly and that all contigs/reads were identified as
182 belonging to the Solanaceae (Figure S2). 86.11% of the assembly is repetitive elements, primarily
183 retroelements such as long terminal repeats (Table S2). Structural gene annotation using the helixer
184 pipeline was able to preserve the majority of gene content represented in our assembly (BUSCO
185 odb10_Solanales: Complete=96.9%, [Single-copy=91.9%, Duplicated=5.0%], Fragmented=0.6%,

186 Missing=2.5.8%; N_{genes} = 45500; Figure 1). Of the structurally annotated genes, 42745 were functionally
187 annotated; 37040 of which have inferred gene names from blastp against the Uniprot/Swissprot
188 database [47]. Detailed annotation statistics are found in Table S3.

189 *Gene Duplication Analyses*

190 We used MCScanX [32] to test for the presence of the ancient Solanaceae WGD [16,19] using a Ks
191 analysis of the colinear gene blocks within our assembly. This analysis identified a large peak (Ks = 0.70),
192 indicative of a rapid burst of gene duplications and consistent with the predictions of ancient WGD
193 (Figure 2A). We used a neighboring lineage of the Solanaceae that does not share the ancient WGD
194 (Solanales: Convolvulaceae: *Ipomoea purpurea*) to calculate the syntenic depth of loci collineary to the
195 *D. wrightii* genome assembly (Figure 2B). A significant number of collinear loci (N = 972) had a syntenic
196 depth ratio of 1:3, consistent with paleohexaploid ancestry. These inferences were further validated
197 using Frackify to identify multi-copy paleologs retained from the Ks peak at 0.70. Among the paleologs
198 retained in duplicate, 17% were retained in the triple copy state. We further used MCScanX to classify all
199 the remaining paralogs within the largest 50 contigs of the genome assembly (Table 2; Table S8). We
200 found that dispersed duplications are most common (N = 20523; Table 2), 75% of which were identified
201 as paleologs by Frackify. Ancient WGDs are likely to have generated a significant amount of the gene
202 content observed in the *D. wrightii* genome as well (N = 30714). All of these results are consistent with
203 prior analyses that the ancient Solanaceae WGD was an ancient hexaploidization [16,19].

204 *Differential expression experiment 1: pairwise comparison & gene set enrichment analysis*

205 Our pairwise analysis found that control and *L. daturaphila*-induced plants had distinct gene
206 expression profiles (Figure 3A) driven by 3555 significantly differentially expression genes (DEGs; Figure
207 3B; Table S5). Most DEGs (N = 1985) were up regulated in response to *L. daturaphila* feeding (Figure 3B).
208 Up-regulated genes had a far greater range of fold changes (Log_2FC max = 10.3) than down-regulated

209 genes (Log_2FC min = -7.3). Functional annotation identified many genes as being involved with known
210 herbivore-response processes such as jasmonic acid signaling and terpene synthesis (Table S5). Gene set
211 enrichment analysis (GSEA) further confirmed this and demonstrated that several functional groups are
212 differentially expressed in *L. daturaphila* -induced plants (Figure 4; Table S6). Terpene synthase activity
213 was notably enriched within our set of up-regulated DEGs alongside other known herbivore-response
214 functions such as peptidase inhibitors, oxidoreductase enzymes, and UDP-glycosyltransferases (Figure
215 4C). UDP-glycosyltransferase and oxidoreductase activity were also found to be significantly enriched in
216 our set of down-regulated DEGs (Figure 4D). Extracellular and apoplastic gene products were found to
217 be enriched in up-regulated DEGs (Figure 4E), consistent with previous findings in other plant-insect
218 interactions [53]. Gene-wise results of pairwise DEG analysis are found in Table S5 whereas the full GSEA
219 results and summary statistics are shown in Table S6.

220 *Differential expression experiment 2: time course*

221 Our time course analysis generated a shorter list of differentially expressed genes (N = 605;
222 Figure 5; Table S7). Principle component analysis (PCA) indicated that *L. daturaphila*-induced samples at
223 24h, 48h, and 96h formed a distinct cluster (Figure 5A). We also noticed a strong differentiation
224 between 0h (pre-herbivory) and 12h post-herbivory treatments (Figure 5A), regardless of treatment
225 type (control or *L. daturaphila*-induced), which may be a sign of background circadian rhythms in gene
226 expression [40]. Most of these genes (N_{total} = 495) clustered into nine different groups based on their
227 expression profiles over time (Figure S3), but only the two largest clusters showed a clear difference
228 between control and *L. daturaphila*-induced treatments (Figure 5B/C). Genes in these clusters were up-
229 regulated in the 24h, 48h, and 96h timepoints, consistent with our PCA results.

230 *Gene duplication over-representation analysis*

231 When assessing the functional groups enriched by different duplications, we found that tandem
232 duplications have expanded the most molecular functions (N = 45; Table 2; Table S9). Many of these
233 functions – such as chitin binding, hydrolase enzymes, peptidase inhibitors, and terpene synthases –
234 have known roles in defense against herbivores [53,54]. When we analyzed only the enriched functions
235 within our list of DEGs, we again found tandem duplications to have played the largest role by enriching
236 eight molecular function GO-terms (Figure 6; Table S10). Many of these functions were also enriched in
237 our genome-wide analysis, including acyl-transferase activity, oxidoreductase activity, and
238 endopeptidase inhibitors.

239 **Discussion**

240 Our genome assembly produced using only PacBio HiFi reads is highly complete and contiguous.
241 It is, however, important to note that it is not chromosome-scale or completely error-free. Despite great
242 improvements in genome assembly algorithms and long-read sequencing accuracy, additional datasets
243 (e.g. Hi-C) are generally required to produce chromosome-scale gapless assemblies without substantial
244 structural errors [55]. That said, our largest contigs (>100Mb in length), may represent entire
245 chromosomes and were more than sufficient to detect the presence of an ancient whole-genome
246 duplication (Figure 2A). Furthermore, most of the gene content was represented, making our assembly
247 suitable for use in mapping-based analysis such as differential expression studies.

248 Our differential expression analyses identified thousands of genes that are differentially
249 expressed in *D. wrightii* plants that have been attacked by herbivores, hundreds of which remain highly
250 expressed for multiple days. Many of the DEGs that we identified are known homologs of functionally
251 important genes in other plant-insect systems [54], although it is possible that some of these genes may
252 be induced by mechanical wounding and are not specifically involved in responses to herbivory. The lack
253 of wounding controls in our study design is a limitation, although mechanisms to impose mechanical

254 wounding that accurately mimics the small amount of tissue damaged per "bite," and accumulated
255 wounding over time remains challenging for most systems (e.g., Mithofer et al. 2005 [58]). Nonetheless,
256 our functional annotation allows us to confidently identify many DEGs as having functional roles in the
257 production of herbivore-induced plant defenses. Furthermore, the presence of genes with long-term
258 (e.g. multiple-day) changes to expression levels is likely indicative of a role in herbivore defense, as prior
259 studies have found that differential expression of wounding-response genes to be most pronounced on
260 short time scales (e.g. less than 24h) [56]. As such, our list of DEGs provides an excellent starting point
261 for future studies that will lead to further insights into the molecular basis of ecological between plants
262 and herbivorous insects over ecologically relevant time periods.

263 Gene and genome duplications have been shown to play an important role in the arms race
264 between plants and insects [20], but overall, the molecular underpinnings of plant-insect co-evolution
265 remain poorly resolved. By examining the role of various gene duplication types in expanding the
266 functional repertoire of herbivore-induced genes, we have helped to close this gap in our
267 understanding. The finding that tandem duplications are the most important is consistent with previous
268 research that has found biosynthetic genes involved with the production of secondary metabolites to
269 form distinct clusters generated via tandem duplication bursts [57]. Future studies of the duplication
270 dynamics in both plant and insects may begin to unravel the role of biotic selective pressures in
271 generating chemical defense diversity and the resources presented here will serve to accelerate this line
272 of research.

273 **Concluding Remarks**

274 In sum, we present a high quality long-read genome assembly for the sacred datura plant
275 (*Datura wrightii*). We then analyze its gene duplication history and use it as the basis for a genome-
276 guided analysis of herbivore-induced gene expression changes. Multiple tools supported the presence of

a well-documented ancient whole-genome duplication event in this species; our analysis identified thousands of differentially expressed genes, some of which have known functional annotations based on comparison to existing reference genomes. We further show, using a GO-term enrichment approach, that tandem duplications have played an important role in the evolution of *D. wrightii*'s herbivore-responsive gene repertoire. Together, these data provide a valuable resource more broadly and will contribute to future studies of angiosperm evolution and the molecular basis of ecological interactions between plants and herbivorous insects.

Declarations

Ethical Approval and consent to participate

No collections were undertaken in protected regions or of protected species, thus no permits were required for fieldwork. In lieu of voucher specimens, observations of the initial plant, alongside the users who identified it, from which seeds were collected in 2016 can be found on iNaturalist (https://www.inaturalist.org/observations/156947070). All methods were carried out in accordance with relevant institutional, national, and international guidelines and legislation.

Consent to Publication – not applicable

Data and Material Availability statement

Assembly and raw HiFi reads can be found on NCBI (BioProject: PRJNA966699; BioSample: SAMN34546691). Supporting datasets, which includes annotation, can be found at https://github.com/caterpillar-coevolution/Datura-wrightii-genome-project alongside scripts used for computational analyses. Raw transcriptome data will be available on NCBI GenBank (BioProject: PRJNA966699). Seeds from the Portal 2016 cohort can be made available on request.

Reviewer link for un-released data:
https://dataview.ncbi.nlm.nih.gov/object/PRJNA966699?reviewer=skgd96q1m91kppueafvea7va5p

Acknowledgments

We thank Professor Luciano Matzkin for his guidance throughout the process of working on the project. We would also like to thank the 2016 staff and volunteers of the Southwestern Research Station for assistance with field work and initial collection of seeds.

Funding

Funding was provided by an NSF postdoctoral research fellowship in biology to JKG (#2010772).

Conflict of Interest

The authors declare no conflict of interest.

Author contributions

JKG, AO, JDH, and MSB conceived of the project and designed the experiments. JKG collected initial samples for genome assembly. AO and JDH collected samples for differential expression analysis. JKG and MM conducted computational analyses. JKG drafted the initial manuscript. All authors reviewed and contributed to the final version of this manuscript.

315 **Literature Cited**

316 [1] Hare, J.D., Walling, L.L., 2006. Constitutive and Jasmonate-Inducible Traits of Datura wrightii. J Chem
317 Ecol 32, 29–47. https://doi.org/10.1007/s10886-006-9349-8

318 [2] Goldberg, J.K., Lively, C.M., Sternlieb, S.R., Pintel, G., Hare, J.D., Morrissey, M.B., Delph, L.F., 2020.
319 Herbivore-mediated negative frequency-dependent selection underlies a trichome dimorphism in
320 nature. Evolution Letters 4, 83–90. https://doi.org/10.1002/evl3.157

321 [3] Bronstein, J.L., Huxman, T., Horvath, B., Farabee, M., Davidowitz, G., 2009. Reproductive biology of
322 Datura wrightii: the benefits of a herbivorous pollinator. Annals of Botany 103, 1435–1443.
323 https://doi.org/10.1093/aob/mcp053

324 [4] Riffell, J.A., Alarcón, R., Abrell, L., 2008. Floral trait associations in hawkmoth-specialized and mixed
325 pollination systems. Communicative & Integrative Biology 1, 6–8. https://doi.org/10.4161/cib.1.1.6350

326 [5] Elle, E., Hare, J.D., 2002. Environmentally Induced Variation in Floral Traits Affects the Mating System
327 in Datura wrightii. Functional Ecology 16, 79–88.

328 [6] Dupin, J., Smith, S.D., 2019. Corrigendum to: "Phylogenetics of Datureae (Solanaceae), including
329 description of the new genus Trompettia and re-circumscription of the tribe" [in Taxon 67: 359–375.
330 2018]. TAXON 68, 419–419. https://doi.org/10.1002/tax.12056

331 [7] Dupin, J., Smith, S.D., 2018. Phylogenetics of Datureae (Solanaceae), including description of the new
332 genus Trompettia and re–circumscription of the tribe. TAXON 67, 359–375.
333 https://doi.org/10.12705/672.6

334 [8] van Dam, N., Hare, J., Elle, E., 1999. Inheritance and distribution of trichome phenotypes in Datura
335 wrightii. Journal of Heredity 90, 220–227. https://doi.org/10.1093/jhered/90.1.220

336 [9] Kariñho-Betancourt, E., Agrawal, A.A., Halitschke, R., Núñez-Farfán, J., 2015. Phylogenetic
337 correlations among chemical and physical plant defenses change with ontogeny. New Phytologist 206,
338 796–806. https://doi.org/10.1111/nph.13300

339 [10] Goldberg, J.K., Sternlieb, S.R., Pintel, G., Delph, L.F., 2021. Observational evidence of herbivore-
340 specific associational effects between neighboring conspecifics in natural, dimorphic populations of
341 Datura wrightii. Ecology and Evolution 11, 5547–5561. https://doi.org/10.1002/ece3.7454

342 [11] Zhang, J., Komail Raza, S.A., Wei, Z., Keesey, I.W., Parker, A.L., Feistel, F., Chen, J., Cassau, S.,
343 Fandino, R.A., Grosse-Wilde, E., Dong, S., Kingsolver, J., Gershenzon, J., Knaden, M., Hansson, B.S., 2022.
344 Competing beetles attract egg laying in a hawkmoth. Current Biology 32, 861-869.e8.
345 https://doi.org/10.1016/j.cub.2021.12.021

346 [12] Hare, J.D., Elle, E., van Dam, N.M., 2003. Costs of Glandular Trichomes in Datura Wrightii: A Three-
347 Year Study. Evolution 57, 793–805. https://doi.org/10.1111/j.0014-3820.2003.tb00291.x

348 [13] Lang, D., Zhang, S., Ren, P., Liang, F., Sun, Z., Meng, G., Tan, Y., Li, X., Lai, Q., Han, L., Wang, D., Hu,
349 F., Wang, W., Liu, S., 2020. Comparison of the two up-to-date sequencing technologies for genome
350 assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore.
351 GigaScience 9, giaa123. https://doi.org/10.1093/gigascience/giaa123

352 [14] Rajewski, A., Carter-House, D., Stajich, J., Litt, A., 2021. Datura genome reveals duplications of
353 psychoactive alkaloid biosynthetic genes and high mutation rate following tissue culture. BMC Genomics
354 22, 201. https://doi.org/10.1186/s12864-021-07489-2

355 [15] De-la-Cruz, I.M., Hallab, A., Olivares-Pinto, U., Tapia-López, R., Velázquez-Márquez, S., Piñero, D.,
356 Oyama, K., Usadel, B., Núñez-Farfán, J., 2021. Genomic signatures of the evolution of defence against its

357 natural enemies in the poisonous and medicinal plant Datura stramonium (Solanaceae). Sci Rep 11, 882.
358 https://doi.org/10.1038/s41598-020-79194-1

359 [16] Xu, S., Brockmöller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., Zhou, W., Kreitzer, C.,
360 Stanke, M., Tang, H., Lyons, E., Pandey, P., Pandey, S.P., Timmermann, B., Gaquerel, E., Baldwin, I.T.,
361 2017. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proceedings of the National
362 Academy of Sciences 114, 6133–6138. https://doi.org/10.1073/pnas.1700073114

363 [17] Schranz, E. M., Mohammadin, S., Edger, P.P., 2012. Ancient whole genome duplications, novelty
364 and diversification: the WGD Radiation Lag-Time Model. Current Opinion in Plant Biology, Genome
365 studies molecular genetics 15, 147–153. https://doi.org/10.1016/j.pbi.2012.03.011

366 [18] Selmecki, A.M., Maruvka, Y.E., Richmond, P.A., Guillet, M., Shoresh, N., Sorenson, A.L., De, S.,
367 Kishony, R., Michor, F., Dowell, R., Pellman, D., 2015. Polyploidy can drive rapid adaptation in yeast.
368 Nature 519, 349–352. https://doi.org/10.1038/nature14187

369 [19] Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham,
370 S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M., Rensing, S.A., Soltis, D.E., Soltis,
371 P.S., Stevenson, D.W., Ullrich, K.K., Wickett, N.J., DeGironimo, L., Edger, P.P., Jordon-Thaden, I.E., Joya,
372 S., Liu, T., Melkonian, B., Miles, N.W., Pokorny, L., Quigley, C., Thomas, P., Villarreal, J.C., Augustin, M.M.,
373 Barrett, M.D., Baucom, R.S., Beerling, D.J., Benstein, R.M., Biffin, E., Brockington, S.F., Burge, D.O.,
374 Burris, J.N., Burris, K.P., Burtet-Sarramegna, V., Caicedo, A.L., Cannon, S.B., Çebi, Z., Chang, Y., Chater, C.,
375 Cheeseman, J.M., Chen, T., Clarke, N.D., Clayton, H., Covshoff, S., Crandall-Stotler, B.J., Cross, H.,
376 dePamphilis, C.W., Der, J.P., Determann, R., Dickson, R.C., Di Stilio, V.S., Ellis, S., Fast, E., Feja, N., Field,
377 K.J., Filatov, D.A., Finnegan, P.M., Floyd, S.K., Fogliani, B., García, N., Gâteblé, G., Godden, G.T., Goh, F.
378 (Qi Y., Greiner, S., Harkess, A., Heaney, J.M., Helliwell, K.E., Heyduk, K., Hibberd, J.M., Hodel, R.G.J.,
379 Hollingsworth, P.M., Johnson, M.T.J., Jost, R., Joyce, B., Kapralov, M.V., Kazamia, E., Kellogg, E.A., Koch,

380 M.A., Von Konrat, M., Könyves, K., Kutchan, T.M., Lam, V., Larsson, A., Leitch, A.R., Lentz, R., Li, F.-W.,
381 Lowe, A.J., Ludwig, M., Manos, P.S., Mavrodiev, E., McCormick, M.K., McKain, M., McLellan, T., McNeal,
382 J.R., Miller, R.E., Nelson, M.N., Peng, Y., Ralph, P., Real, D., Riggins, C.W., Ruhsam, M., Sage, R.F., Sakai,
383 A.K., Scascitella, M., Schilling, E.E., Schlösser, E.-M., Sederoff, H., Servick, S., Sessa, E.B., Shaw, A.J., Shaw,
384 S.W., Sigel, E.M., Skema, C., Smith, A.G., Smithson, A., Stewart, C.N., Stinchcombe, J.R., Szövényi, P.,
385 Tate, J.A., Tiebel, H., Trapnell, D., Villegente, M., Wang, C.-N., Weller, S.G., Wenzel, M., Weststrand, S.,
386 Westwood, J.H., Whigham, D.F., Wu, S., Wulff, A.S., Yang, Y., Zhu, D., Zhuang, C., Zuidof, J., Chase, M.W.,
387 Pires, J.C., Rothfels, C.J., Yu, J., Chen, C., Chen, L., Cheng, S., Li, J., Li, R., Li, X., Lu, H., Ou, Y., Sun, X., Tan,
388 X., Tang, J., Tian, Z., Wang, F., Wang, J., Wei, X., Xu, X., Yan, Z., Yang, F., Zhong, X., Zhou, F., Zhu, Y.,
389 Zhang, Y., Ayyampalayam, S., Barkman, T.J., Nguyen, N., Matasci, N., Nelson, D.R., Sayyari, E., Wafula,
390 E.K., Walls, R.L., Warnow, T., An, H., Arrigo, N., Baniaga, A.E., Galuska, S., Jorgensen, S.A., Kidder, T.I.,
391 Kong, H., Lu-Irving, P., Marx, H.E., Qi, X., Reardon, C.R., Sutherland, B.L., Tiley, G.P., Welles, S.R., Yu, R.,
392 Zhan, S., Gramzow, L., Theißen, G., Wong, G.K.-S., One Thousand Plant Transcriptomes Initiative, 2019.
393 One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574, 679–685.
394 https://doi.org/10.1038/s41586-019-1693-2

395 [20] Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glöckner, G., Platts, A.E., Heckel, D.G., Der,
396 J.P., Wafula, E.K., Tang, M., Hofberger, J.A., Smithson, A., Hall, J.C., Blanchette, M., Bureau, T.E., Wright,
397 S.I., dePamphilis, C.W., Eric Schranz, M., Barker, M.S., Conant, G.C., Wahlberg, N., Vogel, H., Pires, J.C.,
398 Wheat, C.W., 2015. The butterfly plant arms-race escalated by gene and genome duplications.
399 Proceedings of the National Academy of Sciences 112, 8362–8366.
400 https://doi.org/10.1073/pnas.1503926112

401 [21] Olcerst, A. 2017. Variation in Induced Responses of Datura wrightii to Herbivore Attack: Plasticity of
402 Volatile Organic Compound Emissions and Gene Expression Across Genotypes, Ontogeny, and a Single
403 Attack – Doctoral Disseration accessed via ProQuest:

404 https://www.proquest.com/openview/920bf36262446dfa983f0a67e1755bff/1?pq-
405 origsite=gscholar&cbl=18750

406 [22] Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf
407 tissues. Phytochem Bull. 19:11–15.

408 [23] Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021. Haplotype-resolved de novo assembly
409 using phased assembly graphs with hifiasm. Nat Methods 18, 170–175. https://doi.org/10.1038/s41592-
410 020-01056-5

411 [24] Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to Identify Repetitive Elements in
412 Genomic Sequences. Current Protocols in Bioinformatics 25, 4.10.1-4.10.14.
413 https://doi.org/10.1002/0471250953.bi0410s25

414 [25] Palmer, J., Stajich, J., 2019. nextgenusfs/funannotate: funannotate v1.5.3.
415 https://doi.org/10.5281/zenodo.2604804

416 [26] Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell,
417 A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R.,
418 Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–
419 1240. https://doi.org/10.1093/bioinformatics/btu031

420 [27] Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E., 2015. Bandage: interactive visualization of de novo
421 genome assemblies. Bioinformatics 31, 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

422 [28] Seppey, M., Manni, M., Zdobnov, E.M., 2019. BUSCO: Assessing Genome Assembly and Annotation
423 Completeness, in: Kollmar, M. (Ed.), Gene Prediction: Methods and Protocols, Methods in Molecular
424 Biology. Springer, New York, NY, pp. 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

425 [29] Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies.
426 https://doi.org/10.12688/f1000research.12232.1

427 [30] Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–
428 3100. https://doi.org/10.1093/bioinformatics/bty191

429 [31] Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of
430 occurrences of k-mers. Bioinformatics 27, 764–770. https://doi.org/10.1093/bioinformatics/btr011

431 [32] Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H.,
432 Kissinger, J.C., Paterson, A.H., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene
433 synteny and collinearity. Nucleic Acids Research 40, e49. https://doi.org/10.1093/nar/gkr1293

434 [33] Hosmani, P. S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., … Saha,
435 S., 2019. An improved de novo assembly and annotation of the tomato reference genome using single-
436 molecule sequencing, Hi-C proximity ligation and optical maps. Cold Spring Harbor Laboratory.
437 https://doi.org/10.1101/767764

438 [34] The Potato Genome Sequencing Consortium, 2011. Genome sequence and analysis of the tuber
439 crop potato. *Nature* 475, 189–195. https://doi.org/10.1038/nature10158

440 [35] Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9,
441 90–95. https://doi.org/10.1109/MCSE.2007.55

442 [36] Hare, J.D., 2007. Variation in Herbivore and Methyl Jasmonate-Induced Volatiles Among Genetic
443 Lines of Datura wrightii. J Chem Ecol 33, 2028–2043. https://doi.org/10.1007/s10886-007-9375-1

444 [37] Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,
445 Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.
446 https://doi.org/10.1093/bioinformatics/bts635

447 [38] Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for
448 RNA-seq data with DESeq2. Genome Biol 15, 1–21. https://doi.org/10.1186/s13059-014-0550-8

449 [39] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for
450 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

451 [40] Xu, X., Yuan, L., Xie, Q., 2022. The circadian clock ticks in plant stress responses. Stress Biology 2, 15.
452 https://doi.org/10.1007/s44154-022-00040-7

453 [41] Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for
454 reference-free profiling of polyploid genomes. Nat Commun 11, 1432. https://doi.org/10.1038/s41467-
455 020-14998-3

456 [42] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009.
457 BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-
458 10-421

459 [43] Chen, Y., Zhang, Y., Wang, A.Y., Gao, M., Chong, Z., 2021. Accurate long-read de novo assembly
460 evaluation with Inspector. Genome Biol 22, 1–21. https://doi.org/10.1186/s13059-021-02527-4

461 [44] Holst, F., Bolger, A., Günther, C., Maß, J., Triesch, S., Kindel, F., Kiel, N., Saadat, N., Ebenhöh, O.,
462 Usadel, B., Schwacke, R., Bolger, M., Weber, A.P.M., Denton, A.K., 2023. Helixer–de novo Prediction of
463 Primary Eukaryotic Gene Models Combining Deep Learning and a Hidden Markov Model.
464 https://doi.org/10.1101/2023.02.06.527280

465 [45] Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A.P.M., Denton, A.K., 2021. Helixer: cross-
466 species gene annotation of large eukaryotic genomes using deep learning. Bioinformatics 36, 5291–
467 5298. https://doi.org/10.1093/bioinformatics/btaa1044

468 [46] Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell,
469 A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R.,
470 Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–
471 1240. https://doi.org/10.1093/bioinformatics/btu031

472 [47] Schneider, M., Lane, L., Boutet, E., Lieberherr, D., Tognolli, M., Bougueleret, L., Bairoch, A., 2009.
473 The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. Journal of
474 Proteomics, Plant Proteomics 72, 567–573. https://doi.org/10.1016/j.jprot.2008.11.010

475 [48] Dainat, J., Hereñú, D., Murray, D.K.D., Davis, E., Crouch, K., LucileSol, Agostinho, N., pascal-git,
476 Zollman, Z., tayyrov, 2023. NBISweden/AGAT: AGAT-v1.2.0. https://doi.org/10.5281/zenodo.8178877

477 [49] McKibben, M.T.W., Barker, M.S., 2023. Applying Machine Learning to Classify the Origins of Gene
478 Duplications, in: Van de Peer, Y. (Ed.), Polyploidy: Methods and Protocols, Methods in Molecular Biology.
479 Springer US, New York, NY, pp. 91–119. https://doi.org/10.1007/978-1-0716-2561-3_5

480 [50] Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S.,
481 Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The
482 Innovation 2, 100141. https://doi.org/10.1016/j.xinn.2021.100141

483 [51] Hong, G., Zhang, W., Li, H., Shen, X., Guo, Z., 2014. Separate enrichment analysis of pathways for
484 up- and downregulated genes. Journal of The Royal Society Interface 11, 20130950.
485 https://doi.org/10.1098/rsif.2013.0950

486 [52] Pantano, L. 2017. Report of DEG analysis URL https://lpantano.github.io/DEGreport/

487 [53] Kerchev, P.I., Fenton, B., Foyer, C.H., Hancock, R.D., 2012. Plant responses to insect herbivory:
488 interactions between photosynthesis, reactive oxygen species and hormonal signalling pathways. Plant,
489 Cell & Environment 35, 441–453. https://doi.org/10.1111/j.1365-3040.2011.02399.x

490 [54] Howe, G.A., Jander, G., 2008. Plant Immunity to Insect Herbivores. Annual Review of Plant Biology
491 59, 41–66. https://doi.org/10.1146/annurev.arplant.59.032607.092825 [55] chitin binding, hydrolase
492 enzymes, peptidase inhibitors, and terpene synthases

493 [55] Cheng, H., Jarvis, E.D., Fedrigo, O., Koepfli, K.-P., Urban, L., Gemmell, N.J., Li, H., 2022. Haplotype-
494 resolved assembly of diploid genomes without parental data. Nat Biotechnol 40, 1332–1335.
495 https://doi.org/10.1038/s41587-022-01261-x

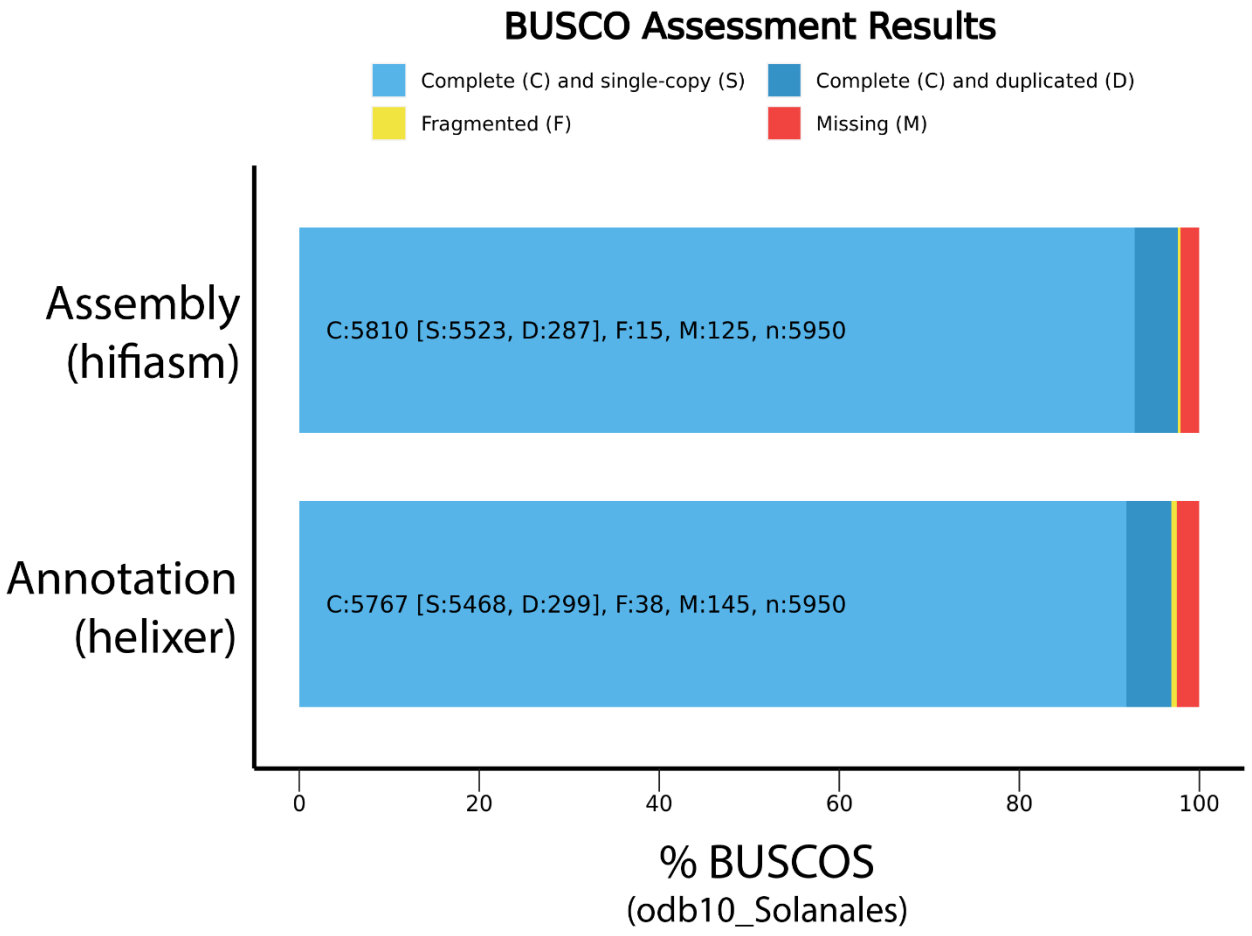
496 [56] Xu, Z., Pu, X., Gao, R., Demurtas, O.C., Fleck, S.J., Richter, M., He, C., Ji, A., Sun, W., Kong, J., Hu, K.,
497 Ren, F., Song, Jiejie, Wang, Z., Gao, T., Xiong, C., Yu, H., Xin, T., Albert, V.A., Giuliano, G., Chen, S., Song,
498 Jingyuan, 2020. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic
499 pathways in plants. BMC Biol 18, 1–14. https://doi.org/10.1186/s12915-020-00795-3

500 [57] Hare, J.D., Sun, J.J., 2011. Production of Induced Volatiles by Datura wrightii in Response to Damage
501 by Insects: Effect of Herbivore Species and Time. J Chem Ecol 37, 751–764.
502 https://doi.org/10.1007/s10886-011-9985-5

503 [58] Mithöfer, A., Wanner, G., Boland, W., 2005. Effects of feeding Spodoptera littoralis on lima bean
504 leaves. II. Continuous mechanical wounding resembling insect feeding is sufficient to elicit herbivory-
505 related volatile emission. Plant Physiol 137, 1160–1168. https://doi.org/10.1104/pp.104.054460

506 [59] Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau,
507 Evgeni Burovski, et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.”
508 Nature Methods 17 (3): 261–72

509 **Figure Legends & Tables**



510

511 **Figure 1.** Bar charts showing the gene content represented in both our assembly (top bar) and our

512 annotation (bottom bar). Annotation assessment was run in proteome mode.

513

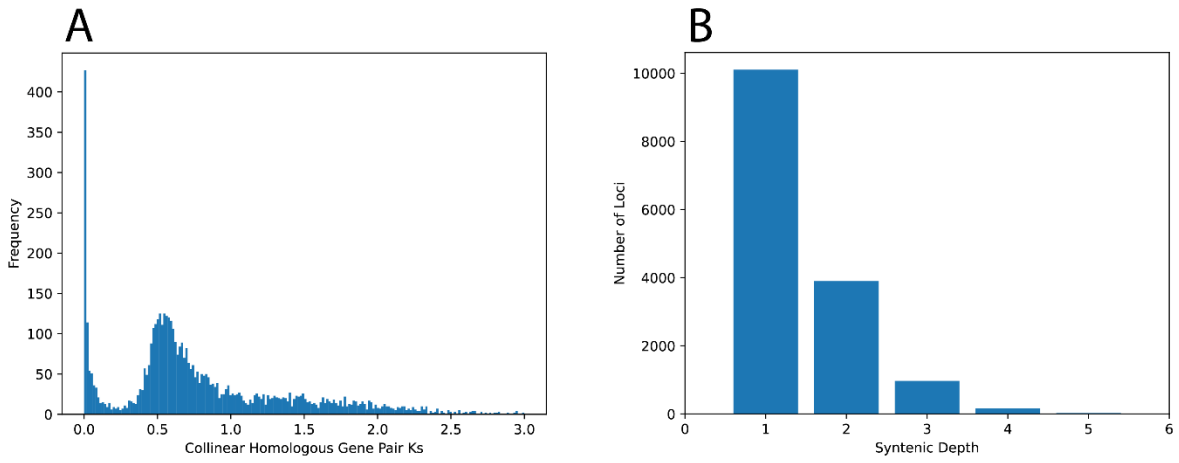


Figure 2. (A) Results of MCSxanX showing the frequency of non-synonymous substitutions in homologous syntenic gene blocks. The large peak at ~0.5Ks indicates a large burst of gene duplications, consistent with the presence of an ancient whole-genome duplication. (B) Frackify showing the syntenic depth of gene blocks in the *Datura wrightii* genome compared to *Ipopmoea purpurea*. The elevated number of triplicated syntenic blocks indicates an ancient hexaploidy state in *D. wrightii* that is not present in *I. purpurea*.

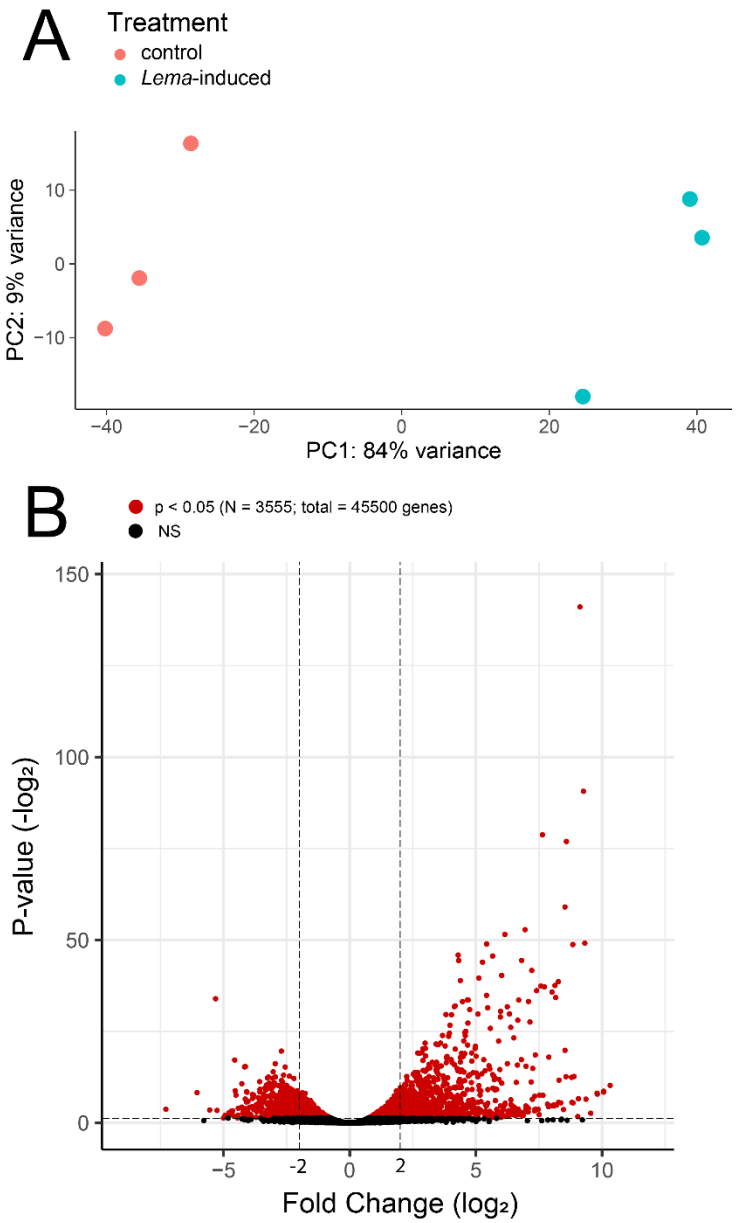
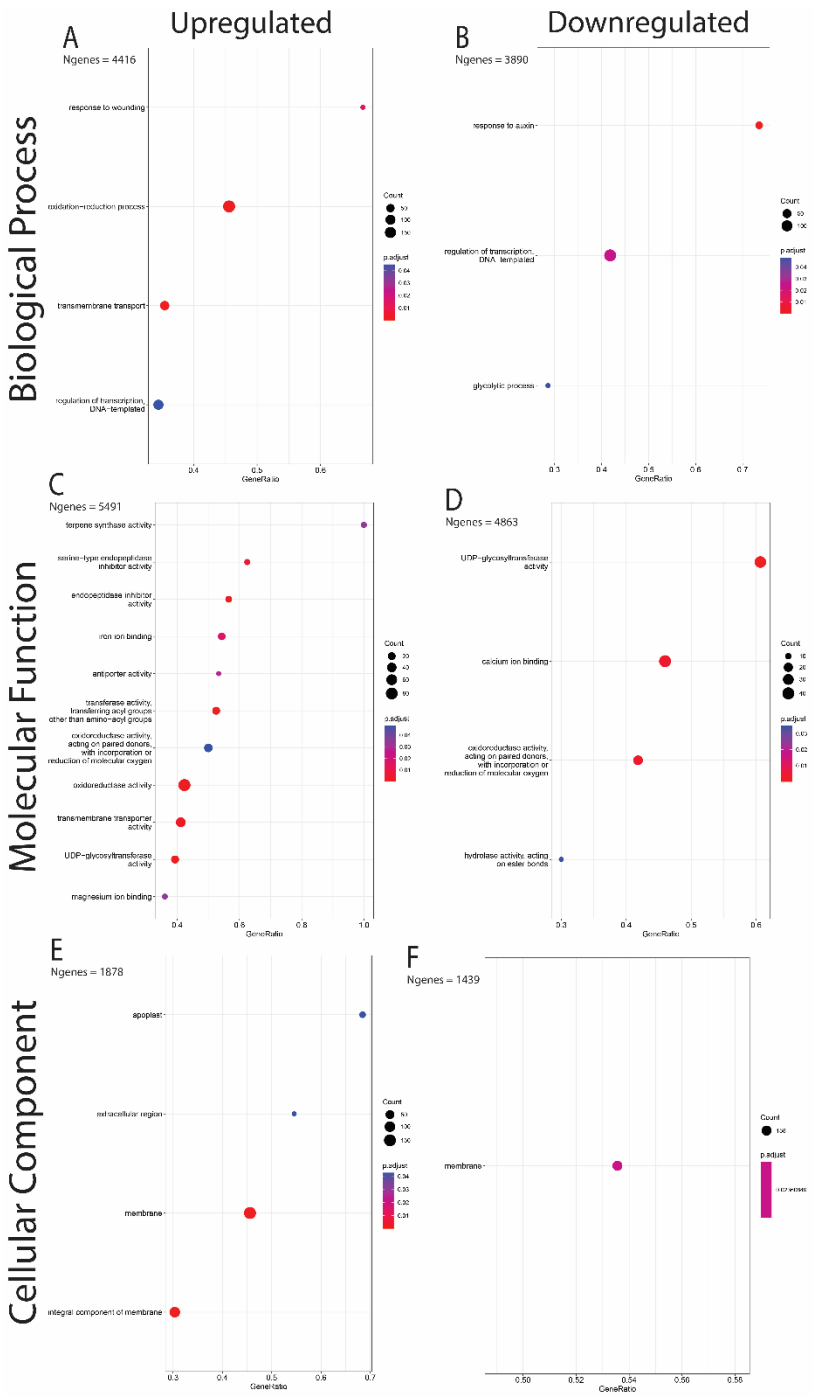


Figure 3. Results of pairwise differential expression analysis using only the samples collected in May 2012. (A) Results of exploratory PCA to qualitatively screen for differences between treatment groups. There was a clear qualitative difference between the control and *L. daturaphila*-induced samples driven by PC1. (B) Volcano plot showing the relationship between significance and fold-change for each gene in our *D. wrightii* genome annotation. A change of 2/-2 is marked for reference but was not used as a cutoff for any analysis. Significant genes ($p < 0.05$) are shown in red, non-significant genes in black.



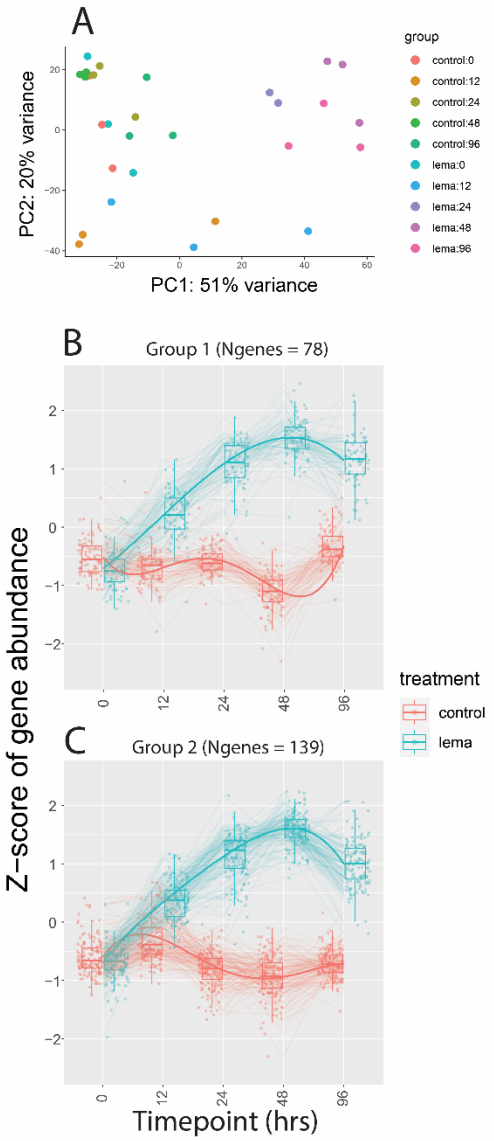
529

530 **Figure 4.** Results of gene set enrichment analysis of pairwise DEGs. X-axes show the ratio of enriched

531 genes vs the total count of genes sharing that GO term. Dot sizes represent the total number of genes

532 sharing each GO term whereas dot colors represent the p-value (adjusted for multiple tests) of each

533 term. Each GO ontology was analyzed separately for up- and down-regulated genes.



534

535 **Figure 5.** (A) Exploratory PCA of time course data. Although our samples did not form clear clusters,

536 there are definitive patterns. The 24h, 48h, and 96h *L. daturaphila*-induced samples cluster in the top

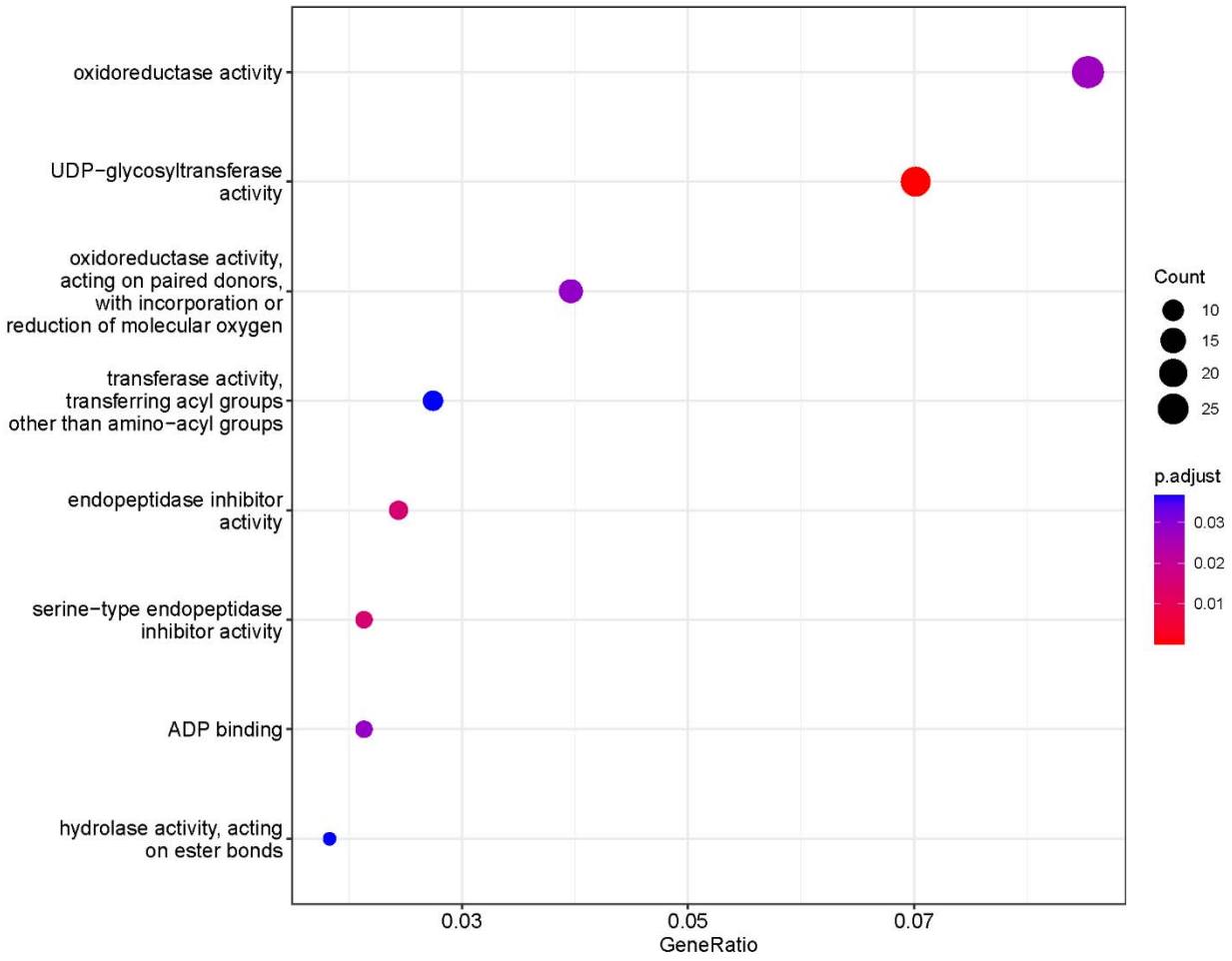
537 right, whereas the controls for these time points cluster with the 0h samples in the top left. Ranges for

538 0h and 12h samples do not overlap at all, suggesting that there may also be circadian cycles occurring.

539 (B/C) Cluster plots showing the two largest gene clusters with similar expression patterns. These are the

540 only gene clusters that showed a clear separation between treatments throughout the experiment time.

541 Extended gene cluster results are shown in Figure S3.



542

543 **Figure 6.** Results of over-representation analysis examining the molecular function GO-terms of

544 differentially expressed genes enriched by tandem duplications. Eight terms were found to be enriched

545 in our dataset, more than any other duplication type (see Table 2 for details).

546

547 **Table 1.** Summary of assembly statistics compared to those of *Datura stramonium* genomes previously
548 published by De-La-Cruz et al. (2021) [15].

	***D. wrightii* (Portal, AZ, USA)**	***D. stramonium* (Ticumán, MX)**	***D. stramonium* (Teotihuacán, MX)**
Total Size (Gb)	2.2	1.48	1.28
Contig number	1144	27915	30392
Largest contig (Mb)	202	3.13	2.11
N50	179Mb	84.1kb	58.2kb
N75	121Mb	44.5kb	32.6kb
L50	5	4557	5713
L75	10	10641	13166
BUSCO complete (%)	97.70	91	81.70
BUSCO db (odb10)	Solanales	Solanaceae	Solanaceae

549

550 **Table 2.** Results of MCScanX gene duplication classifier function and GO-term over-representation
551 analysis by duplication type. Gene counts refer to the total number of genes (ie: with or without
552 assigned GO-terms).

Duplication type	**Gene count**		**Enriched GO-terms (Molecular function)**	
	Whole genome	**DEGs only**	**Whole Genome**	**DEGs only**
Singleton	4268	431	NA	NA
Dispersed	20523	1329	10	1
Proximal	2140	154	28	2
Tandem	4064	487	45	8
Segmental/WGD	14505	1196	13	1

553