Title: A *de novo* long-read genome assembly is contiguous and complete enough to study paleo-polyploidy and herbivore-induced transcriptional responses of the sacred datura plant (*Datura wrightii*)

Authors: Jay K Goldberg[1]*, Aaron Olcerst[2], Michael McKibben[1], J Daniel Hare[2], 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 (*Datura wrightii*) has been used to study plant-herbivore interactions for decades and, due to this wealth of literature regarding it's ecological interactions, has potential as a model system for future research on the ecological and evolutionary genomics of plant/insect interactions. In this paper, we present a *de novo Datura wrightii* genome that we assembled using PacBio HiFi long-reads. Our assembly is highly complete and contiguous (N50 = 179Mb, BUSCO Complete = 97.7%). To demonstrate the utility of our assembly to future studies, we successfully used multiple programs to test for the presence of an ancient whole genome duplication and classify the duplicated genes identified within our assembly. We also used this assembly as the basis for a genome-guided differential expression analysis to identify the induced responses of this plant to one of its specialist herbivores (Coleoptera: Chrysomelidae: *Lema daturaphila*). We find over 6000 significantly differentially expressed genes between control and herbivore-induced plants, and further that these changes appear to be sustained and last multiple days. Overall, our results demonstrate that HiFi-only draft plant genomes are of sufficient quality to serve as the basis for varied biological insights and expand our knowledge of herbivore-induced plant transcriptional responses.

Introduction

Datura wrightii is a fascinating plant species with great potential as a model system for research into the ecology and evolution of various plant traits including both physical and chemical defenses (Hare and Walling 2006; Goldberg et al 2020), tolerance to herbivory (Bronstein et al 2009), floral phenotypes (Elle and Hare 2002; Riffell et al 2008), perennial life history (Dupin and Smith 2018; 2019) and ontogenetic changes in defense production throughout an individual plant's lifetime (van Dam et al 1999; Kariñho-Betancourt et al 2015). This is due to the wealth of ecological knowledge already gathered regarding this common plant and some of its specialist herbivores (Goldberg et al. 2021; Zhang et al 2022). Furthermore, field-based methodologies have already provided some insights into the evolutionary processes maintaining a trichome dimorphism in naturally occurring populations (Hare et al 2007; Goldberg et al 2020). 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.

In this paper, we present a *de novo* genome assembly for the sacred datura plant (*Datura wrightii*) generated using highly accurate long-reads (PacBio HiFi; lang et al 2020). As a result, our assembly is highly complete and contiguous when compared to recent assemblies of closely related species (Rajewski et al 2021; De-la-Cruz et al 2021; Xu et al 2017) 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. Together, they show that our *de novo* assembly will have prolonged value to studies of molecular mechanisms of plant responses to stress and comparative analyses of flowering plant evolution.

Gene and whole genome duplications (WGDs) have broad implications for evolutionary processes including the generation and adaptation of novel traits (Schranz et al 2012; Selmecki et al 2015). WGDs have occurred numerous times throughout the diversification of angiosperms (Landis et al 2018); and these 'ancient' WGDs have been shown to underlie the arms-race dynamic occurring between plants and herbivores (Edger et al 2015). There is widespread evidence of an ancient hexploidy that occurred early in the evolution of the nightshade family (Solanaceae; Xu et al 2017; Landis et al 2018). As such we used two different analyses – one of which uses only contiguous collinear gene blocks – to test for evidence of this ancient WGD in our genome.

We further use our assembly to conduct a differential expression study into the induced responses of *D. wrightii* to one of its specialist herbivores, the three-lined potato beetle (*Lema daturaphila*; Goldberg et al. 2021; Zhang et al 2022). This analysis utilized older data, generated roughly a decade ago (Olcerst 2017). Using both pairwise and time course analyses, we identified 1000s of genes that are significantly differentially expressed when plants are under attack by one of their closest natural enemies and further show that many of them remain upregulated for multiple days.

Methods

Initial sample collection and preparation, and sequencing

Datura wrightii seeds were originally collected in June 2016 from a wild plant growing at the intersection of Portal Rd. and Foothills Rd. in Portal, Arizona (see

64 https://www.inaturalist.org/observations/156947070 for more details). A cohort of seeds from this
65 collection were germinated in May 2021. Leaf tissue collected for genome sequencing originated from a
66 single individual, whereas the bud and root tissue for annotation were collected from other individuals
67 in this cohort. All samples were flash frozen and ground under liquid nitrogen immediately before
68 storage at -80C.

69 Nucleic acid extractions and sequencing strategies

70 DNA extraction and SMRT bell library preparation, and sequencing (PacBio HiFi, Pacific
71 Bioscience, San Francisco, CA, USA) were performed by the Arizona Genomics Institute (University of
72 Arizona, Tucson, AZ, USA). High molecular weight DNA was extracted from young leaves using the
73 protocol of Doyle and Doyle (1987) with minor modifications. Flash-frozen young leaves were ground to
74 a fine powder in a frozen mortar with liquid nitrogen followed by very gentle extraction in 2% CTAB
75 buffer (that included proteinase K, PVP-40 and beta-mercaptoethanol) for 30min to 1h at 50 °C. After
76 centrifugation, the supernatant was gently extracted twice with 24:1 chloroform:isoamyl alcohol. Take
77 out the upper phase and add 1/10th volume 3 M NaAc, gently mixed, and DNA precipitated with iso-
78 propanol. DNA was collected by centrifugation, washed with 70% ethanol, air dried for 20 minutes and
79 dissolved thoroughly in EB at room temperature followed by RNAse treatment. DNA purity was
80 measured with Nanodrop, DNA concentration measured with Qubit HS kit (Invitrogen) and DNA size was
81 validated by Femto Pulse System (Agilent).

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

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

93 *Genome assembly/annotation and gene duplication analyses*

94 CCS output (ie: HiFi reads; 3952061 reads; 65.26Gb total; mean length = 16524) were assembled
95 using hifiasm-0.16.0 (Cheng et al. 2021) with default settings. Repetitive elements were identified using
96 RepeatModeler v2.0.1 and softmasked using RepeatMasker v4.1.0 (Tarailo-Graovac and Chen 2009).
97 Structural annotation was carried out using the funannotate pipeline v1.8.14 (Palmer and Stajich 2019)
98 using all mRNA-seq libraries (N=2, 150bp paired-end; N=36 50bp single-end) as evidence during the
99 'funannotate train' step. Protein sequences from a publicly available *D. wrightii* transcriptome
100 (downloaded from medplantrnaseq.com) were used as evidence in the 'funanotate predict' step.
101 Default settings were used, except for max intron length which was set to 10000bp. Functional
102 annotation was performed in InterProScan v5.57-90.0 (Jones et al 2014). Genome quality was examined
103 using Bandage v0.8.1 (Wick et al 2015), BUSCO v5.1.3 (Figure 1; Seppey et al 2019), and the blobtools
104 v1.1 pipeline (Figure 2; Laetsch and Blaxter 2017) employing minimap v2-2.24 (Li 2018) for alignment

and the nt sequence database for taxonomic identification (BLAST 2.13.0). We used two programs, DupPipe (Barker et al 2010) and McScanX (Wang et al 2012), to investigate the history of gene duplication events that generated our observed *D. wrightii* gene content and test for signatures of ancient WGDs. We used our *D. wrightii* transcriptome in addition to transcriptomes from *Solanum lycopersicum* (Hosmani et al 2019), and *S. tuberosum* (The Potato Genome Consortium 2011) as a reference protein database for DupPipe. MCscanX was run on default settings following the removal of isoforms and Ka/Ks values were calculated using the add_kaks_to_synteny.pl script (Wang et al 2012). Synonymous (Ks) divergence calculations for the DupPipe and MCscanX were then visualized using matplotlib (Hunter 2007).

Differential expression analysis

Samples for differential expression analysis were collected from two experiments conducted at UC Riverside in 2012/13 (Olcerst 2017). Both studies were conducted with the MVV6 line, a backcrossed line originating from seeds originally collected in southern California and which induces production of defenses in response to herbivore attack (Hare, 2007). The first consisted of three replicates of control/*Lema*-induced leaf samples (N = 6 total RNAseq libraries) collected from greenhouse grown plants in May 2012. The second was a time course experiment in which samples were taken at 5 timepoints following herbivory (0h [before treatment], 12h, 24h, 48h, 96h; 6 per timepoint, except 24h which was excluded due to lack of replication). Additional details regarding growing conditions and sample collection can be found in Olcerst (2017).

Raw RNAseq reads were aligned to the annotated reference genome and counted using STAR (Dobin et al 2013) 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 S3. Untransformed read counts were passed to DESeq2 (Love et al 2014) in R (R Core Team 2022) for statistical analysis of differential gene expression. Pairwise analysis was conducted using only the samples collected in May 2012 (N = 6; Figure 3). Time course dataset (N = 24) was also analyzed in DESeq2 using a likelihood ratio test (LRT) approach. Our full model consisted of an ANCOVA design, with treatment, timepoint, and their interaction term as variables. Our reduced model lacked the interaction term; thus, this analysis tested the significance of the interaction term specifically (i.e. genes for which expression over time differed by treatment). Due to a lack of replication for the 24h timepoint, it was excluded from the time course analysis. Data visualization was performed in the DEseq2 (PCA plots), EnhancedVolcano, and pheatmap packages in R (R Core Team 2022).

Results and Discussion

Genome assembly and annotation

Our *de novo* assembly is highly complete and contiguous (N50 = 179Mb; Longest contig = 202Mb; 1144 contigs; BUSCO Complete=97.7% [Single-copy=92.9%, Duplicated=4.8%], Fragmented=0.2%, Missing=2.1%; Figure 1). The total length of our assembly is 2.2Gb, comparable to other *Datura* assemblies (Rajewski et al 2021; De-la-Cruz et al 2021) and making our final coverage

27.7X across the whole genome (18.1X when counting only reads that mapped back to the assembly; Figure 1). When compared to previous *Datura* assemblies we find that our assembly contains far fewer contigs that are orders of magnitude longer. This improvement is due entirely to our use of PacBio HiFi reads, rather than previous generation technologies. Blobtools analysis determined that no contamination was present in our assembly and that all contigs/reads were identified as belonging to the Solanaceae family (Figure 1). 86.11% of the assembly is repetitive elements, primarily retroelements such as long terminal repeats (Table S1).

Our annotation, which incorporated both mRNA and protein evidence, identified 49439 genes, but BUSCO analysis showed that it was less complete than our assembly (Complete=85.7%, [Single-copy=81.1%, Duplicated=4.6%], Fragmented=5.5%, Missing=8.8%; Figure 1). This is likely due to the shortcomings of our annotation pipeline choice (funannotate) as it is known to be conservative and fragment genes with incomplete evidence. Even so, with over 85% of BUSCOs present, our annotation proved suitable for downstream applications.

Gene Duplication Analyses

We used two separate tools to test for the presence of ancient whole genome duplications (WGD) within our genome. Both use a similar Ks approach, but one looks at all genes in the genome with greater than 40% sequence similarity (DupPipe; Barker et al 2010) whereas the other looks specifically at gene pairs within collinear blocks of genes (MCSCanX; Wang et al 2012). Both analyses provided similar results (Figure 2) that confirm the presence of the ancient WGD that has been previously documented in Solanceaous plant genomes (Xu et al 2017; Landis et al 2018). We further used MCScanX to classify all the gene duplications within the largest 50 contigs of our genome (Table 1, S2). We found that dispersed duplications are most common (N = 17951; Table 1), although it is important to note that McScanX tends to inflate this number with improperly classified duplications. Ancient whole genome duplications are likely to have generated a significant amount of the gene content observed in the *D. wrightii* genome as well (N = 7714 genes; Table 1) and this number is likely underestimated.

Differential Expression analysis

Our pairwise analysis of *Lema*-induced transcriptional changes showed that control and *Lema*-induced samples did not have separate profiles, but that induction increased the variance in plant gene expression profiles (Figure 3A). Our analyses identified 6358 significantly differentially expressed genes between control and *Lema*-induced plants. (Figure 3; Table S4). Roughly half of these (N = 3114) were up regulated in response to *L. daturaphila* feeding (Figure 3B). We did not conduct any dedicated GO term enrichment analysis, but our functional annotation identified some of these genes as playing a role in responses to herbivory by other plant species (Table S4, S5).

Our time course analysis generated a shorter list of differentially expressed genes (N = 368; Figure 4; Table S5), but similar results to our pairwise analysis, in terms of most genes being upregulated in response to herbivory (Figure 4B). We did notice a strong differentiation between 0h (pre-herbivory) and 12h post-herbivory treatments (Figure 4A), regardless of treatment type (control or herbivory), which may be a sign of background circadian rhythms in gene expression (Xu et al 2022). Our results show clear signs of sustained long-term herbivore-induced changes to gene expression (shown in our 48h and 96h timepoints; Figure 4B).

186 **Concluding Remarks**

187 In sum, we present a high quality long-read genome assembly for the sacred datura plant
188 (*Datura wrightii*). We then analyze its gene duplication history and use it as the basis for a genome-
189 guided analysis of herbivore-induced gene expression changes. Multiple tools supported the presence of
190 a well-documented ancient whole-genome duplication event in *Datura wrightii*; and our analysis
191 identified 1000s of differentially expressed genes, some of which have known functional annotations
192 based on comparison to existing reference genomes. Together these data provide a valuable resource
193 for plant scientists more broadly and will contribute to future studies of angiosperm evolution and
194 plant-insect interactions.

195 **Data Availability**

196 Assembly and raw HiFi reads can be found on NCBI (BioProject: PRJNA966699; BioSample:
197 SAMN34546691). Supporting datasets, which includes annotation, can be found at
198 https://github.com/caterpillar-coevolution/Datura-wrightii-genome-project alongside scripts used for
199 computational analyses. Raw transcriptome data will be available on NCBI GenBank (BioProject:
200 PRJNA966699)

201 **Acknowledgments & Funding**

202 We thank Professor Luciano Matzkin and Mike Barker for their guidance throughout the process
203 of working on the project. We would also like to thank the 2016 staff and volunteers of the
204 Southwestern Research Station. Funding was provided by an NSF postdoctoral research fellowship in
205 biology to JKG (#2010772).

206 **Conflict of Interest**

207 The authors declare no conflict of interest.

208 **Literature Cited**

209 Barker, M.S., Dlugosch, K.M., Dinh, L., Challa, R.S., Kane, N.C., King, M.G., Rieseberg, L.H., 2010.
210 EvoPipes.net: Bioinformatic Tools for Ecological and Evolutionary Genomics. Evol Bioinform Online 6,
211 EBO.S5861. https://doi.org/10.4137/EBO.S5861

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

215 Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021. Haplotype-resolved de novo assembly using
216 phased assembly graphs with hifiasm. Nat Methods 18, 170–175. https://doi.org/10.1038/s41592-020-
217 01056-5

218 De-la-Cruz, I.M., Hallab, A., Olivares-Pinto, U., Tapia-López, R., Velázquez-Márquez, S., Piñero, D.,
219 Oyama, K., Usadel, B., Núñez-Farfán, J., 2021. Genomic signatures of the evolution of defence against its
220 natural enemies in the poisonous and medicinal plant Datura stramonium (Solanaceae). Sci Rep 11, 882.
221 https://doi.org/10.1038/s41598-020-79194-1

222 Dupin, J., Smith, S.D., 2019. Corrigendum to: "Phylogenetics of Datureae (Solanaceae), including
223 description of the new genus Trompettia and re-circumscription of the tribe" [in Taxon 67: 359–375.
224 2018]. TAXON 68, 419–419. https://doi.org/10.1002/tax.12056

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

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

231 Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf
232 tissues. Phytochem Bull. 19:11–15.

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

239 Elle, E., Hare, J.D., 2002. Environmentally Induced Variation in Floral Traits Affects the Mating System in
240 Datura wrightii. Functional Ecology 16, 79–88.

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

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

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

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

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

253 Hosmani, P. S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., … Saha, S.,
254 2019. An improved de novo assembly and annotation of the tomato reference genome using single-
255 molecule sequencing, Hi-C proximity ligation and optical maps. Cold Spring Harbor Laboratory.
256 https://doi.org/10.1101/767764

257 Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A.,
258 Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter,

S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

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

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

Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., Soltis, P.S., 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. American Journal of Botany 105, 348–363. https://doi.org/10.1002/ajb2.1060

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

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

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

Olcerst, A. 2017. Variation in Induced Responses of Datura wrightii to Herbivore Attack: Plasticity of Volatile Organic Compound Emissions and Gene Expression Across Genotypes, Ontogeny, and a Single Attack – Doctoral Disseration accessed via ProQuest: https://www.proquest.com/openview/920bf36262446dfa983f0a67e1755bff/1?pq-origsite=gscholar&cbl=18750

Palmer, J., Stajich, J., 2019. nextgenusfs/funannotate: funannotate v1.5.3. https://doi.org/10.5281/zenodo.2604804

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

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

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

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

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

297 Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to Identify Repetitive Elements in Genomic
298 Sequences. Current Protocols in Bioinformatics 25, 4.10.1-4.10.14.
299 https://doi.org/10.1002/0471250953.bi0410s25

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

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

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

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

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

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

317

318 **Figure Legends & Table**

319 **Figure 1.** (left) Bar charts showing the gene content represented in both our assembly (top bar) and our
320 annotation (bottom bar). (right) Blobtools coverage plot showing that 64.79% of raw reads mapped to
321 our final assembly, and that all mapped reads were taxonomically identified as Solanaceae, indicating
322 that our assembly is contaminant free. The 35.21% of raw reads that did not map to our assembly is
323 likely the result of the haplotig purging step of the hifiasm process, which produces a haplotype-specific
324 assembly.

325 **Figure 2.** Results of DupPipe (left) and MCScanX (right) showing the frequency of non-synonymous
326 substitions throughout the genome. DupPipe analyzes global gene content, whereas MCScanX only
327 examines genes within syntenic blocks. Both analyses show a large peak at Ks = 0.5, indicative of ancient
328 whole-genome duplication.

329 **Figure 3.** Results of pairwise differential expression analysis using only the samples collected in May
330 2012. (A) Results of exploratory PCA to qualitatively screen for differences between treatment groups.
331 There was a clear qualitative difference between the control and *Lema*-induced samples. (B) Volcano
332 plot showing the relationship between significance and fold-change for each gene in our *D. wrightii*
333 genome annotation. A change of 2/-2 is marked for reference but was not used as a cutoff for any
334 analysis.

335 **Figure 4.** (A) Exploratory PCA of time course data. Although our samples did not form clear clusters,
336 there are definitive patterns. The 48h and 96h *Lema*-induced samples cluster in the top right, whereas
337 the controls for these time points cluster with the 0h samples in the top left. Ranges for 0h and 12h
338 samples do not overlap at all, suggesting that there may also be circadian cycles occurring. (B) Heatmap
339 showing the fold-change of top 50 most significant genes from our time course analysis (N = 368 total
340 significant genes).

341 **Table 1.** Results of McScanX gene duplication classifier function. Most of the gene content is the results
342 of dispersed duplications, but ancient whole-genome-duplications (WGD) have also generated
343 thousands of genes within the *D. wrightii* genome. IDs refer to duplication identifiers in the full gene list
344 (Table S2).

Duplication Type	***ID***	***Gene count***
Singleton	0	8301
Dispersed	1	17951
Proximal	2	2649
Tandem	3	3556
WGD/Segmental	4	7714

345

346

347 **Supplemental Materials**

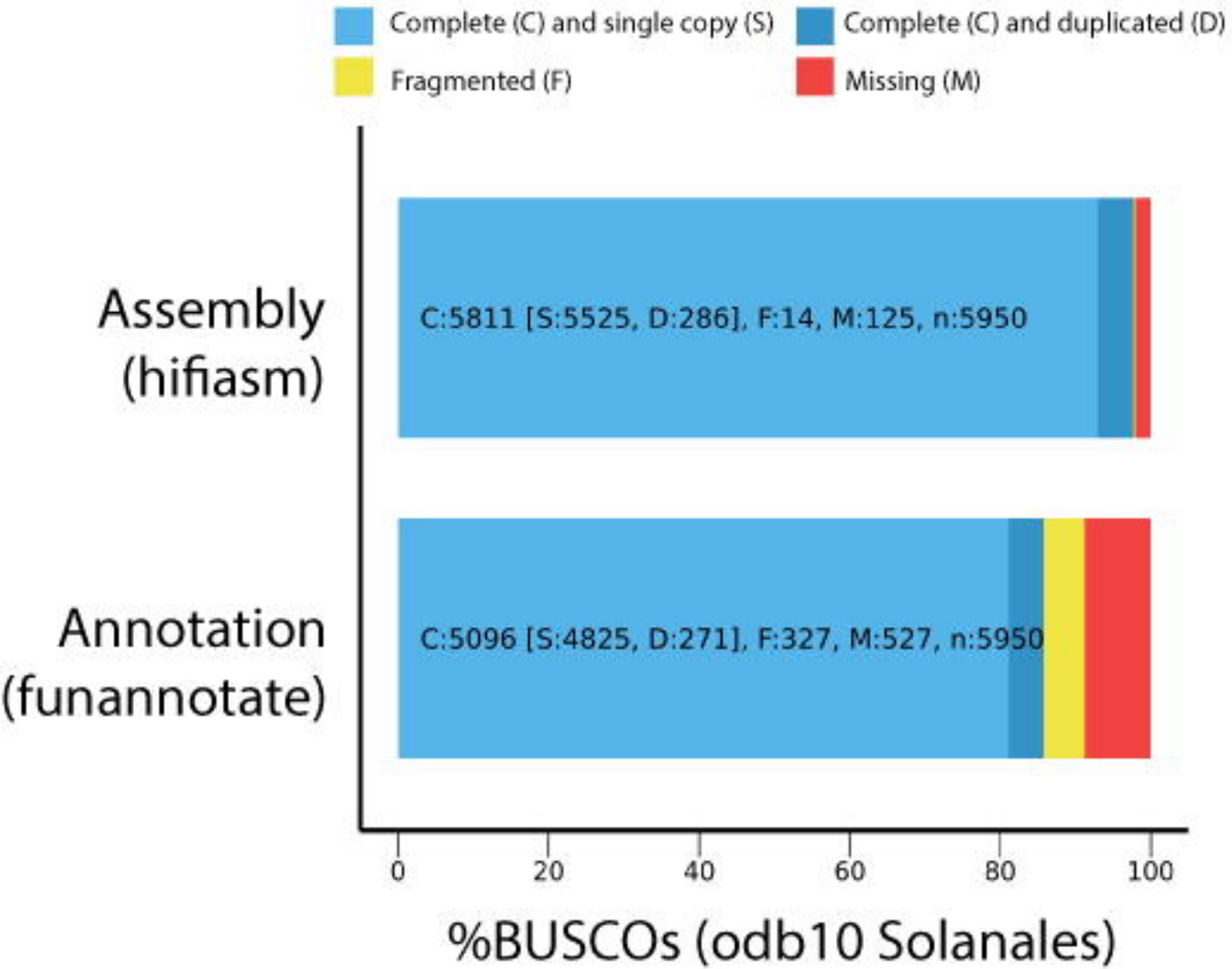
348 Table S1. Detailed output table from RepeatMasker.

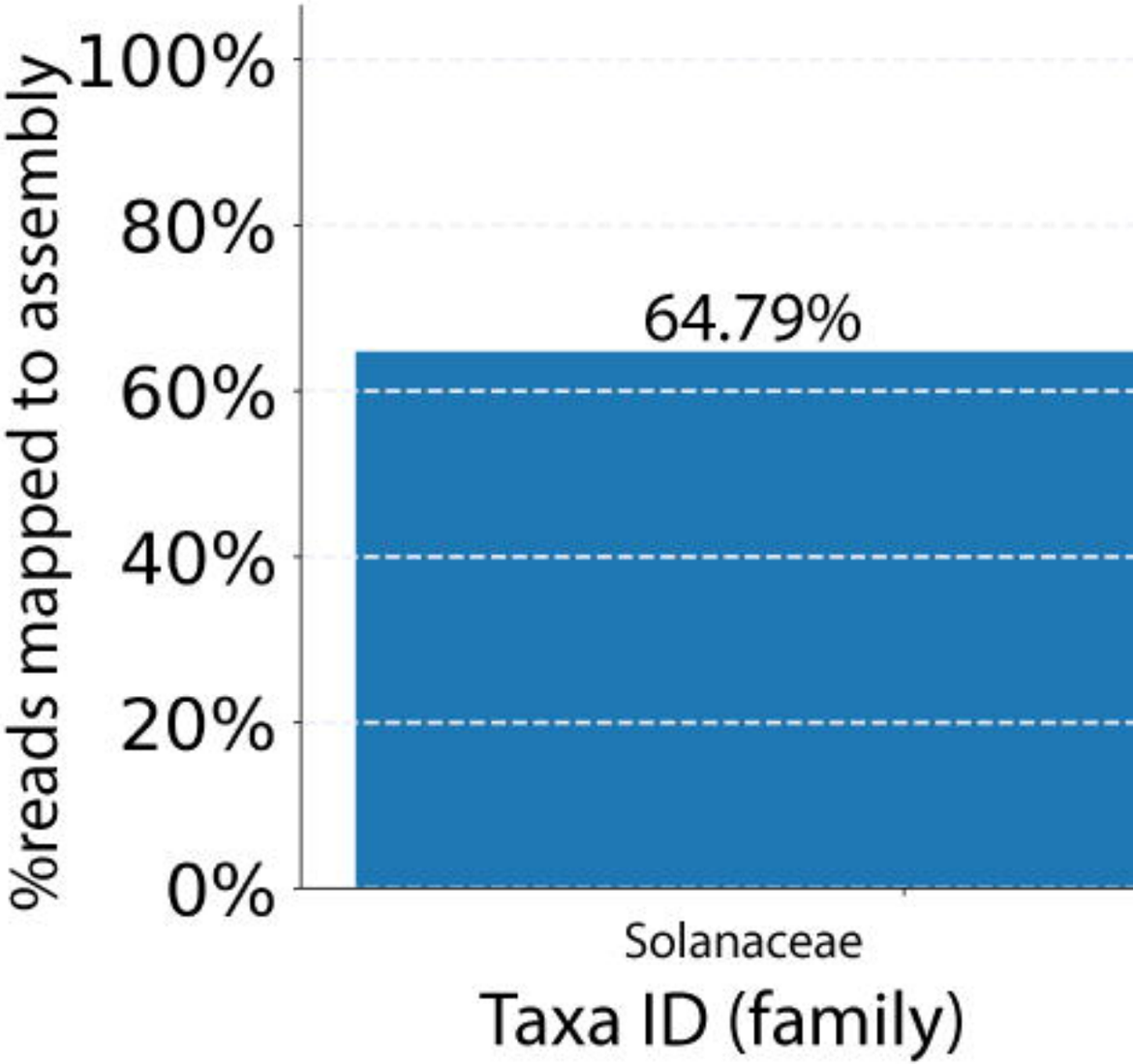
349 Table S2. Detailed results of MCScanX gene duplication classifier function

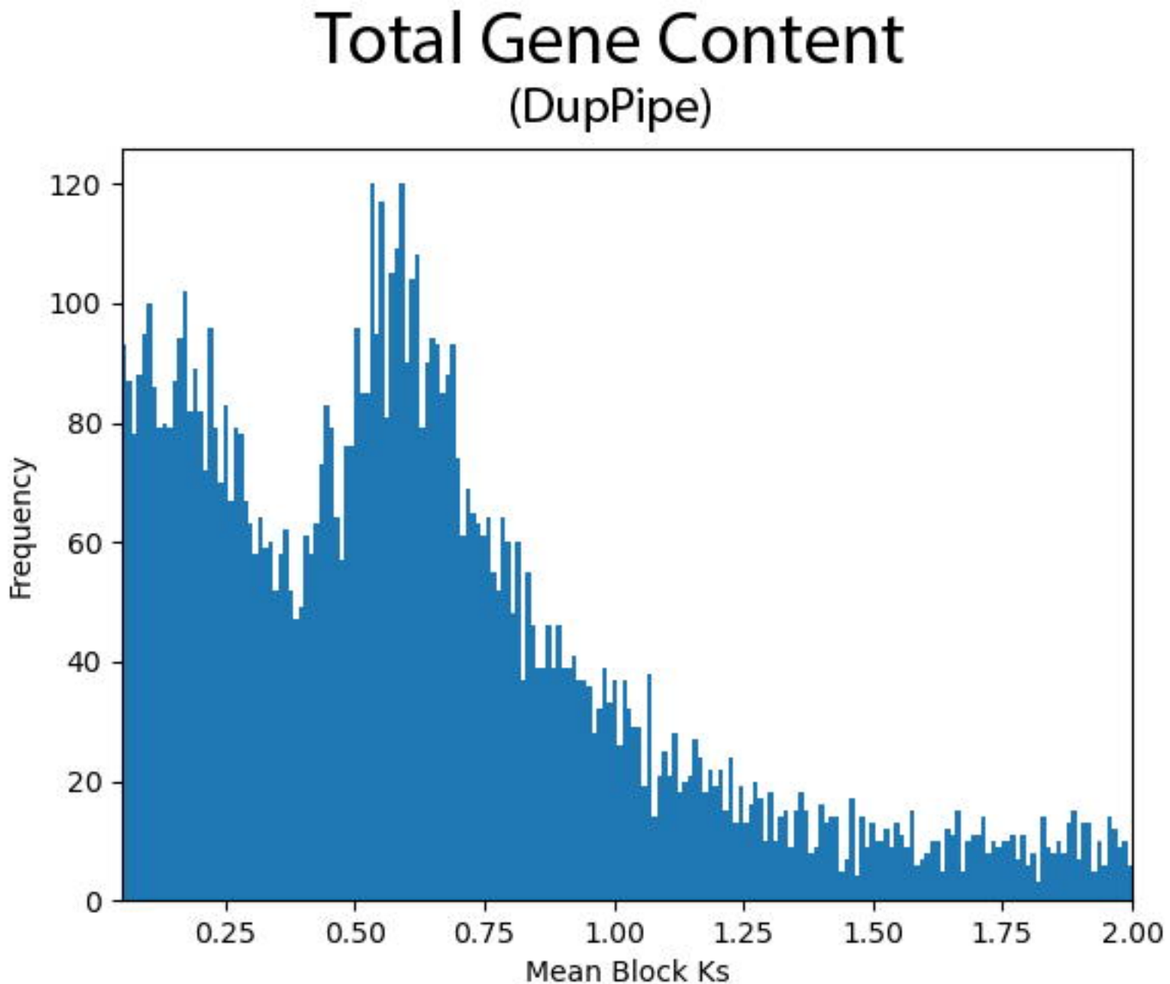
350 Table S3. Detailed output of STAR alignments

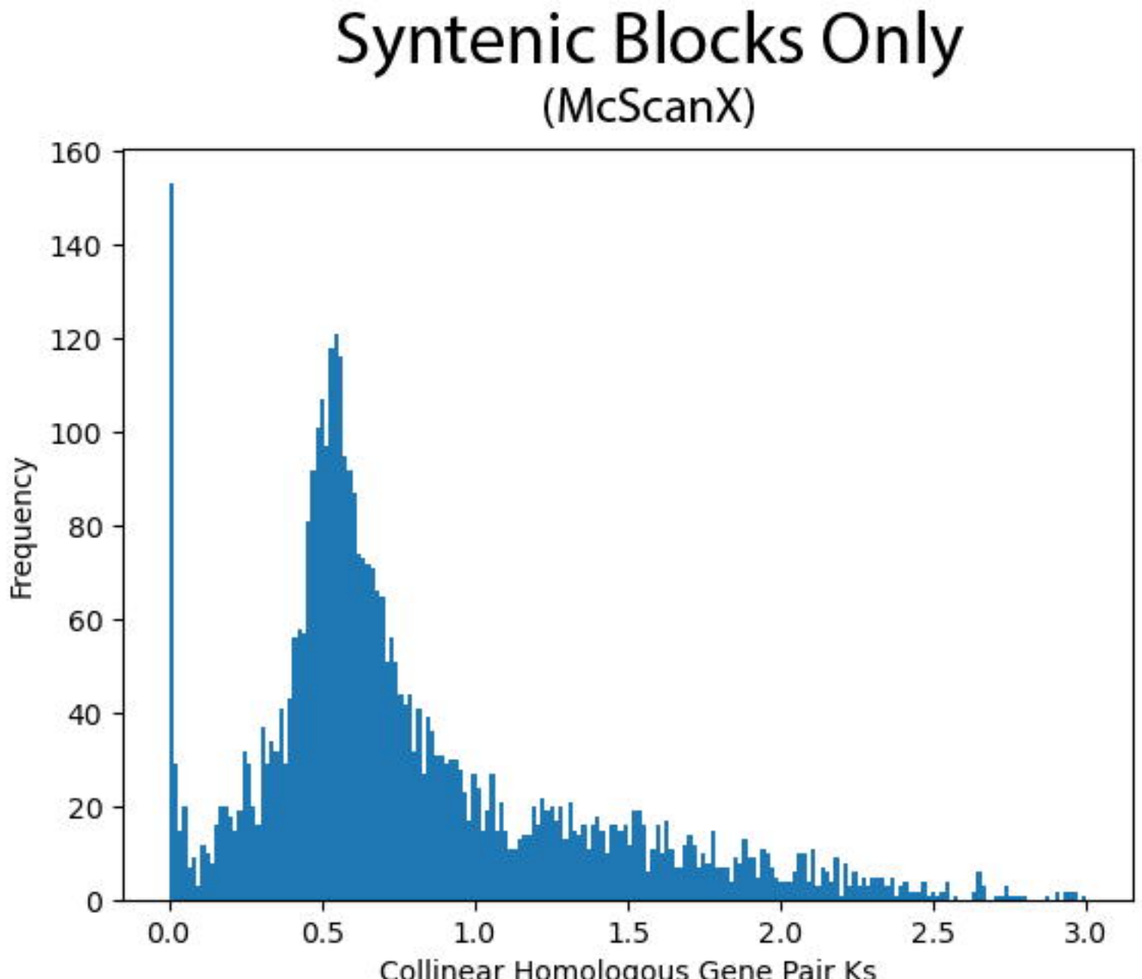
351 Table S4. Complete list of significant genes from pairwise differential expression analysis

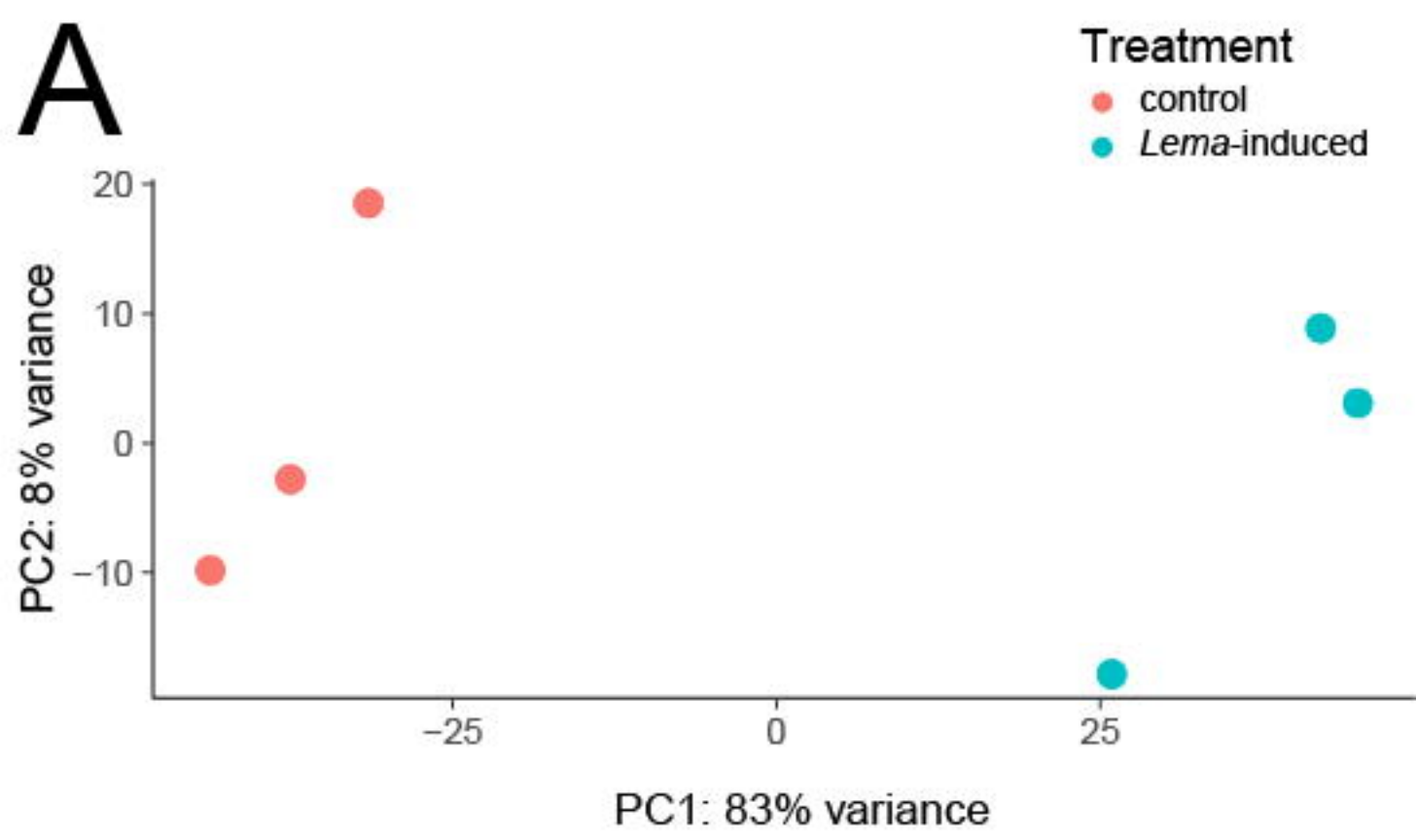
352 Table S5. Complete list of significant genes from time course expression analysis

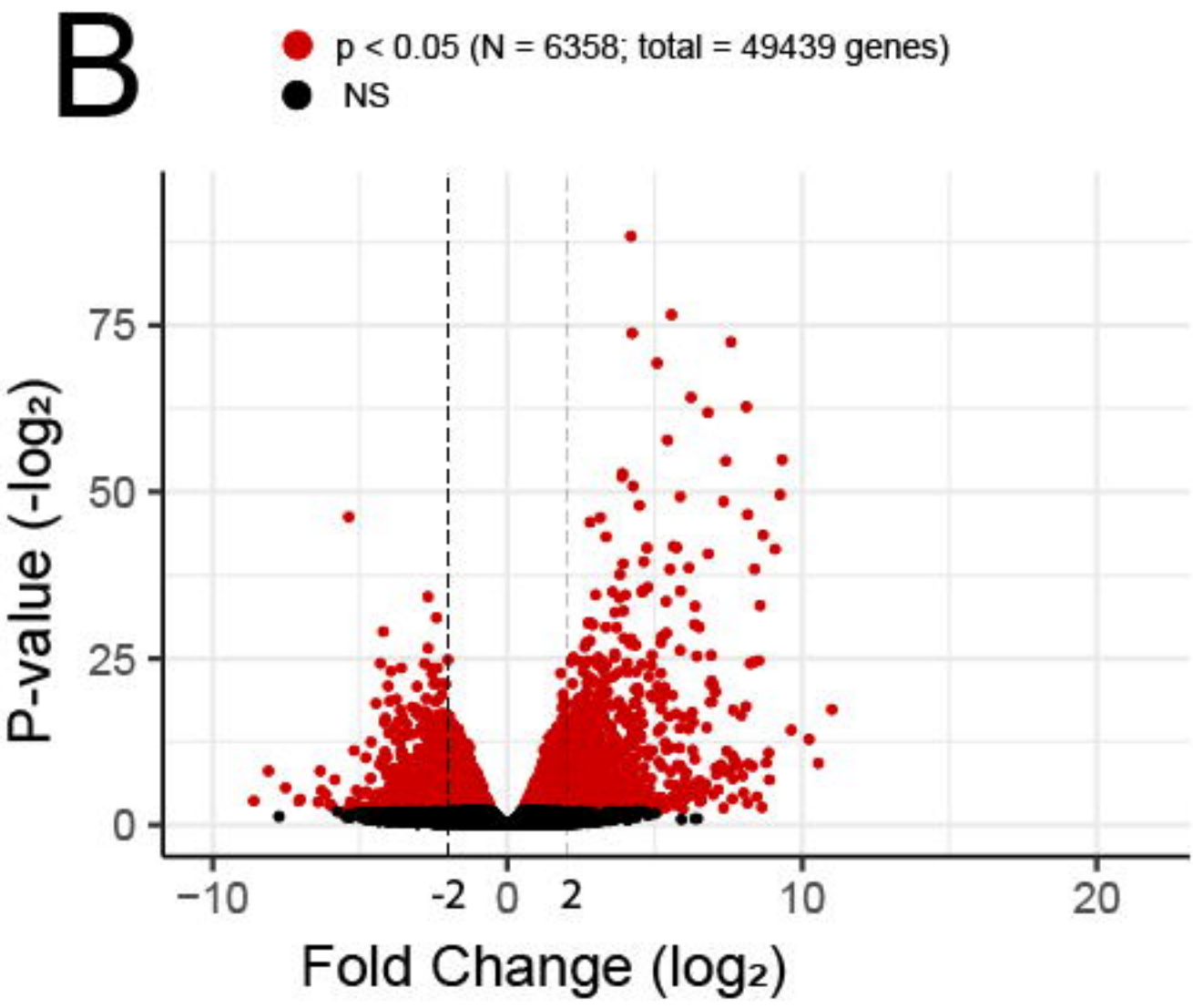

Complete (C) and single copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
Assembly (hifiasm)
C:5811 [S:5525, D:286], F:14, M:125, n:5950
Annotation (funannotate)
C:5096 [S:4825, D:271], F:327, M:527, n:5950
0
20
40
60
80
100
%BUSCOs (odb10 Solanales)

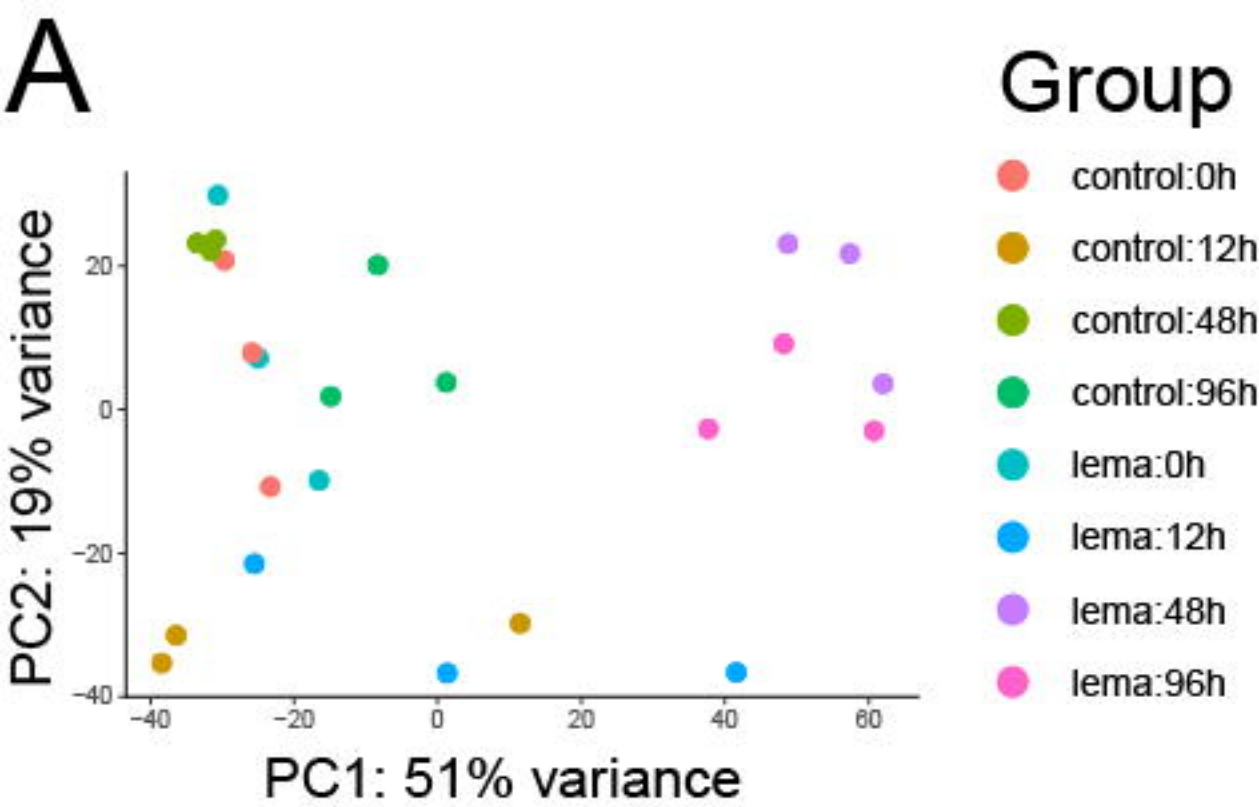

100%
80%
60%
40%
20%
0%
%reads mapped to assembly
64.79%
Solanaceae
Taxa ID (family)

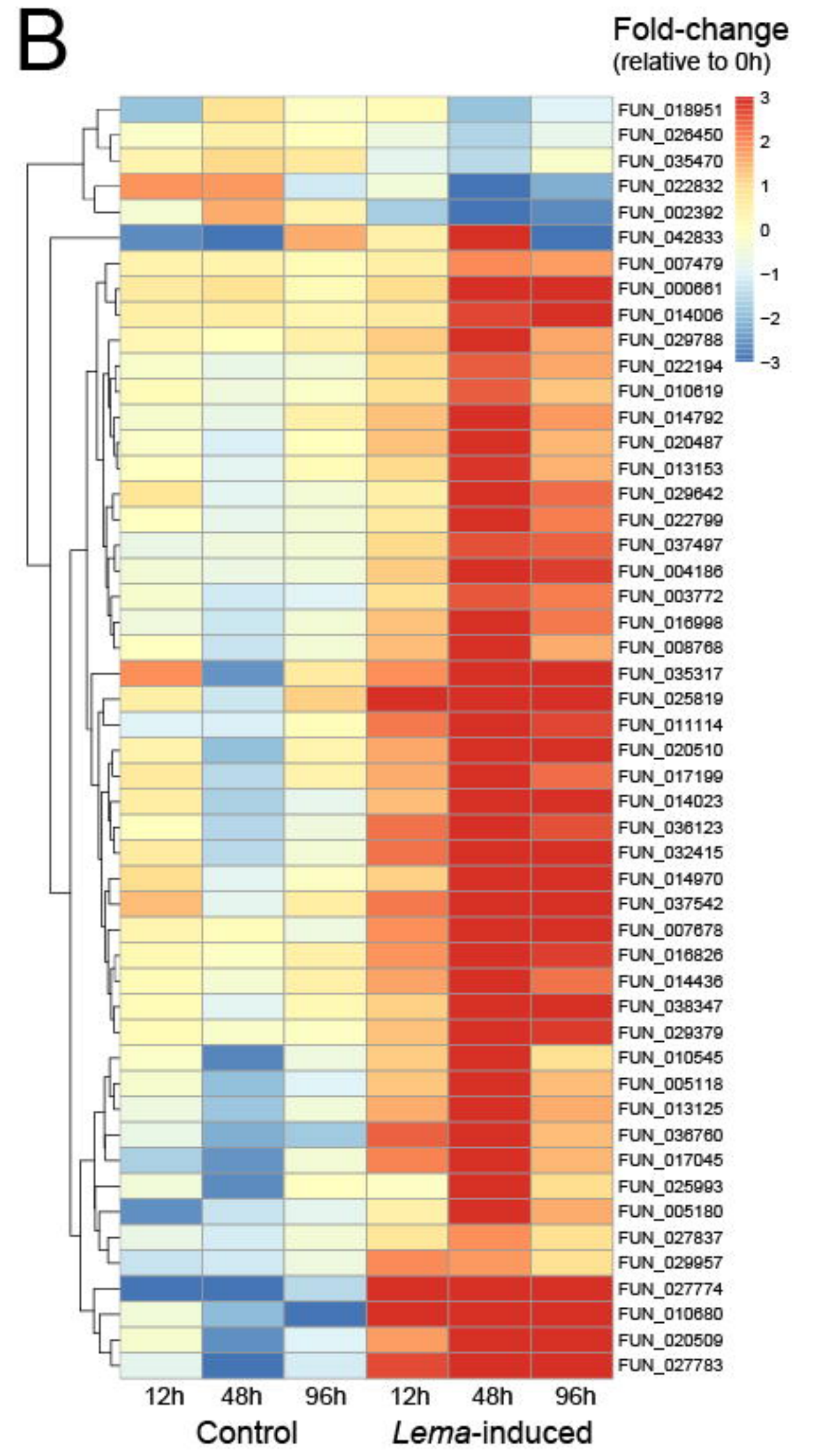

Total Gene Content
(DupPipe)
Frequency
Mean Block Ks
0
20
40
60
80
100
120
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00


Syntenic Blocks Only
(McScanX)
Frequency
Collinear Homologous Gene Pair Ks
0
20
40
60
80
100
120
140
160
0.0
0.5
1.0
1.5
2.0
2.5
3.0


A
Treatment
control
Lema-induced
20
10
0
−10
PC2: 8% variance
−25
0
25
PC1: 83% variance


B
p < 0.05 (N = 6358; total = 49439 genes)
NS
75
50
25
0
P-value (-log2)
−10
-2
0
2
10
20
Fold Change (log2)


A
Group
control:0h
control:12h
control:48h
control:96h
lema:0h
lema:12h
lema:48h
lema:96h
PC2: 19% variance
PC1: 51% variance
20
0
−20
−40
−40
−20
0
20
40
60


B
Fold-change
(relative to 0h)
3
2
1
0
−1
−2
−3
FUN_018951
FUN_026450
FUN_035470
FUN_022832
FUN_002392
FUN_042833
FUN_007479
FUN_000661
FUN_014006
FUN_029788
FUN_022194
FUN_010619
FUN_014792
FUN_020487
FUN_013153
FUN_029642
FUN_022799
FUN_037497
FUN_004186
FUN_003772
FUN_016998
FUN_008768
FUN_035317
FUN_025819
FUN_011114
FUN_020510
FUN_017199
FUN_014023
FUN_036123
FUN_032415
FUN_014970
FUN_037542
FUN_007678
FUN_016826
FUN_014436
FUN_038347
FUN_029379
FUN_010545
FUN_005118
FUN_013125
FUN_036760
FUN_017045
FUN_025993
FUN_005180
FUN_027837
FUN_029957
FUN_027774
FUN_010680
FUN_020509
FUN_027783
12h
48h
96h
12h
48h
96h
Control
Lema-induced