

1 **The genomic basis of the plant island syndrome in Darwin's giant daisies**

2

3 **[Longer title] A chromosome-resolved polyploid assembly for Darwin's giant daisy radiation**
4 **(*Scalesia*, Galápagos) reveals the genomic basis of the plant island syndrome**

5

6 José Cerca^{*1}, Bent Petersen^{2,3}, José Miguel Lazaro Guevara⁴, Angel Rivera-Colón⁵, Siri Birkeland⁶,
7 Joel Vizueta⁸, Siyu Li⁹, João Loureiro¹⁰, Chatchai Kosawang¹¹, Patricia Jaramillo Díaz^{12,13},
8 Gonzalo Rivas-Torres^{14,15,16}, Mario Fernández-Mazuecos¹⁷, Pablo Vargas¹⁸, Ross McCauley¹⁹,
9 Gitte Petersen²⁰, Luisa Santos-Bay², Nathan Wales²¹, Julian Catchen⁵, Daniel Machado²², Michael
10 D. Nowak⁷, Alexander Suh^{23,24}, Neelima Sinha⁹, Lene R. Nielsen¹¹, Ole Seberg²⁵, M. Thomas P.
11 Gilbert^{1,2}, James H. Leebens-Mack²⁶, Loren Rieseberg⁴, Michael D. Martin^{*.1}

12

13 1 - Department of Natural History, NTNU University Museum, Norwegian University of Science
14 and Technology, Trondheim, Norway

15 2 - Centre for Evolutionary Hologenomics, The GLOBE Institute, Faculty of Health and Medical
16 Sciences, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen, Denmark

17 3 - Centre of Excellence for Omics-Driven Computational Biodiscovery, Faculty of Applied
18 Sciences, AIMST University, Kedah, Malaysia

19 4 - Department of Botany and Biodiversity Research Centre, University of British Columbia,
20 Vancouver, BC V6T 1Z4, Canada

21 5 - Department of Evolution, Ecology, and Behavior, University of Illinois at Urbana-Champaign,
22 IL, USA

23 6 - Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life
24 Sciences, Ås, Norway

25 7 - Natural History Museum, University of Oslo, Oslo, Norway

26 8 - Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of
27 Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark

28 9 - Department of Plant Biology, University of California, Davis, Davis, CA 95616

29 10 - Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada
30 Martim de Freitas 3000-095 Coimbra, Portugal

31 11 - Department of Geosciences and Natural Resource Management, University of Copenhagen,
32 Rolighedsvej 23, 1958, Frederiksberg C, Denmark

33 12 - Estación Científica Charles Darwin, Fundación Charles Darwin, Santa Cruz, Galápagos,
34 Ecuador

35 13 - Department of Botany and Plant Physiology, University of Malaga, Spain.

36 14 - Colegio de Ciencias Biológicas y Ambientales COCIBA & Galapagos Institute for the Arts and
37 Sciences GAIAS, Universidad San Francisco de Quito USFQ, Quito, Ecuador

38 15 - Estación de Biodiversidad Tiputini, Colegio de Ciencias Biológicas y Ambientales,
39 Universidad San Francisco de Quito USFQ, Quito, Ecuador

40 16 - Estación de Biodiversidad Tiputini, Colegio de Ciencias Biológicas y Ambientales,
41 Universidad San Francisco de Quito - USFQ, Quito, Ecuador

42 17 - Departamento de Biología, Universidad Autónoma de Madrid, 28049 Madrid, Spain

43 18 - Departamento de Biodiversidad y Conservación, Real Jardín Botánico (RJB-CSIC), Plaza de
44 Murillo 2, 28014 Madrid, Spain

45 19 - Department of Biology, Fort Lewis College, Durango, CO 81301, USA

46 20 - Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91
47 Stockholm, Sweden

48 21 - Department of Archaeology, University of York, York, UK
49 22 - Department of Biotechnology and Food Science, Norwegian University of Science and
50 Technology, Trondheim, 7491, Norway.
51 23 - School of Biological Sciences, University of East Anglia, Norwich Research Park, NR4 7TU,
52 Norwich, UK
53 24 - Department of Organismal Biology, Evolutionary Biology Centre (EBC), Science for Life
54 Laboratory, Uppsala University, 75236 Uppsala, Sweden
55 25 - The Natural History Museum of Denmark, University of Copenhagen, Denmark
56 26 - Department of Plant Biology, University of Georgia, Athens GA 30602, U.S.A.
57 * Corresponding authors: José Cerca jose.cerca@gmail.com; Michael D. Martin
58 mike.martin@ntnu.no

59 Abstract

60 Oceanic archipelagos comprise multiple disparate environments over small geographic
61 areas and are isolated from other biotas. These conditions have led to some of the most
62 spectacular adaptive radiations, which have been key to our understanding of evolution, and
63 offer a unique chance to characterise the genomic basis underlying rapid and pronounced
64 phenotypic changes. Repeated patterns of evolutionary change in plants on oceanic
65 archipelagos, i.e. the plant island syndrome, include changes in leaf morphology, acquisition of
66 perennial life-style, and change of ploidy. Here, we describe the genome of the critically
67 endangered and Galápagos endemic *Scalesia atractyloides* Arnot., obtaining a chromosome-
68 resolved 3.2-Gbp assembly with 43,093 candidate gene models. Using a combination of fossil
69 transposable elements, *k*-mer spectra analyses and orthologue assignment, we identify the two
70 ancestral subgenomes and date their divergence and the polyploidization event, concluding that
71 the ancestor of all *Scalesia* species on the Galápagos was an allotetraploid. There are a
72 comparable number of genes and transposable elements across the two subgenomes, and while
73 their synteny has been mostly conserved, we find multiple inversions that may have facilitated
74 adaptation. We identify clear signatures of selection across genes associated with vascular
75 development, life-growth, adaptation to salinity and changes in flowering time, thus finding
76 compelling evidence for a genomic basis of island syndrome in Darwin's giant daisy radiation.
77 This work advances understanding of factors influencing subgenome divergence in polyploid
78 genomes, and characterizes the quick and pronounced genomic changes in a spectacular and
79 diverse radiation of an iconic island plant radiation.

80 Introduction

81 As naturalists set sail to explore the world, the distinctiveness of insular species stood
82 out from the remaining biota. The collections carried out in the Galápagos, Cape Verde and
83 Malay archipelagos were key for the development of the theory of natural selection (C. Darwin
84 1859; B. Y. C. Darwin et al. n.d.) and biogeography (Wallace 1962). More recently, Ernst Mayr's
85 work, which set the scene for the modern synthesis (Mayr 1942), focused heavily on island
86 biota (Emerson 2008). The central role of remote archipelagos in our understanding of
87 evolution is not coincidental. Organisms colonizing these regions encounter highly distinct
88 microenvironments that provide abundant ecological niches and thus ideal conditions for rapid
89 adaptive radiation (Lomolino, Riddle, and Whittaker 2017). The 'island syndrome hypothesis'

90 predicts the repeated and pronounced phenotypic shifts that species may undergo after
91 colonizing islands, as a result of novel selective pressures and empty ecological space (Baeckens
92 and Van Damme 2020). While the island syndrome hypothesis has been well established (Burns
93 2019; Baeckens and Van Damme 2020), its integration with genomic evidence still lags. For
94 instance, while body size differences in animal lineages are the textbook example of an island
95 syndrome (e.g., pygmy mammoths and giant tortoises), the extent to which these changes are
96 hereditary (genetic) or induced by different food sources (diet) has yet to be documented for
97 many lineages. Considering the rapid and drastic changes characteristic of these radiations, it
98 can be expected that rearrangements in genome structure contribute to the adaptation to novel
99 environmental conditions.

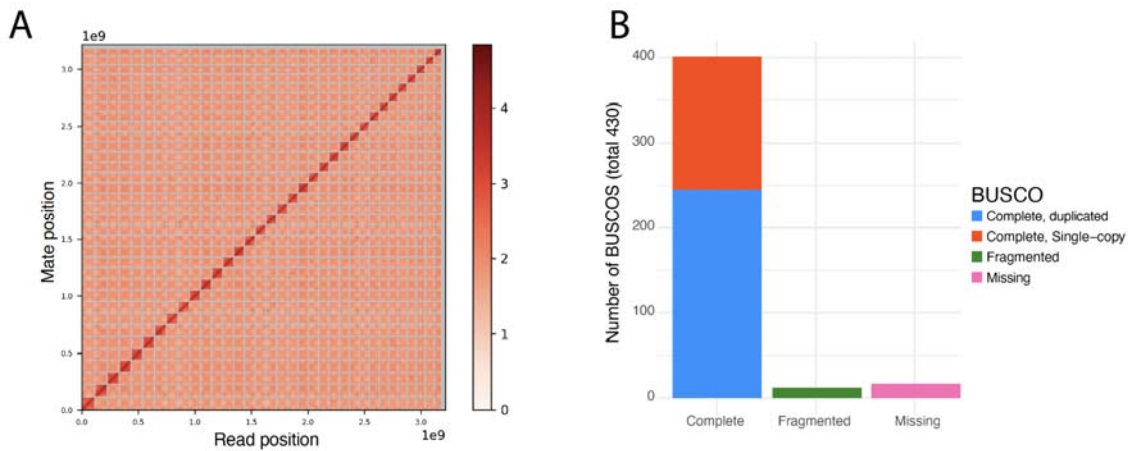
100 Because the most prominent examples of adaptive radiation and island syndromes
101 feature animal lineages, such as Darwin's finches, our understanding of these phenomena in
102 plant lineages lags (Burns 2019). As plants colonize archipelagos, they typically and repeatedly
103 undergo shifts in leaf morphology, dispersal ability, lifespan and size (Burns 2019). This is
104 exemplified by the daisy family (Asteraceae) which has radiated in Hawai'i (*Bidens* radiation
105 and silversword radiations) (Knape et al. 2012; Baldwin and Sanderson 1998; Knape et al.
106 2020), Macaronesia (*Sonchus*, *Tolpis*, *Argyranthemum* and *Cheirolophus* radiations) (S. C. Kim
107 et al. 1996; Gruenstaedl, Santos-Guerra, and Jansen 2013; Vitales et al. 2014; White et al. 2018,
108 2020), Juan Fernández (*Erigeron* and *Robinsonia* radiation), Ryukyu (*Ainsliaea* radiation)
109 (Mitsui and Setoguchi 2012), Tristan da Cunha (*Commidendrum* and *Melanodendron* radiation)
110 (Eastwood, Gibby, and Cronk 2004), Mauritius (*Psiadia* radiation) (Besse et al. 2003), and
111 Polynesia (*Tetramolopium* radiation)(Whitkus 1998).

112 One iconic, yet understudied, plant radiation is the remarkable diversification of daisies
113 in the genus *Scalesia* (Blaschke and Sanders 2009; Fernández-Mazuecos et al. 2020; Crawford et
114 al. 2009; U. Eliasson and U 1974). This genus consists of ca. 15 species, which have colonized
115 moist forest, littoral, arid, dry forest, volcanic soil, lava gravel and fissured environments across
116 varied elevations (Itow 1995; Blaschke and Sanders 2009). This adaptive ability has been linked
117 to *Scalesia's* exceptional variation in leaf morphology, which may be associated with the
118 adaptation to dry and humid environments (Stöcklin 2009; Fernández-Mazuecos et al. 2020),
119 capillum morphology and habit. The outstanding phenotypic and ecological variation has led
120 previous authors to refer to it as the 'Darwin finches of the plant world' (Stöcklin 2009). All
121 *Scalesia* species are ancestrally tetraploid ($2n=4x=68$) (Ono 1967; Uno Eliasson 1974), and the
122 polyploid genetics may have provided the genetic grist for adaptive radiation, as has been
123 speculated for other island floras (Meudt et al. 2021).

124 Here, we describe a high-quality chromosomal reference genome assembly and
125 annotation for *Scalesia atractyloides*. This species was chosen since it is a critically endangered
126 species and because it belongs to the most basal lineage in the endemic radiation (Fernández-
127 Mazuecos et al. 2020). A chromosome-resolved assembly has allowed us to identify and
128 separate the two ancestral genomes that united in the polyploidization event, and to compare
129 gene and transposable element distribution across and between these subgenomes. Annotation
130 of genes using PacBio IsoSeq RNA afforded a high-quality annotation of the genome, and the
131 detection of selection and gene-family expansions that implicate the genomic basis for island
132 syndrome traits in plants.

133 Results and discussion

134 Genome assembly, annotation, and quality control



135

136

137 **Figure 1.** Chromosome-resolved assembly of the *Scalesia atractyloides* nuclear genome. A) Link
138 density histogram, with 34 linkage groups (chromosome models) identified by contiguity
139 ligation sequencing. The x and y axes show mapping positions of the first and second read in
140 read pairs. B) Viridiplantae BUSCO set, which offers a characterization of universally conserved
141 orthologue genes.

142

143 The *Scalesia atractyloides* genome assembly is of remarkably high contiguity (Figure
144 1A), consisting of 3,216,878,694 base pairs (3.22 Gbp) distributed over 34 chromosome models,
145 in line with previous cytological evidence (Spring, Heil, and Vogler 1997; Uno Eliasson and
146 Others 1974; Schilling, Panero, and Eliasson 1994). The N_{90} was of 31, corresponding to all but
147 the three smallest chromosomes ($n = 34$) and LN_{90} was 81.66 Mbp. Flow cytometry estimates
148 (Supplementary Information; Supplementary Table 01), however, suggest a genome size of ca.
149 3.9 Gbp, and thus ~700 Mbp were likely collapsed by the assembler or removed by
150 *purgehaplotigs* (Peona, Weissensteiner, and Suh 2018). Despite this likely collapse of repeats,
151 we were able to annotate 76.22% of the genome as repeats, which were masked by
152 *RepeatMasker* (~2.5 Gbp). Considering the whole genome, 47.9% of the genome was composed
153 of long terminal repeat (LTR) retroelements, of which 16.2% were Copia and 31.54% Gypsy
154 elements (Supplementary Information; Supplementary Table 2), and 26.32% were unclassified
155 repeats.

156

157 The IsoSeq transcriptome recovered 46,375 genes and 224,234 isoforms
158 (Supplementary Information; Supplementary Figure 01). Using this as evidence and *ab initio*
159 models, we retrieved 43,093 genes from the annotation. Of the 430 Viridiplantae odb 10 BUSCO
160 groups used in a search of the genome (Figure 1B), 401 were found as complete (93.3%), of
161 which 245 were found as duplicate (57%), 156 as complete and single-copy (36.3%), and 12 as
162 fragmented (2.8%). Only 17 were absent (3.9%). When running OrthoFinder including *Scalesia*
163 and five other Asteraceae chromosome-resolved assemblies, we found that 34% of all the
164 orthogroups included genes from the five genomes. This overlap indicates a high-quality gene
165 annotation (Supplementary Information; Supplementary Figure 02). The proportion of
annotated repeats and number of genes is within the variation reported for Asteraceae. For

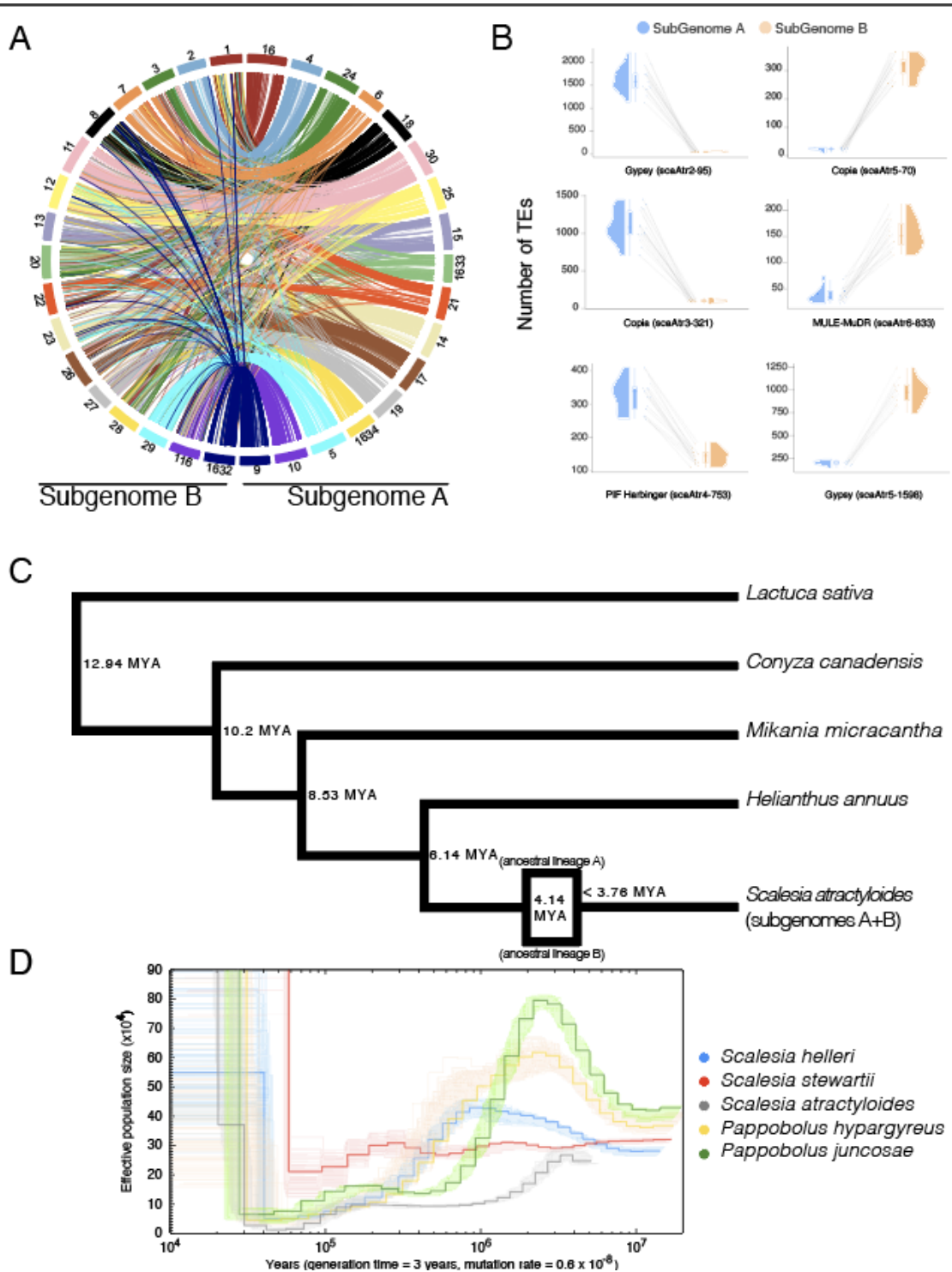
166 instance, the assembly of the closely related sunflower (*Helianthus annuus*) reference genome
 167 includes ~52,000 protein coding genes and has a repeat content of 74% (Badouin et al. 2017).

168

169 **Subgenome identification & evolution**

170

171



172

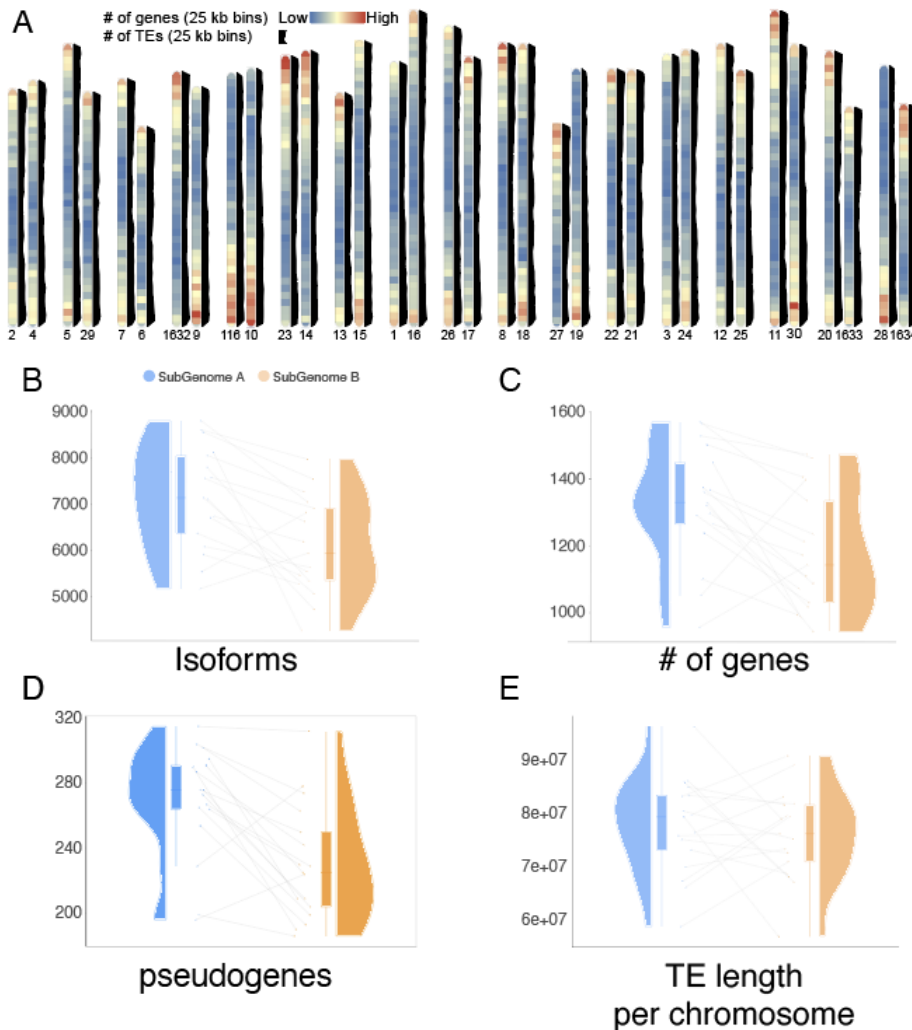
173

174 **Figure 2.** Subgenomes and evolutionary history of *Scalesia*. A) Circos plots displaying the 34
175 pseudo-chromosomes in the assembly. Pairs are organized to the left and right from the top, and
176 have the same colour-coding; B) Families of transposable elements (TEs) that are differently
177 represented on each subgenome. These TE families were likely active while the two
178 subgenomes were separated and thus confirm subgenome identification. Each data point
179 corresponds to a chromosome in a subgenome (subgenome A in blue and B in orange).
180 Chromosome pairs are linked by grey lines; C) Single-copy ortholog phylogeny of the studied
181 Asteraceae genome assemblies. Node ages are provided to the right of each node, as well as the
182 predicted time for the polyploidization event. D) Pairwise sequentially Markovian coalescent
183 (PSMC) estimation of the demographic history of *Scalesia atractyloides*, two other *Scalesia*
184 species, and two members of the *Pappobolus* genus, which is the sister taxon to *Scalesia*.
185

186 The identification of subgenomes was carried in two steps. On the first step, we assigned
187 the 34 chromosomes into 17 homeolog pairs by identifying and mapping duplicated conserved
188 orthologous sequences (COS; Supplementary Information; Supplementary Table 03; Figure 2 A).
189 This step only allowed the identification of pairs (homeologs), and did not allow the assignment
190 of subgenome identity within pairs. Homeolog exchanges were therefore not a concern here
191 (Edger et al. 2018). On the second step, we used the *k*-mer spectrum to identify ‘fossil
192 transposable elements’, which are transposable elements that were replicating while both
193 subgenomes were separate (i.e. after the speciation event, before the polyploidization event).
194 Since different genomes accumulate different transposable elements, we hypothesized that
195 some transposable element families in either subgenome will maintain frequency biases
196 (Session et al. 2016; Mitros et al. 2020). In short, transposable element families active before the
197 divergence of the two parental lineages are predicted to be approximately equally represented
198 in either subgenome, whereas elements activated after the divergence of the parental species
199 are predicted to be differently represented on the *Scalesia* subgenomes. Using the *k*-mer
200 spectrum, we selected *k*-mers that were in high numbers (i.e. repeats/TEs) and unevenly
201 represented between chromosome-pairs identified in the previous step (i.e. active during the
202 separation period). Using this selection of *k*-mers we ran a hierarchical clustering approach that
203 grouped chromosomes into two groups (two subgenomes; Supplementary Information;
204 Supplementary Figure 03). To confirm this assignment, we explored the output from
205 *RepeatMasker*, finding transposable element families unevenly represented across subgenomes
206 (Figure 2B), as predicted. The identification of differently represented transposable element
207 families also provides compelling evidence that the *Scalesia* radiation is of allopolyploid lineage.
208 Island floras are characterized by a high degree of paleo-allopolyploids, where the variation
209 brought forth by ploidy may underpin the diversification to multiple environments (Julca et al.
210 2020; te Beest et al. 2012) - a scenario which is line with the evolutionary history of *Scalesia*.

211 Using four other chromosome-level assemblies from Asteraceae (*Helianthus annuus*,
212 *Conyza canadensis*, *Mikania micrantha*, and *Lactuca sativa*) and the two subgenomes, we
213 estimated groups of orthologous genes using OrthoFinder. We obtained 710 orthogroups in
214 which each genome had only a single member, tolerating no missing data, and used this data to
215 construct a phylogenetic tree. The tree topology agrees with the placement of the Asteraceae
216 lineages from a recent and comprehensive set of genomic analyses (Mandel et al. 2019). Dating
217 of this tree was done by constraining the node separating the *Scalesia* subgenomes and
218 *Helianthus* as 6.14 Mya (Figure 2C), following recent literature (Mandel et al. 2019). This
219 suggests that the subgenomes diverged from their MRCA roughly 4.14 Mya, but the separation
220 of the ancestral lineages only lasted ~0.5 My, as calculated by LTR-family divergence

221 (Supplementary Information; Supplementary Figure 04). Specifically, the ancestral genomes
222 reunited in a single polyploid genome at least 3.76 Mya (Figure 2C; Supplementary Information;
223 Supplementary Figure 04). These dates are concordant with the *PSMC* analysis which roughly
224 indicate that the three *Scalesia* species had concordant population sizes of 250,000-300,000
225 circa 4 MYA (Figure 2D). Mismatches between the three genomes could result from variation in
226 generation time in *Scalesia* (see results and discussion below), and bottlenecks suffered by
227 populations as a result of climatic shifts in the Galapagos (Whittaker, School of Geography
228 Robert J Whittaker, and Fernandez-Palacios 2007). These estimates are concordant with a
229 recent dating analysis that estimated the divergence between *Pappobolus* and *Scalesia* occurred
230 ~3 Mya (Fernández-Mazuecos et al. 2020).
231
232



233
234 **Figure 3.** Subgenome evolution and characterization. A) Ideogram with gene and
235 transposable element distribution in 25-kbp bins. Gene density is plotted in chromosome
236 representations and transposable element distribution is plotted to the side of each
237 chromosome in black. Chromosomes are arranged in homoelogenous pairs. B) Number of isoforms
238 detected for each subgenome. Each data point corresponds to a chromosome in a subgenome
239 (subgenome A in blue and B in orange). Chromosome pairs are linked by grey lines; C) Number

240 of genes detected for each subgenome; D) Number of pseudogenes detected for each
241 subgenome; E) Length of transposable elements detected for each subgenome.

242

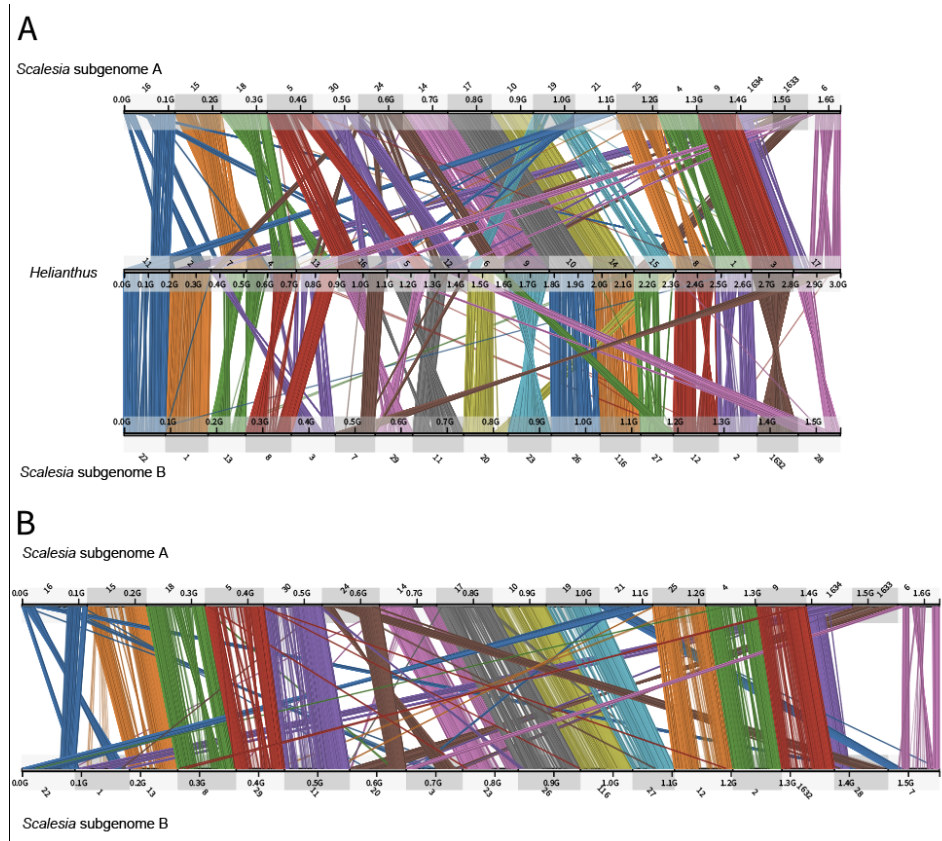
243 The identification of subgenomes allowed comparing the genes and transposable
244 element distribution across chromosome pairs. We find that gene density is highest near the
245 telomeres on both subgenomes, while transposable elements are more evenly distributed
246 throughout chromosomes (Figure 3A). This even distribution of transposable elements is
247 different from most other vascular plants, in which transposable element load is highest near
248 the center and decreases towards the ends of the chromosome, and rather is reminiscent of
249 observations in bryophyte genomes (Diop et al. 2020; F.-W. Li et al. 2020; Lang et al. 2018).
250 Even distributions of transposable elements were also observed in the sunflower genome
251 (Badouin et al. 2017), which may be indicative of particular transposable element regulation in
252 the Heliantheae.

253 As two genomes unite to form a single hybrid genome, an accommodation of the two
254 subgenomes, the process of 'diploidization' takes place (Bird et al. 2018; Freeling, Scanlon, and
255 Fowler 2015; Wolfe 2001). This process can occur very quickly, with changes in transcription
256 between subgenomes observed in 2-3 generations (Bird et al. 2021), and result in pronounced
257 changes in gene numbers. Whereas subgenome dominance in gene expression and retention has
258 been documented paleopolyploid plant genomes (Alger and Edger 2020; Renny-Byfield et al.
259 2015; Douglas et al. 2015), *Scalesia* subgenomes contain roughly equal gene and isoform
260 contents (Figure 3 B, C), as well as pseudogene numbers and transposable element load (Figure
261 3 D, E). In addition to this, when running the Viridiplantae BUSCO set for each subgenome
262 separately, we find 82.7% complete BUSCOs on subgenome A (76.6% single-copy, 6%
263 duplicates), and 81.9% complete BUSCOs (77% single-copy, 4.9% duplicates) on subgenome B.
264 Finally, both subgenomes are roughly the same length (subgenome A = 1,629,251,263 bp;
265 subgenome B = 1,554,170,668 bp), and have retained the same number of chromosomes
266 (Figure 3A). This indicates that during the past ~3.76 million years, during which the two
267 subgenomes have been unified in the same organism, there has not been a drastic
268 rearrangement of either subgenome, despite a smaller accumulation of genes and pseudogenes
269 on subgenome A. To explain this, we speculate that *Scalesia's* adaptation to insular
270 environments has benefitted from the genetic variation and diversity stemming from the
271 allopolyploidization event.

272

273 **Fast evolutionary rates in Heliantheae**

274



275
276

277 **Figure 4.** Chromosome stability plots reveal the role of inversions and translocations in the
278 differential in each subgenome. A) Chromosome stability plot between the two *Scalesia*
279 subgenomes and the *Helianthus annuus* genome. Each line connects a pair of orthologous genes,
280 colour-coded by chromosome pair. B) Chromosome stability between the two *Scalesia*
281 subgenomes. Each line connects orthologous genes in subgenome A and B, colour-coded by
282 chromosome pair.

283

284 To further dissect the mode and tempo of polyploid subgenome evolution, we used
285 *Synolog* (Catchen, Conery, and Postlethwait 2009) to create chromosome stability plots that
286 would allow us to detect translocations and inversions and to quantify their impact (Figure 4).
287 This method establishes clusters of conserved synteny by identifying single-copy orthologs
288 shared between two genomes via reciprocal BLAST searching between all annotated protein-
289 coding genes. From the identified synteny clusters, we calculated statistics on the orientation
290 (forward/inverted) and chromosome location. We thereby classified genes into four categories:
291 “Forward pair” (FP; i.e. not inverted, and the single-copy orthologs are in chromosomes from
292 the same pair), “Inverted pair” (IP; i.e. inverted, and the single-copy orthologs are in
293 chromosomes from the same pair), “Forward translocated” (FT; i.e. not inverted, and the
294 orthologues are not in chromosomes from the same pair), “Inverted translocated” (IT; i.e.
295 inverted, and the orthologues are not in chromosomes from the same pair). Comparing the two
296 *Scalesia* subgenomes, we found 2,284 FP genes (17%), 1,760 IP genes (13%), 3,539 FT genes
297 (27%), and 5,641 IT genes (43%), totalling 13,224 genes included in the analysis (Figure 4B).
298 Thus, the majority of the genes have been translocated (70%), and/or inverted (56%).
299 Interestingly, there is a mismatch between the length of these regions in the genome and the

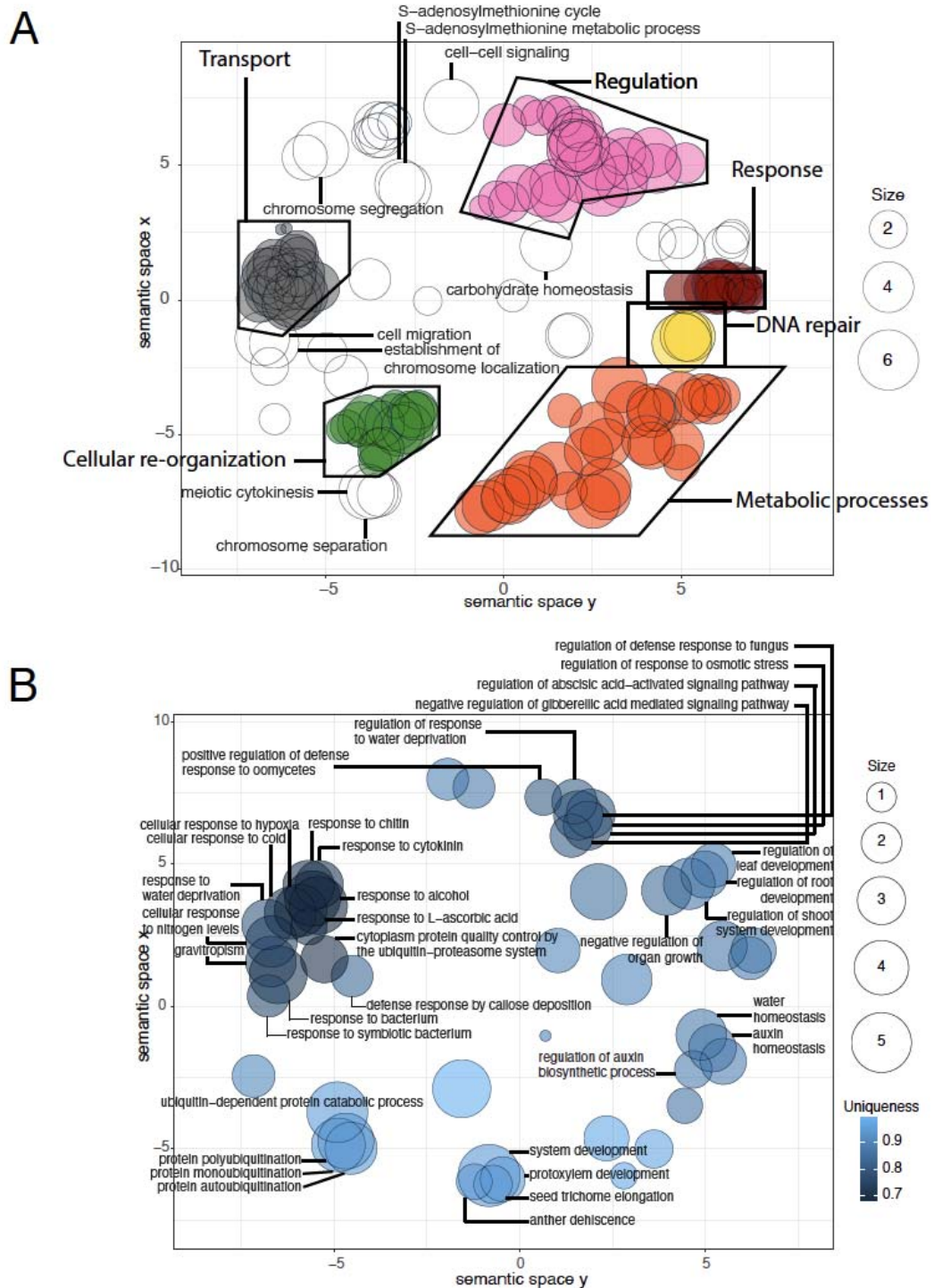
300 proportion of genes they contain. Specifically, we classified 434.6 Mbp as FP (23%), 189 Mbp as
301 IP (36%), 693,2 Mbp as FT (31%), and 586,5 Mbp as IT (10%). These results are in line with the
302 inference of rapid rates of chromosomal rearrangements in the Asteraceae based on
303 comparisons of the sunflower and lettuce genomes (Badouin et al. 2017). The discordance
304 between the fraction of genes, and the fraction of genome length (e.g., IT includes 43% of the
305 genes but only occupies 10% of the genome; 46% of the subgenomes is found as inverted, but
306 contains 56% of the genes) implicates inversions as having influenced the retention of
307 duplicated genes after formation of *Scalesia's* allopolyploid ancestor.

308 Despite its polyploid ancestry, the *Scalesia* genome has roughly the same length (3.22
309 Gbp) as that of the diploid sunflower (3.6 Gbp assembly). Further, *Scalesia's* annotated gene
310 model number (43k) is smaller than the number of gene models annotated in the sunflower
311 genome (52k) (Badouin et al. 2017). The sunflower has 17 chromosome pairs, whereas the
312 tetraploid *Scalesia* has 34 chromosome pairs. This indicates that *Scalesia* has likely undergone a
313 reduction in genome length and gene numbers without a reduction in chromosome number.
314 This evidence is consistent with hypothesized genome miniaturization in island species (J. Suda,
315 Kyncl, and Jarolímová 2005). We cannot, however, rule out a scenario in which *Helianthus* has
316 increased its number of genes and genome length.

317

318 **Evolutionary history of *Scalesia* & evidence for island syndromes**

319



320

321 **Figure 5.** Positive selection and gene family expansion across the *Scalesia atractyloides* genome.

322 A) GO term enrichment of the genes under selection across the genome. GO terms assigned to at

323 least four genes are labelled. Size refers to the number of genes associated with a particular GO

324 term. B) GO term enrichment of the genes belonging to expanded gene families across the

325 genome according to a *CAFE* analysis. Only GOs within a group of three or more overlapping

326 circles are included. Uniqueness measures the degree to which a particular GO term is distinct
327 relative to the whole list.

328

329 We identified 920 genes under selection ($p < 0.05$) in the *Scalesia* genome, after
330 correcting dN/dS ratios using a Holm-Bonferroni FDR correction. To understand their function,
331 we extracted the functional annotation using a Gene Ontology (GO) term enrichment analysis,
332 and the results were visualised using *Revigo*. GO classifications relating to many metabolic
333 processes under selection (Figure 5A, orange group), cellular reorganization (green group),
334 DNA repair (yellow group), response to protein folding (maroon group), and regulation
335 (regulation of metabolic processes, translation, gene expression, translation, nuclear division,
336 chromosome segregation, among others; pink group; Figure 5A; Supplementary Information;
337 Supplementary Tables 04-6). Genes inferred to have evolved under positive selection are also
338 associated with meiosis, chromosome arrangement, and chromatin status (meiotic cytokinesis,
339 establishment of chromosome location, chromosome separation, and chromosome segregation,
340 among other GO classifications; Figure 5A, Supplementary Information; Supplementary Tables
341 04-6), and this may indicate selection at genes associated with the coexistence of two genomes.

342 To chart the landscape of *Scalesia*'s adaptive revolution, we randomly selected 100
343 genes under selection and performed a systematic investigation of homolog function in
344 *Arabidopsis thaliana* literature. The classification of these genes into non-mutually exclusive
345 categories showed 23 genes implicated in photosynthesis and leaf morphology/metabolism, 17
346 in stress response, 15 in fertility and reproductive organ function, 13 in life-phase transition
347 and growth, 11 in the embryo, 10 in root functions, eight in immunity, seven in cellular
348 functions (mitochondrial functions, cytoplasm, DNA repair, nucleus, cell elongation or
349 autophagy), five in meiosis/mitosis, three in the vascular system, two in general metabolism,
350 and one in methylation (Supplementary Information; Supplementary Table 07). As leaf
351 morphology was an important innovation in the *Scalesia* radiation, it is particularly interesting
352 that we found selection on potential regulators of leaf morphology, including genes well known
353 to determine leaf cell number in *A. thaliana* (E2F1) (Ószi et al. 2020; Berckmans et al. 2011),
354 cell fate in leaves (YABBY5) (Kojima et al. 2011; Husbands et al. 2016), leaf senescence
355 (RANBPM, LARP1C, PEN3) (Crane et al. 2019; Fu et al. 2019; Zhang et al. 2012), leaf variegation
356 (THF1) (Ma et al. 2015; Z. Wang et al. 2016), and leaf growth (PAC) (Jörg Meurer et al. 2017;
357 Holding 2000; J. Meurer et al. 1998). It is also interesting to note that many *Scalesia* genes under
358 selection are affected by light stimulus. A STRING analysis showed that there was selection at
359 multiple points in the light regulatory pathways including responses to R/FR and blue light
360 responses. These include an inhibitor of red and far-red light photoreceptor (PHL) (Endo et al.
361 2014, 2013), a lysine-tRNA ligase that regulates photomorphogenic responses (G. Li et al. 2011),
362 an amino acid aminotransferase-like PLP-dependent enzymes superfamily protein that is
363 regulated under light conditions and is associated with the photorespiration process (Basset et
364 al. 2004; Smeekens 2006), and genes for which knock-out mutants experience alterations in
365 light reception (DJC69, COX15) (Oravec et al. 2006; Dal Bosco et al. 2004). This may underlie
366 the natural history observations where *Scalesia* individuals growing in the absence of
367 permanent light conditions show substantially retarded growth (Lawesson 1988) or higher
368 mortality (Rivas-Torres, *unpublished data*).

369 Many of the stress-response genes under selection in *Scalesia* are associated with
370 osmotic stress in *A. thaliana*, concomitant with evidence that the *Scalesia atractyloides* habitat is
371 characterized by arid conditions such as the Galápagos' arid zone, the littoral zone, and fissured
372 lava areas (Blaschke and Sanders 2009; Itow 1995). For instance, we found five genes (

373 “Leucine-rich repeat protein kinase family protein”, MPPBETA, “leaf osmotic stress elongation
374 factor 1- β -1”, AT2G21250, VAP27-1) associated with osmotic stress (Tan et al. 2010; J. Y. Kim et
375 al. 2007; ten Hove et al. 2011; Jakoby et al. 2008; Fox et al. 2020; P. Wang et al. 2019), but also
376 heat shock proteins (Valverde, Groover, and Romero 2018), and regulators of stomatal closure
377 (THF1) (Ma et al. 2015; Z. Wang et al. 2016) under selection. Other stress-associated genes
378 under selection include those involved in response to high irradiation (ZAT10, AT1G06690,
379 DDB2) (Bittner, Hause, and Baier 2021; Kuki et al. 2020; Czarnocka et al. 2020; Kleine et al.
380 2007; Castells et al. 2011; Lahari et al. 2018).

381 Some genes under selection are associated with growth and transitions between life
382 stages. *Scalesia* plants’ fast rates of growth have earned them the name ‘weedy trees’, and these
383 genes may regulate these plants’ exceptionally fast growth and tree-like habits. We find three
384 genes under selection that cause the transition between embryonic and vegetative traits
385 (RING1A, SWC4, ABCI20) (*A. Kim et al. 2020*), and four genes that regulate flowering time in *A.*
386 *thaliana* (ELF8, RING1A, Short-Vegetative-Phase, NRP1) (D. Chen et al. 2010; Shen et al. 2014; J.
387 Li et al. 2017; An et al. 2020; Gómez-Zambrano et al. 2018), and height or size of the plant
388 (CLAVATA, GH9C2, ELF8, NSL1, TUA6) (Glass et al. 2015; Markakis et al. 2012; Noutoshi et al.
389 2006; Fukunaga et al. 2017; Singh et al. 2021; Fal et al. 2017; He 2004; Hoson et al. 2014; Xiong
390 et al. 2013; Whitewoods et al. 2020).

391 Genes under selection include some genes associated with increased sensitivity to
392 bacteria and fungi (JAM3, ABCG16, NMT1, AT5G05790) (Kapos et al. 2015; Kessler et al. 2010;
393 Swain, Singh, and Nandi 2015; Sasaki-Sekimoto et al. 2013; Ji et al. 2014) as well as
394 WRKY70, which is central to immunity in *A. thaliana* (*Noh et al. 2021; H. Liu et al. 2021; S. Chen*
395 *et al. 2021*). These may indicate the importance of rapid evolutionary response to the new
396 enemies and symbionts colonizing plants encounter upon their arrival to volcanic archipelagos.

397 Finally, we assessed the expansion and contraction of gene families in the *Scalesia*
398 genome, finding a total of 37 significantly contracted families and 26 significantly expanded
399 families (Figure 5B). GO enrichment testing of the expanded families uncovered significantly
400 enriched functions associated with vascularization (secondary cell wall biogenesis, shoot
401 system development, negative regulation of organ growth, xylem vessel member cell
402 differentiation, protoxylem development), likely associated with plant growth in *Scalesia*
403 (Fernández-Mazuecos et al. 2020). We also find evidence of evolutionary responses to aridity
404 and changes in osmotic pressure in significantly expanded families (regulation of stomatal
405 closure, response to water deprivation, response to osmotic stress, water homeostasis), similar
406 to the genes under selection (Figure 5B). Interestingly, we detect contraction in gene families
407 with GO terms associated with tree habits (shoot system development, regulation of organ
408 growth, regulation of root development, xylem vessel member cell differentiation,
409 gravitropism), adaptation to arid environments (water deprivation, stomatal closure, regulation
410 to osmotic stress) and cold tolerance (cellular response to cold; Supplementary Information;
411 Supplementary Tables 08-13). While this may seem contradictory, it suggests that different
412 families have redundant functions, and the expansion of a family may lead to redundancy in
413 another family and consequent gene loss through pseudo-gene formation.

414 Conclusions

415 In this study, we were able to elucidate patterns of genome evolution in the critically
416 endangered Darwin’s giant daisy tree (*Scalesia atractyloides*) by attaining a chromosome-

417 resolved genome and by subsequently identifying two ancient genomes underlying its polyploid
418 state. We found that both subgenomes retain a relatively similar number of genes as well as
419 other genetic features, such as pseudogenes and transposable elements, which lead us to
420 speculate on the role of insular evolution underlying these changes. Moreover, we uncovered
421 the role of inversions in gene accumulation, suggesting these have played an important role in
422 the maintenance of genes in subgenomes, and found a relatively unique pattern of transposable
423 element accumulation within flowering of plants. Ultimately, expanded gene families and genes
424 under positive selection indicate the first and solid evidence for genomic island syndrome in a
425 plant, revealing an underlying genomic basis of the outstanding phenotypic variation in *Scalesia*.

426 Methods

427 Plant material, flow cytometry, DNA extraction, library preparation and sequencing

428 Tissues used for the *de-novo* genome assembly and annotation were sampled from
429 living *Scalesia atractyloides* plant P2000-5406/C2834 cultivated in the greenhouse of the
430 University of Copenhagen Botanical Garden collections. This plant was originally germinated
431 from a seed collected from Santiago Island. Fresh tissue was collected and flash-frozen in dry ice
432 or liquid nitrogen and then stored at -80C for later use.

433 To assist with sequencing coverage strategy and to inform genome assembly, we
434 obtained estimates of genome size using flow cytometry following (Galbraith et al. 1983).
435 Briefly, 50 mg of freshly collected leaves from the sample material and from the reference
436 standard (*Solanum lycopersicum* ‘Stupické’; 2C = 1.96 pg; (Jaroslav Dolezel, Sgorbati, and
437 Lucretti 1992) were chopped with a razor blade in a Petri dish containing 1 ml of Woody Plant
438 Buffer (Loureiro et al. 2007). The nuclear suspension was filtered through a 30- μ m nylon filter,
439 and nuclei were stained with 50 mg ml⁻¹ propidium iodide (PI) (Fluka, Buchs, Switzerland).
440 Fifty mg ml⁻¹ of RNase (Sigma, St Louis, MO, USA) was added to the nuclear suspension to
441 prevent staining of double-stranded RNA. After a 5 minute incubation period, samples were
442 analysed in a Sysmex CyFlow Space flow cytometer (532 nm green solid-state laser, operating at
443 30 mW). At least 1,300 particles in G1 peaks were acquired using the FloMax software v2.4d
444 (Jan Suda et al. 2007). The average coefficient of variation for the G1 peak was below 5% (mean
445 CV value = 2.72%). The holoploid genome size in mass units (2C in pg; *sensu* (Greilhuber et al.
446 2005) was obtained as follows: sample 2C nuclear DNA content (pg) = (sample G1 peak mean /
447 reference standard G1 peak mean) * genome size of the reference standard. Conversion into
448 base-pair numbers was performed using the factor: 1 pg = 0.978 Gbp (J. Dolezel et al. 2003).
449 Three replicates were performed on two different days, to account for instrumental artefacts.

450 The commercial provider Dovetail Genomics extracted and purified high-molecular-
451 weight DNA from flash-frozen leaf tissue using the CTAB protocol, and the concentration of DNA
452 was measured by Qubit. For long-read sequencing, they constructed a PacBio SMRTbell library
453 (~20kb) using the SMRTbell Template Prep Kit 1.0 (PacBio, CA, USA) following the
454 manufacturer recommended protocol. This library was bound to polymerase using the Sequel
455 Binding Kit 2.0 (PacBio) and loaded onto the PacBio Sequel sequencing machine using the
456 MagBeadKit v2 (PacBio). Sequencing was performed on the PacBio Sequel SMRT cell, using
457 Instrument Control Software v5.0.0.6235, Primary analysis software v5.0.0.6236, and SMRT
458 Link Version 5.0.0.6792. PacBio sequencing yielded 41,322,824 reads, resulting in a total of 197-
459 fold coverage of the nuclear genome. For contiguity ligation, they prepared two Chicago
460 libraries as described in (Putnam et al. 2016). Briefly, for each Dovetail Omni-C library,

461 chromatic is fixed in place with formaldehyde in the nucleus and then extracted. Fixed
462 chromatin is digested with DNase I, and chromatin ends are repaired and ligated to a
463 biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After
464 proximity ligation, crosslinks are reversed and the DNA is purified. Purified DNA is then treated
465 to remove biotin that was not internal to ligated fragments. Sequencing libraries were
466 generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing
467 fragments were isolated using streptavidin beads before PCR enrichment of each library. These
468 libraries were then sequenced on an Illumina HiSeq 2500 instrument, producing a total of
469 1,463,389,090 sequencing reads.

470 To obtain RNA transcript sequences for annotation of the genome, we extracted RNA
471 from five tissues (root, stem, young leaf, old leaf, and floral head) of *S. attractyloides* plant
472 P2000-5406/C2834 using a Spectrum Plant Total RNA Kit (Sigma, USA) with on-column DNA
473 digestion following the manufacturer's protocol. RNA extracts from all five tissues were pooled.
474 mRNA was enriched using oligo (dT) beads, and the first strand cDNA was synthesized using the
475 Clontech SMARTer PCR cDNA Synthesis Kit, followed by first-strand synthesis with
476 SMARTScribe™ Reverse Transcriptase. After cDNA amplification, a portion of the product was
477 used directly as a non-size selected SMRTbell library. In parallel, the rest of amplification was
478 first selected using either BluePippin or SageELF, and then used to construct a size-selected
479 SMRTbell library after size fractionation. DNA damage and ends were then repaired, followed by
480 hairpin adaptor ligation. Finally, sequencing primers and polymerase were annealed to
481 SMRTbell templates, and IsoSeq isoform sequencing was performed by Novogene Europe
482 (Cambridge, UK) using a PacBio Sequel II instrument, yielding 223,051,882 HiFi reads.

483

484 **Genome assembly and annotation**

485 An overview of the bioinformatic methods is provided in [IOSES GITHUBPAGE](#). We
486 assembled the genome using *wtdbg2* (Ruan and Li 2020), specifying a genome size of 3.7 Gbp,
487 PacBio Sequel reads, and minimum read length of 5,000. The *wtdbg2* assembly consisted of
488 contigs with 3.62 Gbp total length. This assembly was then assessed for contamination using
489 *Blobtools* v1.1.1 (Laetsch and Blaxter 2017a) against the NT database, detecting a removing a
490 fraction of the scaffolds. This filtered assembly was used as input to and *purge_dups* v1.1.2,
491 which removed duplicates based on sequence similarity and read depth (Guan et al. 2020),
492 reducing the assembly length was reduced to 3.22 Gbp. This assembly and the Dovetail OmniC
493 library reads were used as input data for *HiRise* by aligning the Chicago library sequences to the
494 input assembly. After aligning the reads on the reference genome using *bwa*, *HiRise* produces a
495 likelihood model for genomic distance between read pairs, and the model was used to identify
496 misjoints, prospective joints, and make joins. After *HiRise* scaffolding, the N_{50} increased to 16,
497 and the N_{90} to 31, corresponding to all but the three smallest chromosomes ($n=34$), while the
498 LN_{50} was 94.2 Mbp and the LN_{90} was 81.66 Mbp. The largest scaffold was 116.23 Mbp. In total,
499 *HiRise* scaffolding joined 1,329 scaffolds (suppl, assemblathon script). We used the
500 Assemblathon 2 script (<https://github.com/ucdavis-bioinformatics/assemblathon2-analysis>)
501 (Bradnam et al. 2013) to assess assembly quality.

502 To annotate genes, we first masked repeats and low complexity DNA using
503 RepeatMasker v4.1.1 (Smit, Hubley, and Green 2013) using the 'Asteraceae' repeat database
504 with Repbase database. After this first round, we ran RepeatModeler v2.0.1 (Flynn et al. 2020)
505 on the masked genome to obtain a database of novel elements (*denovo*). This database was
506 subsequently used as input to *RepeatMasker* for a second round of masking the genome. To find
507 gene models, we first assembled a transcriptome using PacBio HiFi data and following the

508 IsoSeq3 pipeline (Pacific Biosciences). Processing of the RNA data involved clipping of
509 sequencing barcodes (lima v2.0.0), removal of poly(A) tails and artificial concatemers (Isoseq3
510 refine v3.4.0), clustering of isoforms (Isoseq3 cluster v3.4.0), alignment of the reads to the
511 reference genome using (pbmm2 align v1.4.0), characterization and filtering of transcripts
512 (*SQANTI3* v1.0.0) (Tardaguila et al. 2018). Genome annotation was carried out using the
513 *MAKER2* pipeline v2.31.9 (Holt and Yandell 2011; Moore et al. 2008), using a combination of *ab-*
514 *initio* and homology-based gene predictions (using Asteraceae protein sets). Since no training
515 gene models were available for *Scalesia atractyloides*, we used *CEGMA* (Parra, Bradnam, and
516 Korf 2007) to train the *ab-initio* gene prediction software *SNAP* (Korf 2004). In addition to the
517 *ab-initio* features, we used the PacBio-based transcriptome as a training set for the gene
518 predictor *AUGUSTUS* (Keller et al. 2011), and as direct RNA evidence to *MAKER2*. Finally, when
519 running *MAKER2* we specified, model_org=simle, softmask=1, augustus_species=arabidopsis
520 and specifying *snapphmm* to training of *SNAP*. To assess the quality of the gene models we
521 started using BUSCO and the viridiplantae odb v10 set (Simão et al. 2015; Waterhouse et al.
522 2018; Seppey, Manni, and Zdobnov 2019).

523

524 **Demographic reconstruction using PSMC**

525 To complement the *S. atractyloides* genome, we generated shotgun genomic data from
526 DNA extracts of specimens of *S. helleri* B. L. Rob. and *S. stewartii* Riley, as well as the outgroup
527 species *Pappobolus hypargyreus* and *P. juncosae*. Briefly, the *S. helleri* and *S. stewartii*
528 specimens were extracted with a Qiagen DNeasy 96 Plant Kit. The *P. hypargyreus* and *P.*
529 *juncosae* extracts were previously reported (Fernández-Mazuecos et al 2020). DNA extracts for
530 these four specimens were sent to the commercial provider Novogene for dsDNA library
531 preparation, and they were sequenced on the Illumina NovaSeq platform in 150-bp PE mode.
532 For these sequence data, we used *FastQC* v0.11.8 to check for quality of raw reads (Andrews
533 2017), identified adapters using *AdapterRemoval* v2.3.1, and removed them using *Trimmomatic*
534 v0.39 (Schubert, Lindgreen, and Orlando 2016; Bolger, Lohse, and Usadel 2014). These
535 sequences were then aligned to the *S. atractyloides* genome using the *mem* algorithm of *bwa* (H.
536 Li and Durbin 2009), and reads with a mapping quality below 30 were removed, resulting in a
537 final high-quality sequencing depth of about ~15x. Alignments were then processed and
538 analysed using *PSMC* (Heng Li and Durbin 2011). Specifically, this processing involved calling
539 variants using the *bcftools mpileup* and *call* algorithms, considering base and mapping qualities
540 above 30 and read depths above 5 (Danecek et al. 2021), and posterior processing of the files
541 using *fq2psmcf*. For the *PSMC* run we specified a maximum of 25 iterations, initial theta ratio
542 of 5, bootstrap, and a pattern of "4+25*2+4+6". To plot files we used the util *psmc_plot.pl*
543 specifying a generation time of 3 years and a mutation rate of $6e^{-9}$, and constrained the y- and x-
544 axes to 50 and 20,000,000, respectively.

545

546 **Determination of subgenomes, and testing for subgenome dominance**

547 We reasoned that homologous chromosomes would share conserved orthologue sets
548 (COS). We used the Compositae-COS as baits (available through
549 github.com/Smithsonian/Compositae-COS-workflow/raw/master/COS_probes_phyluce.fasta
550 (Mandel et al. 2014), running *phyluce* to mine for COS in the genome assembly (Faircloth 2016;
551 Faircloth et al. 2012). This pipeline, however, is designed for single-copy COS, and we modified
552 the python script to provide duplicates. We then constructed a matrix of COS-assignment using
553 double-copy COS (Supplementary Information; Supplementary Table 03).

554 Duplicated-COS provided a solid determination of chromosome pairings but did not
555 reveal which member of the pair belongs to either subgenome (subgenome A or B *hereafter*). To
556 distinguish this, we analysed the *k*-mer spectrum (Session et al. 2016). We hypothesized that
557 given a period of separation between the two subgenomes, they have accumulated different
558 repeat content and transposable elements. To do so, we ran the software *Jellyfish* (Marçais and
559 Kingsford 2011) for each chromosome independently, thus obtaining a chromosome-by-
560 chromosome frequency of 13-mers. To ensure we obtained only repeats, we selected 13-mers
561 represented only >100 times in each chromosome, and kept only *k*-mers which were twice
562 represented on a member of each pair. Using R, we computed a distance matrix and a
563 hierarchical clustering, which neatly separated members of each pair into two groups
564 (Supplementary Information; Supplementary Figure 03).

565 To confirm the quality of subgenome assignment we took two independent approaches.
566 First, we did a circos plot using the masked regions of the genome. To do the circos plot we
567 aligned the masked subgenomes to each other using mummer (Delcher et al. 1999; Kurtz et al.
568 2004), and plotted the circos using the ‘Circos, round is beautiful’ software (Krzywinski et al.
569 2009). Second, we studied transposable element representation in each subgenome benefiting
570 from the transposable element identification done by *RepeatMasker*. In specific, we obtained
571 the list of different annotated transposable elements from *RepeatMasker* (e.g. RTE-BovB, LINE-
572 L1, LINE-L2, Helitron, PIF-Harbinger Gypsy, Copia, CRE), and separated the families within
573 these groups. For each family, we counted the number of elements present on each subgenome,
574 and plotted all the families using raincloud plots (Allen et al. 2019). To visualize genes and
575 transposable elements along chromosomes we used the R package Ideogram (Hao et al. 2020).
576 After finding out each subgenome, we ran BUSCO separately for each subgenome as a way of
577 understanding subgenome specific gene loss (*Viridiplantae odb10* as specified above).

578

579 **Evolutionary history of the *Scalesia atrectyloides* subgenomes and comparative genomics**

580 We searched the literature and NCBI for chromosome-level assemblies of the Asteraceae
581 (05 / February / 2021), downloading the genomes of the sunflower (*Helianthus annuus*
582 (Badouin et al. 2017)), the Canada fleabane (*Conyza canadensis*; (Laforest et al. 2020)), the
583 ‘mile-a-minute’ weed (*Mikania micrantha*; (B. Liu et al. 2020)), and the lettuce (*Lactuca sativa*;
584 (Reyes-Chin-Wo et al. 2017)). We downloaded the *Arabidopsis thaliana* genome from
585 Arabidopsis.org.

586 To obtain sets of orthologous genes, we ran OrthoFinder (Emms and Kelly 2015) on
587 predicted amino acid sequences (faa) and coding sequences (cds). Before running this software,
588 we selected only the longest isoforms of both files, and removed sequences with stop codons. On
589 the amino acids file we removed sequences with lengths below 30 bp using kinfin’s
590 filter_fastas_before_clustering.py script (Laetsch and Blaxter 2017b). We ran OrthoFinder on
591 various combinations of the genomes, including: 1) All Asteraceae, with subgenomes separated,
592 (*S. atrectyloides* subgenomeA, *S. atrectyloides* subgenomeB, *C. canadensis*, *H. annuus*, *L. sativa*,
593 *M. micrantha*); 2) All Asteraceae, and the *Scalesia* genome (*S. atrectyloides* (complete), *C.*
594 *canadensis*, *H. annuus*, *L. sativa*, *M. micrantha*); 3) *A. thaliana* and subgenomes (*S. atrectyloides*
595 subgenomeA, *S. atrectyloides* subgenomeB, *A. thaliana*). A representation of run 2 and its
596 processed results using an upset plot (Conway, Lex, and Gehlenborg 2017).

597 To obtain a tree of the two subgenomes we ran OrthoFinder with the two subgenomes
598 and obtained the tree of the single-copy orthologs. This tree was made ultrametric using r8s
599 (Sanderson 2003). To date the tree, we converted branch lengths to time estimates using a
600 calibration point of 6.14 Mya between *H. annuus* and *S. atrectyloides* following recent literature

601 (Mandel et al. 2019). To date the divergence of the subgenomes we followed the approach of
602 (Session et al. 2016; Lovell et al. 2021; Mitros et al. 2020). Briefly, this approach has a simple
603 assumption: before the speciation event (which separates the ancestral lineages) and after the
604 polyploidization event (which brings the ancestral genomes together), the accumulation of
605 transposable elements will be similar on both subgenomes. Transposable element families
606 which are equally represented on both subgenomes will therefore represent the pre-speciation
607 and post-allopolyploidization period. We focused on long-terminal repeats (LTRs) given their
608 prevalence along the genome. We used LTRharvest to identify LTR elements (Ellinghaus, Kurtz,
609 and Willhoeft 2008), and LTRdisgest to process these (i.e. annotating features such as genes
610 inside LTRs). To find these features we downloaded various PFAM domains provided in
611 (Steinbiss et al. 2009), and downloaded “Gypsy” and “Copia” domains from the PFAM online
612 database. We converted the domains to HMMs using hmmconvert (Eddy 1992), and added
613 HMMs from the Gypsy Database (Llorens et al. 2011). The identification and annotation of LTRs
614 was done for the *S. atractyloides* and *H. annuus* genomes. Instead of using the whole LTR-
615 element (i.e. whole transposable element including repeated regions and genes insides) we
616 used only the LTR-region (long terminal repeat), and ran OrthoFinder to group closely related
617 LTRs. We processed the orthofinder data selecting orthogroups which were in equal
618 representation on both subgenomes, and that were also present in *Helianthus*. After this, we
619 aligned the selected orthogroups using mafft, and cleaned poorly aligned regions using Gblocks
620 (Castresana 2000; Talavera and Castresana 2007), with not stringent options (i.e. “allow
621 smaller final blocks”, “allow gap positions within the final blocks”, and “allow less strict flanking
622 regions”). After this, we removed sequences with more than 50% missing data, and re-checked
623 whether numbers of TEs were still balanced between subgenomes. We then re-aligned the data
624 using mafft and inferred a tree for each ortholog. We kept only orthogroups where the *S.*
625 *atractyloides* sequences were monophyletic, and where both subgenomes were non-
626 monophyletic. For the final set of orthogroups passing all this filtering, we calculated pairwise
627 Jukes Cantor distance between each *S. atractyloides* LTR-region; and between each *S.*
628 *atractyloides* and *H. annuus*. The Jukes Cantor distances were plotted in R and we analysed the
629 overall frequency and converted it to million of years distance by a simple three rule with the
630 *Helianthus* divergence with *Scalesia* of 6.14 Mya (Supplementary Information; Supplementary
631 Figure 04).

632

633 **Signatures of selection and expanded gene regions**

634 Using the *Scalesia* genome together with the remaining Asteraceae genomes we ran
635 CAFE analyses (De Bie et al. 2006; Mendes et al. 2020) to estimate significant gene family
636 expansions and contractions. Briefly, we did an all-by-all BLAST to identify orthologues in the
637 dataset and estimated significantly expanded and contracted families using CAFÉ. To interpret
638 the data we relied on Gene Ontology Annotation. We obtained GOs for the annotated *Scalesia*
639 genes by means of two complementary approaches: 1) by using the Interproscan command-line
640 version (Jones et al. 2014), using the NCBI’s Conserved Domains Database (CDD), Prediction of
641 Coiled Coil Regions in Proteins (COILS), Protein Information Resource (PIRSF), PRINTS, PFAM,
642 ProDom, ProSitePatterns and ProSiteProfiles, the Structure–Function Linkage Database (SFLD),
643 Simple Modular Architecture Research Tool (SMART), SUPERFAMILY, and TIGRFAMs
644 databases; 2) by extracting the curated Swiss Prot database from UniProt (Viridiplantae). We
645 blasted the *Scalesia* genes to this database and kept hits with an e-value below 1e-10. We then
646 extracted the GOs from each gene from the database and assigned these to *Scalesia*’s
647 correspondent orthologs (e-value below 1e-10). Genes belonging to significantly expanded gene

648 families in the *S. attractyloides* genome were analysed using a GO enrichment analysis. To do so,
649 we used the TopGO package using the 'elim' algorithm which takes GO hierarchy into account
650 (Alexa, Rahnenführer, and Others 2010; Alexa and Rahnenführer 2009), this were then plotted
651 with *REVIGO* (Supek et al. 2011).

652 To test which genes are under positive selection in *S. attractyloides* genome, we retrieved
653 the orthogroups from all Asteraceae, and aligned the cds from each orthogroup using prank
654 (Löytynoja 2014). Considering the divergence in the genomes, as well as evidence for fast
655 evolution in Asteraceae genomes (including this paper), we ran zorro (Wu, Chatterji, and Eisen
656 2012), to assess alignments. Zorro scores each alignment position with a score between 0-10,
657 and we selected only alignments with an average score position of 5 or greater. For each of
658 these, we inferred a tree using iqtree and ran HyPhy using its aBSREL selection test (Smith et al.
659 2015; Pond, Frost, and Muse 2005). To summarize these results we: 1) ran a GO enrichment (as
660 specified above) and plotted results using *REVIGO*; 2) identified the *Arabidopsis* ortholog to
661 each of the *Scalesia* gene under selection using BLAST, and analysed the *Arabidopsis* literature
662 for that particular gene (Supplementary Information; Supplementary Table 07); 3) we ran a
663 STRING analysis using the *Arabidopsis* ortholog (Szklarczyk et al. 2015), exploring the potential
664 protein-protein interactions among genes under selection. Interaction scores of edges were
665 calculated based on the parameters Experiments, Co-expression, Neighborhood, Gene fusion
666 and Co-occurrence. Edges with interaction score higher than 0.400 were kept in the network.
667 After excluding genes with no physical connection, the STRING network had 627 nodes with
668 470 edges (PPI enrichment p-value <0.001). To simplify the densely connected network into
669 potential biologically functional clusters, we used the distance matrix obtained from the STRING
670 global scores as the input to perform a k-Means clustering analysis (number of clusters = 6). 4
671 out of the 6 clusters are enriched for biological processes related GO terms. Cluster 1 (red
672 bubbles) were enriched for the GO term metabolic processes, cluster 3 (lime green bubbles) for
673 histone modifications and chromosome organization, cluster 4 (green bubbles) for response to
674 light, and cluster 6 (purple bubbles) for ribosomal large subunit biogenesis and RNA-
675 processing.

676

677 **Data Availability**

678 We are in the process of submitting raw reads to ENA and the reference genome to
679 Dryad.

680

681 **Author contributions**

682 J.Ce. designed the experiment, processed and analysed the data and wrote the manuscript. B.P.,
683 J.M.L.G., A. R.-C., J.Ca., S.B., J.V., S.L., D.M. helped analysed the data. J.L. was responsible for flow
684 cytometry analyses. C.K., L. S.-B. helped retrieving DNA/RNA, N. W., M.N., P.J.D.,G.R.-T. obtained
685 permits. M.F.-M., P.V., R.M. obtained the outgroups. G.P., A.S., N.S., N.R.N., O.S., M.T.P.G., J.H. L.-M.,
686 L.R. contributed with senior expertise in data generation and interpretation, Alexander
687 Suh22,23. M.D.M. obtained funding, supervised J.C. and wrote the manuscript. All the authors
688 revised and approved the manuscript.

689

690 **Acknowledgements**

691 JC is grateful to Simen R. Sandve for fruitful discussion, Martin LaForest for sharing
692 genome annotations for his organism, and Henning Adersen for botanical expertise and
693 logistical support in utilizing the University of Copenhagen botanical collections. Jennifer
694 Mandel kindly shared the Asteraceae COS. The collection and photography of specimens, and

695 the preparation of this manuscript benefited enormously from the cooperative assistance of the
696 personnel of the Charles Darwin Foundation Research Station, who made arrangements for
697 collecting trips, arranged laboratory space, and offered encouragement and support throughout
698 the project. *Scalesia* specimens were collected under the Galápagos National Park research
699 permit number PC-001/98 PNG and MAAE-DBI-CM-2021-0213. This publication is contribution
700 number 2426 of the Charles Darwin Foundation for the Galápagos Islands. This work was
701 supported by the Norwegian Research Council via project number 287327 awarded to MDM.

702

703 References

- 704 Alexa, Adrian, and Jörg Rahnenführer. 2009. "Gene Set Enrichment Analysis with topGO."
705 *Bioconductor Improv*27: 1–26.
- 706 Alexa, Adrian, Jorg Rahnenfuhrer, and Others. 2010. "topGO: Enrichment Analysis for Gene
707 Ontology." *R Package Version 2* (0): 2010.
- 708 Alger, Elizabeth I., and Patrick P. Edger. 2020. "One Subgenome to Rule Them All: Underlying
709 Mechanisms of Subgenome Dominance." *Current Opinion in Plant Biology*54 (April): 108–
710 13.
- 711 Allen, Micah, Davide Poggiali, Kirstie Whitaker, Tom Rhys Marshall, and Rogier A. Kievit. 2019.
712 "Raincloud Plots: A Multi-Platform Tool for Robust Data Visualization." *Wellcome Open*
713 *Research* 4 (April): 63.
- 714 Andrews, S. 2017. "FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010."
715 An, Zengxuan, Liufan Yin, Yuhao Liu, Maolin Peng, Wen-Hui Shen, and Aiwu Dong. 2020. "The
716 Histone Methylation Readers MRG1/MRG2 and the Histone Chaperones NRP1/NRP2
717 Associate in Fine-Tuning Arabidopsis Flowering Time." *The Plant Journal: For Cell and*
718 *Molecular Biology*103 (3): 1010–24.
- 719 Badouin, Hélène, Jérôme Gouzy, Christopher J. Grassa, Florent Murat, S. Evan Staton, Ludovic
720 Cottret, Christine Lelandais-Brière, et al. 2017. "The Sunflower Genome Provides Insights
721 into Oil Metabolism, Flowering and Asterid Evolution." *Nature* 546 (7656): 148–52.
- 722 Baeckens, Simon, and Raoul Van Damme. 2020. "The Island Syndrome." *Current Biology: CB*30
723 (8): R338–39.
- 724 Baldwin, B. G., and M. J. Sanderson. 1998. "Age and Rate of Diversification of the Hawaiian
725 Silversword Alliance (Compositae)." *Proceedings of the National Academy of Sciences of*
726 *the United States of America* 95 (16): 9402–6.
- 727 Basset, Gilles J. C., Stéphane Ravel, Eoin P. Quinlivan, Ruth White, James J. Giovannoni, Fabrice
728 Rébeillé, Brian P. Nichols, et al. 2004. "Folate Synthesis in Plants: The Last Step of the P-
729 Aminobenzoate Branch Is Catalyzed by a Plastidial Aminodeoxychorismate Lyase." *The*
730 *Plant Journal: For Cell and Molecular Biology*40 (4): 453–61.
- 731 Berckmans, Barbara, Tim Lammens, Hilde Van Den Daele, Zoltan Magyar, Laszlo Bögre, and
732 Lieven De Veylder. 2011. "Light-Dependent Regulation of DEL1 Is Determined by the
733 Antagonistic Action of E2Fb and E2Fc." *Plant Physiology*157 (3): 1440–51.
- 734 Besse, Pascale, Denis DaSilva, Laurence Humeau, Joyce Govinden-Soulange, Ameenah Gurib-
735 Fakim, and Hippolyte Kodja. 2003. "A Genetic Diversity Study of Endangered *Psiadia*
736 Species Endemic from Mauritius Island Using PCR Markers." *Biochemical Systematics and*
737 *Ecology*31 (12): 1427–45.
- 738 Bird, Kevin A., Chad E. Niederhuth, Shujun Ou, Malia Gehan, J. Chris Pires, Zhiyong Xiong, Robert
739 VanBuren, and Patrick P. Edger. 2021. "Replaying the Evolutionary Tape to Investigate
740 Subgenome Dominance in Allopolyploid Brassica *Napus*." *The New Phytologist*230 (1):
741 354–71.
- 742 Bird, Kevin A., Robert VanBuren, Joshua R. Puzey, and Patrick P. Edger. 2018. "The Causes and
743 Consequences of Subgenome Dominance in Hybrids and Recent Polyploids." *The New*
744 *Phytologist*220 (1): 87–93.
- 745 Bittner, Andras, Bettina Hause, and Margarete Baier. 2021. "Cold-Priming Causes Oxylipin

- 746 Dampening during the Early Cold and Light Response of Arabidopsis Thaliana." *Journal of*
747 *Experimental Botany*, June. <https://doi.org/10.1093/jxb/erab314>.
- 748 Blaschke, Jeremy D., and Roger W. Sanders. 2009. "PRELIMINARY INSIGHTS INTO THE
749 PHYLOGENY AND SPECIATION OF SCALEZIA (ASTERACEAE), GALÁPAGOS ISLANDS."
750 *Journal of the Botanical Research Institute of Texas* 3 (1): 177–91.
- 751 Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer
752 for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.
- 753 Bradnam, Keith R., Joseph N. Fass, Anton Alexandrov, Paul Baranay, Michael Bechner, Inanç
754 Birol, Sébastien Boisvert, et al. 2013. "Assemblathon 2: Evaluating de Novo Methods of
755 Genome Assembly in Three Vertebrate Species." *GigaScience* 2 (1): 10.
- 756 Burns, Kevin C. 2019. *Evolution in Isolation: The Search for an Island Syndrome in Plants*.
757 Cambridge University Press.
- 758 Castells, Enric, Jean Molinier, Giovanna Benvenuto, Clara Bourbousse, Gerald Zabulon, Antoine
759 Zalc, Stefano Cazzaniga, Pascal Genschik, Fredy Barneche, and Chris Bowler. 2011. "The
760 Conserved Factor DE-ETIOLATED 1 Cooperates with CUL4-DDB1DDB2 to Maintain
761 Genome Integrity upon UV Stress." *The EMBO Journal* 30 (6): 1162–72.
- 762 Castresana, J. 2000. "Selection of Conserved Blocks from Multiple Alignments for Their Use in
763 Phylogenetic Analysis." *Molecular Biology and Evolution* 17 (4): 540–52.
- 764 Catchen, J. M., J. S. Conery, and J. H. Postlethwait. 2009. "Automated Identification of Conserved
765 Synteny after Whole-Genome Duplication." *Genome Research*.
766 <https://doi.org/10.1101/gr.090480.108>.
- 767 Chen, Donghong, Anne Molitor, Chunlin Liu, and Wen-Hui Shen. 2010. "The Arabidopsis PRC1-
768 like Ring-Finger Proteins Are Necessary for Repression of Embryonic Traits during
769 Vegetative Growth." *Cell Research* 20 (12): 1332–44.
- 770 Chen, Siyu, Yuli Ding, Hainan Tian, Shucai Wang, and Yuelin Zhang. 2021. "WRKY54 and
771 WRKY70 Positively Regulate SARD1 and CBP60g Expression in Plant Immunity." *Plant*
772 *Signaling & Behavior* 16 (10): 1932142.
- 773 Conway, Jake R., Alexander Lex, and Nils Gehlenborg. 2017. "UpSetR: An R Package for the
774 Visualization of Intersecting Sets and Their Properties." *Bioinformatics* 33 (18): 2938–40.
- 775 Crane, Renee A., Marielle Cardénas Valdez, Nelly Castaneda, Charidan L. Jackson, Ciairra J. Riley,
776 Islam Mostafa, Wenwen Kong, Shweta Chhajed, Sixue Chen, and Judy A. Brusslan. 2019.
777 "Negative Regulation of Age-Related Developmental Leaf Senescence by the IAOx Pathway,
778 PEN1, and PEN3." *Frontiers in Plant Science* 10 (October): 1202.
- 779 Crawford, Daniel J., Timothy K. Lowrey, Gregory J. Anderson, Gabriel Bernardello, Arnaldo
780 Santos-Guerra, and Tod F. Stuessy. 2009. "Genetic Diversity in Asteraceae Endemic to
781 Oceanic Islands: Baker's Law and Polyploidy." *Syst Evol Biogeogr Compos* 139: 151.
- 782 Czarnocka, Weronika, Yosef Fichman, Maciej Bernacki, Elżbieta Różańska, Izabela Sańko-
783 Sawczenko, Ron Mittler, and Stanisław Karpiński. 2020. "FM01 Is Involved in Excess Light
784 Stress-Induced Signal Transduction and Cell Death Signaling." *Cells* 9 (10).
785 <https://doi.org/10.3390/cells9102163>.
- 786 Dal Bosco, Cristina, Lina Lezhneva, Alexander Biehl, Dario Leister, Heinrich Strotmann, Gerd
787 Wanner, and Jorg Meurer. 2004. "Inactivation of the Chloroplast ATP Synthase Gamma
788 Subunit Results in High Non-Photochemical Fluorescence Quenching and Altered Nuclear
789 Gene Expression in Arabidopsis Thaliana." *The Journal of Biological Chemistry* 279 (2):
790 1060–69.
- 791 Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard,
792 Andrew Whitwham, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10
793 (2). <https://doi.org/10.1093/gigascience/giab008>.
- 794 Darwin, B. Y. Charles, Es Q., R. S. F., L. S. F., G. S. F., and Alfred Wallace. n.d. "I. On the Tendency of
795 Species to Form Varieties; And on the Perpetuation of Varieties and Species by Natural
796 Means of Selection." Accessed June 18, 2021.
797 [http://anthro198.pbworks.com/f/Darwin+&+Wallace+\(1858\)+-+On+the+Tendency+of+Species+to+form+Varieties.pdf](http://anthro198.pbworks.com/f/Darwin+&+Wallace+(1858)+-+On+the+Tendency+of+Species+to+form+Varieties.pdf).
- 798 Darwin, Charles. 1859. "On the Origin of Species by Means of Natural Selection, Or, The
799

- 800 Preservation of Favoured Races in the Struggle for Life /.”
801 <https://doi.org/10.5962/bhl.title.68064>.
- 802 De Bie, Tijn, Nello Cristianini, Jeffery P. Demuth, and Matthew W. Hahn. 2006. “CAFE: A
803 Computational Tool for the Study of Gene Family Evolution.” *Bioinformatics* 22 (10):
804 1269–71.
- 805 Delcher, A. L., S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg. 1999.
806 “Alignment of Whole Genomes.” *Nucleic Acids Research* 27 (11): 2369–76.
- 807 Diop, Seydina I., Oliver Subotic, Alejandro Giraldo-Fonseca, Manuel Waller, Alexander Kirbis,
808 Anna Neubauer, Giacomo Potente, et al. 2020. “A Pseudomolecule-Scale Genome Assembly
809 of the Liverwort *Marchantia Polymorpha*.” *The Plant Journal: For Cell and Molecular*
810 *Biology* 101 (6): 1378–96.
- 811 Dolezel, Jaroslav, Sergio Sgorbati, and Sergio Lucretti. 1992. “Comparison of Three DNA
812 Fluorochromes for Flow Cytometric Estimation of Nuclear DNA Content in Plants.”
813 *Physiologia Plantarum*. <https://doi.org/10.1034/j.1399-3054.1992.850410.x>.
- 814 Dolezel, J., J. Bartos, H. Voglmayr, and J. Greilhuber. 2003. “Nuclear DNA Content and Genome
815 Size of Trout and Human.” *Cytometry. Part A: The Journal of the International Society for*
816 *Analytical Cytology*.
- 817 Douglas, Gavin M., Gesseca Gos, Kim A. Steige, Adriana Salcedo, Karl Holm, Emily B. Josephs,
818 Ramesh Arunkumar, et al. 2015. “Hybrid Origins and the Earliest Stages of Diploidization in
819 the Highly Successful Recent Polyploid *Capsella Bursa-Pastoris*.” *Proceedings of the*
820 *National Academy of Sciences of the United States of America* 112 (9): 2806–11.
- 821 Eastwood, A., M. Gibby, and Q. C. B. Cronk. 2004. “Evolution of St Helena Arborescent Asteraceae
822 (Asteraceae): Relationships of the Genera *Commidendrum* and *Melanodendron*.” *Botanical*
823 *Journal of the Linnean Society. Linnean Society of London* 144 (1): 69–83.
- 824 Eddy, Sean. 1992. “HMMER User’s Guide.” *Department of Genetics, Washington University*
825 *School of Medicine* 2 (1): 13.
- 826 Edger, Patrick P., Michael R. McKain, Kevin A. Bird, and Robert VanBuren. 2018. “Subgenome
827 Assignment in Allopolyploids: Challenges and Future Directions.” *Current Opinion in Plant*
828 *Biology* 42 (April): 76–80.
- 829 Eliasson, Uno. 1974. “Studies in Galapagos Plants. XIV. The Genus *Scalesia* Arn.” *Opera Botanica*
830 36: 1–117.
- 831 Eliasson, Uno, and Others. 1974. “Studies in Galápagos Plants. XIV. The Genus *Scalesia* Arn.”
832 [https://pascal-](https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7537000304)
833 [francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7537000304](https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7537000304).
- 834 Eliasson, U., and Eliasson U. 1974. “Studies in Galapagos Plants. Xiv. The Genus *Scalesia* Arn.”
835 [https://pascal-](https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7537000304)
836 [francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7537000304](https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7537000304).
- 837 Ellinghaus, David, Stefan Kurtz, and Ute Willhoeft. 2008. “LTRharvest, an Efficient and Flexible
838 Software for de Novo Detection of LTR Retrotransposons.” *BMC Bioinformatics* 9 (January):
839 18.
- 840 Emerson, Brent C. 2008. “Speciation on Islands: What Are We Learning?” *Biological Journal of*
841 *the Linnean Society. Linnean Society of London* 95 (1): 47–52.
- 842 Emms, David M., and Steven Kelly. 2015. “OrthoFinder: Solving Fundamental Biases in Whole
843 Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy.” *Genome*
844 *Biology* 16 (August): 157.
- 845 Endo, Motomu, Daiki Kudo, Tomoko Koto, Hanako Shimizu, and Takashi Araki. 2014. “Light-
846 Dependent Destabilization of PHL in *Arabidopsis*.” *Plant Signaling & Behavior* 9 (3):
847 e28118.
- 848 Endo, Motomu, Yoshiyasu Tanigawa, Tadashi Murakami, Takashi Araki, and Akira Nagatani.
849 2013. “PHYTOCHROME-DEPENDENT LATE-FLOWERING Accelerates Flowering through
850 Physical Interactions with Phytochrome B and CONSTANS.” *Proceedings of the National*
851 *Academy of Sciences of the United States of America* 110 (44): 18017–22.
- 852 Faircloth, Brant C. 2016. “PHYLUCe Is a Software Package for the Analysis of Conserved
853 Genomic Loci.” *Bioinformatics* 32 (5): 786–88.

- 854 Faircloth, Brant C., John E. McCormack, Nicholas G. Crawford, Michael G. Harvey, Robb T.
855 Brumfield, and Travis C. Glenn. 2012. "Ultraconserved Elements Anchor Thousands of
856 Genetic Markers Spanning Multiple Evolutionary Timescales." *Systematic Biology* 61 (5):
857 717–26.
- 858 Fal, Kateryna, Mengying Liu, Assem Duisebekova, Yassin Refahi, Elizabeth S. Haswell, and
859 Olivier Hamant. 2017. "Phyllotactic Regularity Requires the Paf1 Complex in Arabidopsis."
860 *Development* <https://doi.org/10.1242/dev.154369>.
- 861 Fernández-Mazuecos, Mario, Pablo Vargas, Ross A. McCauley, David Monjas, Ana Otero, Jaime A.
862 Chaves, Juan Ernesto Guevara Andino, and Gonzalo Rivas-Torres. 2020. "The Radiation of
863 Darwin's Giant Daisies in the Galápagos Islands." *Current Biology: CB* 30 (24): 4989–98.e7.
- 864 Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte,
865 and Arian F. Smit. 2020. "RepeatModeler2 for Automated Genomic Discovery of
866 Transposable Element Families." *Proceedings of the National Academy of Sciences of the
867 United States of America* 117 (17): 9451–57.
- 868 Fox, Ana Romina, Florencia Scochera, Timothée Laloux, Karolina Filik, Hervé Degand, Pierre
869 Morsomme, Karina Alleva, and François Chaumont. 2020. "Plasma Membrane Aquaporins
870 Interact with the Endoplasmic Reticulum Resident VAP27 Proteins at ER-PM Contact Sites
871 and Endocytic Structures." *The New Phytologist* 228 (3): 973–88.
- 872 Freeling, Michael, Michael J. Scanlon, and John E. Fowler. 2015. "Fractionation and
873 Subfunctionalization Following Genome Duplications: Mechanisms That Drive Gene
874 Content and Their Consequences." *Current Opinion in Genetics & Development*
875 <https://doi.org/10.1016/j.gde.2015.11.002>.
- 876 Fukunaga, Satoshi, Miho Sogame, Masaki Hata, Suthitar Singkaravanit-Ogawa, Mariola
877 Piślewska-Bednarek, Mariko Onozawa-Komori, Takumi Nishiuchi, et al. 2017. "Dysfunction
878 of Arabidopsis MACPF Domain Protein Activates Programmed Cell Death via Tryptophan
879 Metabolism in MAMP-Triggered Immunity." *The Plant Journal*.
880 <https://doi.org/10.1111/tpj.13391>.
- 881 Fu, Mengni, Changshun Yuan, Aihua Song, Jun Lu, Xiaojing Wang, and Shulan Sun. 2019.
882 "AtWDS1 Negatively Regulates Age-Dependent and Dark-Induced Leaf Senescence in
883 Arabidopsis." *Plant Science: An International Journal of Experimental Plant Biology* 285
884 (August): 44–54.
- 885 Galbraith, D. W., K. R. Harkins, J. M. Maddox, N. M. Ayres, D. P. Sharma, and E. Firoozabady. 1983.
886 "Rapid Flow Cytometric Analysis of the Cell Cycle in Intact Plant Tissues." *Science* 220
887 (4601): 1049–51.
- 888 Glass, Magdalena, Sarah Barkwill, Faride Unda, and Shawn D. Mansfield. 2015. "Endo-β-1,4-
889 Glucanases Impact Plant Cell Wall Development by Influencing Cellulose Crystallization."
890 *Journal of Integrative Plant Biology*. <https://doi.org/10.1111/jipb.12353>.
- 891 Gómez-Zambrano, Ángeles, Pedro Crevillén, José M. Franco-Zorrilla, Juan A. López, Jordi
892 Moreno-Romero, Pawel Roszak, Juan Santos-González, et al. 2018. "Arabidopsis SWC4
893 Binds DNA and Recruits the SWR1 Complex to Modulate Histone H2A.Z Deposition at Key
894 Regulatory Genes." *Molecular Plant* 11 (6): 815–32.
- 895 Greilhuber, Johann, Jaroslav Dolezel, Martin A. Lysák, and Michael D. Bennett. 2005. "The Origin,
896 Evolution and Proposed Stabilization of the Terms 'Genome Size' and 'C-Value' to Describe
897 Nuclear DNA Contents." *Annals of Botany* 95 (1): 255–60.
- 898 Gruenstaeudl, Michael, Arnoldo Santos-Guerra, and Robert K. Jansen. 2013. "Phylogenetic
899 Analyses of Tolpis Adans. (Asteraceae) Reveal Patterns of Adaptive Radiation, Multiple
900 Colonization and Interspecific Hybridization." *Cladistics: The International Journal of the
901 Willi Hennig Society* 29 (4): 416–34.
- 902 Guan, Dengfeng, Shane A. McCarthy, Jonathan Wood, Kerstin Howe, Yadong Wang, and Richard
903 Durbin. 2020. "Identifying and Removing Haplotypic Duplication in Primary Genome
904 Assemblies." *Bioinformatics* 36 (9): 2896–98.
- 905 Hao, Zhaodong, Dekang Lv, Ying Ge, Jisen Shi, Dolf Weijers, Guangchuan Yu, and Jinhui Chen.
906 2020. "Rldeogram: Drawing SVG Graphics to Visualize and Map Genome-Wide Data on the
907 Idiograms." *PeerJ. Computer Science* 6 (January): e251.

- 908 He, Y. 2004. "PAF1-Complex-Mediated Histone Methylation of FLOWERING LOCUS C Chromatin
909 Is Required for the Vernalization-Responsive, Winter-Annual Habit in Arabidopsis." *Genes
910 & Development* <https://doi.org/10.1101/gad.1244504>.
- 911 Holding, D. 2000. "The Chloroplast and Leaf Developmental Mutant, Pale Cress, Exhibits Light-
912 Conditional Severity and Symptoms Characteristic of Its ABA Deficiency." *Annals of Botany*.
913 <https://doi.org/10.1006/anbo.2000.1263>.
- 914 Holt, Carson, and Mark Yandell. 2011. "MAKER2: An Annotation Pipeline and Genome-Database
915 Management Tool for Second-Generation Genome Projects." *BMC Bioinformatics* 12
916 (December): 491.
- 917 Hoson, T., K. Soga, K. Wakabayashi, T. Hashimoto, I. Karahara, S. Yano, F. Tanigaki, et al. 2014.
918 "Growth Stimulation in Inflorescences of an Arabidopsis Tubulin Mutant under
919 Microgravity Conditions in Space." *Plant Biology* 16 Suppl 1 (January): 91–96.
- 920 Hove, Colette A. ten, Zoltán Bochnanovits, Vera M. A. Jansweijer, Fenne G. Koning, Lidija Berke,
921 Gabino F. Sanchez-Perez, Ben Scheres, and Renze Heidstra. 2011. "Probing the Roles of LRR
922 RLK Genes in Arabidopsis Thaliana Roots Using a Custom T-DNA Insertion Set." *Plant
923 Molecular Biology* 76 (1): 69–83.
- 924 Husbands, Aman Y., Anna H. Benkovics, Fabio T. S. Nogueira, Mukesh Lodha, and Marja C. P.
925 Timmermans. 2016. "The ASYMMETRIC LEAVES Complex Employs Multiple Modes of
926 Regulation to Affect Adaxial-Abaxial Patterning and Leaf Complexity." *The Plant Cell*.
927 <https://doi.org/10.1105/tpc.15.00454>.
- 928 Itow, Syuzo. 1995. "Phytogeography and Ecology of Scalesia (compositae) Endemic to the
929 Galapagos Islands!" *Pacific Science* 49 (1): 17–30.
- 930 Jakoby, Marc J., Doris Falkenhan, Michael T. Mader, Ginger Brininstool, Elisabeth Wischnitzki,
931 Nicole Platz, Andrew Hudson, Martin Hülskamp, John Larkin, and Arp Schnittger. 2008.
932 "Transcriptional Profiling of Mature Arabidopsis Trichomes Reveals That NOECK Encodes
933 the MIXTA-like Transcriptional Regulator MYB106." *Plant Physiology* 148 (3): 1583–1602.
- 934 Ji, Hao, Yanhui Peng, Nicole Meckes, Sara Allen, C. Neal Stewart Jr, and M. Brian Traw. 2014.
935 "ATP-Dependent Binding Cassette Transporter G Family Member 16 Increases Plant
936 Tolerance to Abscisic Acid and Assists in Basal Resistance against Pseudomonas Syringae
937 DC3000." *Plant Physiology* 166 (2): 879–88.
- 938 Jones, Philip, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla,
939 Hamish McWilliam, et al. 2014. "InterProScan 5: Genome-Scale Protein Function
940 Classification." *Bioinformatics* 30 (9): 1236–40.
- 941 Kapos, Paul, Fang Xu, Thierry Meinel, Carmela Giglione, and Xin Li. 2015. "N-Terminal
942 Modifications Contribute to Flowering Time and Immune Response Regulations." *Plant
943 Signaling & Behavior* 10 (10): e1073874.
- 944 Keller, Oliver, Martin Kollmar, Mario Stanke, and Stephan Waack. 2011. "A Novel Hybrid Gene
945 Prediction Method Employing Protein Multiple Sequence Alignments." *Bioinformatics* 27
946 (6): 757–63.
- 947 Kessler, Sharon A., Hiroko Shimosato-Asano, Nana F. Keinath, Samuel E. Wuest, Gwyneth
948 Ingram, Ralph Panstruga, and Ueli Grossniklaus. 2010. "Conserved Molecular Components
949 for Pollen Tube Reception and Fungal Invasion." *Science* 330 (6006): 968–71.
- 950 Kim, Areum, Jilin Chen, Deepa Khare, Jun-Young Jin, Yasuyo Yamaoka, Masayoshi Maeshima,
951 Yunde Zhao, Enrico Martinoia, Jae-Ung Hwang, and Youngsook Lee. 2020. "Non-Intrinsic
952 ATP-Binding Cassette Proteins ABCI19, ABCI20 and ABCI21 Modulate Cytokinin Response
953 at the Endoplasmic Reticulum in Arabidopsis Thaliana." *Plant Cell Reports* 39 (4): 473–87.
- 954 Kim, Joo Y., Su J. Park, Boseung Jang, Che-Hun Jung, Sung J. Ahn, Chang-Hyo Goh, Kyoungwon
955 Cho, Oksoo Han, and Hunseung Kang. 2007. "Functional Characterization of a Glycine-Rich
956 RNA-Binding Protein 2 in Arabidopsis Thaliana under Abiotic Stress Conditions." *The Plant
957 Journal*. <https://doi.org/10.1111/j.1365-313x.2007.03057.x>.
- 958 Kim, S. C., D. J. Crawford, J. Francisco-Ortega, and A. Santos-Guerra. 1996. "A Common Origin for
959 Woody Sonchus and Five Related Genera in the Macaronesian Islands: Molecular Evidence
960 for Extensive Radiation." *Proceedings of the National Academy of Sciences of the United
961 States of America* 93 (15): 7743–48.

- 962 Kleine, Tatjana, Peter Kindgren, Catherine Benedict, Luke Hendrickson, and Asa Strand. 2007.
963 "Genome-Wide Gene Expression Analysis Reveals a Critical Role for CRYPTOCHROME1 in
964 the Response of Arabidopsis to High Irradiance." *Plant Physiology* 144 (3): 1391–1406.
- 965 Knope, Matthew L., M. Renee Bellinger, Erin M. Datlof, Timothy J. Gallaher, and Melissa A.
966 Johnson. 2020. "Insights into the Evolutionary History of the Hawaiian Bidens (Asteraceae)
967 Adaptive Radiation Revealed Through Phylogenomics." *The Journal of Heredity* 111 (1):
968 119–37.
- 969 Knope, Matthew L., Clifford W. Morden, Vicki A. Funk, and Tadashi Fukami. 2012. "Area and the
970 Rapid Radiation of Hawaiian Bidens (Asteraceae)." *Journal of Biogeography* 39 (7): 1206–
971 16.
- 972 Kojima, Shoko, Mayumi Iwasaki, Hiro Takahashi, Tomoya Imai, Yoko Matsumura, Delphine
973 Fleury, Mieke Van Lijsebettens, Yasunori Machida, and Chiyoko Machida. 2011.
974 "Asymmetric leaves2 and Elongator, a Histone Acetyltransferase Complex, Mediate the
975 Establishment of Polarity in Leaves of Arabidopsis Thaliana." *Plant & Cell Physiology* 52
976 (8): 1259–73.
- 977 Korf, Ian. 2004. "Gene Finding in Novel Genomes." *BMC Bioinformatics* 5 (May): 59.
- 978 Krzywinski, Martin, Jacqueline Schein, Inanç Birol, Joseph Connors, Randy Gascoyne, Doug
979 Horsman, Steven J. Jones, and Marco A. Marra. 2009. "Circos: An Information Aesthetic for
980 Comparative Genomics." *Genome Research* 19 (9): 1639–45.
- 981 Kuki, Yasunobu, Ryoko Ohno, Kentaro Yoshida, and Shigeo Takumi. 2020. "Heterologous
982 Expression of Wheat WRKY Transcription Factor Genes Transcriptionally Activated in
983 Hybrid Necrosis Strains Alters Abiotic and Biotic Stress Tolerance in Transgenic
984 Arabidopsis." *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie
985 Vegetale* 150 (May): 71–79.
- 986 Kurtz, Stefan, Adam Phillippy, Arthur L. Delcher, Michael Smoot, Martin Shumway, Corina
987 Antonescu, and Steven L. Salzberg. 2004. "Versatile and Open Software for Comparing
988 Large Genomes." *Genome Biology* 5 (2): R12.
- 989 Laetsch, Dominik R., and Mark L. Blaxter. 2017a. "BlobTools: Interrogation of Genome
990 Assemblies." *F1000Research* 6 (1287): 1287.
- 991 ———. 2017b. "KinFin: Software for Taxon-Aware Analysis of Clustered Protein Sequences." *G3*
992 7 (10): 3349–57.
- 993 Laforest, Martin, Sara L. Martin, Katherine Bisailon, Brahim Soufiane, Sydney Meloche, and Eric
994 Page. 2020. "A Chromosome-Scale Draft Sequence of the Canada Fleabane Genome." *Pest
995 Management Science* 76 (6): 2158–69.
- 996 Lahari, Triparna, Janelle Lazaro, Jeffrey M. Marcus, and Dana F. Schroeder. 2018. "RAD7
997 Homologues Contribute to Arabidopsis UV Tolerance." *Plant Science: An International
998 Journal of Experimental Plant Biology* 277 (December): 267–77.
- 999 Lang, Daniel, Kristian K. Ullrich, Florent Murat, Jörg Fuchs, Jerry Jenkins, Fabian B. Haas,
1000 Mathieu Piednoel, et al. 2018. "ThePhyscomitrella Patenschromosome-Scale Assembly
1001 Reveals Moss Genome Structure and Evolution." *The Plant Journal: For Cell and Molecular
1002 Biology* 93 (3): 515–33.
- 1003 Lawesson, Jonas Erik. 1988. "Stand-Level Dieback and Regeneration of Forests in the Galápagos
1004 Islands." *Temporal and Spatial Patterns of Vegetation Dynamics*.
1005 https://doi.org/10.1007/978-94-009-2275-4_10.
- 1006 Li, Fay-Wei, Tomoaki Nishiyama, Manuel Waller, Eftychios Frangedakis, Jean Keller, Zheng Li,
1007 Noe Fernandez-Pozo, et al. 2020. "Anthoceros Genomes Illuminate the Origin of Land Plants
1008 and the Unique Biology of Hornworts." *Nature Plants* 6 (3): 259–72.
- 1009 Li, Gang, Hamad Siddiqui, Yibo Teng, Rongcheng Lin, Xiang-Yuan Wan, Jigang Li, On-Sun Lau, et
1010 al. 2011. "Coordinated Transcriptional Regulation Underlying the Circadian Clock in
1011 Arabidopsis." *Nature Cell Biology* 13 (5): 616–22.
- 1012 Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler
1013 Transform." *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btp324>.
- 1014 Li, Heng, and Richard Durbin. 2011. "Inference of Human Population History from Individual
1015 Whole-Genome Sequences." *Nature* 475 (7357): 493–96.

- 1016 Li, Jian, Zheng Wang, Yugang Hu, Ying Cao, and Ligeng Ma. 2017. "Polycomb Group Proteins
1017 RING1A and RING1B Regulate the Vegetative Phase Transition in Arabidopsis." *Frontiers in*
1018 *Plant Science* 8 (May): 867.
- 1019 Liu, Bo, Jian Yan, Weihua Li, Lijuan Yin, Ping Li, Hanxia Yu, Longsheng Xing, et al. 2020. "Mikania
1020 Micrantha Genome Provides Insights into the Molecular Mechanism of Rapid Growth."
1021 *Nature Communications* 11 (1): 340.
- 1022 Liu, Huanhuan, Bao Liu, Shangling Lou, Hao Bi, Hu Tang, Shaofei Tong, Yan Song, et al. 2021.
1023 "CHYR1 Ubiquitinates the Phosphorylated WRKY70 for Degradation to Balance Immunity
1024 in Arabidopsis Thaliana." *New Phytologist*. <https://doi.org/10.1111/nph.17231>.
- 1025 Llorens, Carlos, Ricardo Futami, Laura Covelli, Laura Domínguez-Escribá, Jose M. Viu, Daniel
1026 Tamarit, Jose Aguilar-Rodríguez, et al. 2011. "The Gypsy Database (GyDB) of Mobile Genetic
1027 Elements: Release 2.0." *Nucleic Acids Research* 39 (Database issue): D70–74.
- 1028 Lomolino, Mark V., Brett R. Riddle, and Robert J. Whittaker. 2017. "Biogeography."
1029 https://digitalscholarship.unlv.edu/sls_fac_articles/433/.
- 1030 Loureiro, João, Eleazar Rodriguez, Jaroslav Dolezel, and Conceição Santos. 2007. "Two New
1031 Nuclear Isolation Buffers for Plant DNA Flow Cytometry: A Test with 37 Species." *Annals of*
1032 *Botany* 100 (4): 875–88.
- 1033 Lovell, John T., Alice H. MacQueen, Sujana Mamidi, Jason Bonnette, Jerry Jenkins, Joseph D.
1034 Napier, Avinash Sreedasyam, et al. 2021. "Genomic Mechanisms of Climate Adaptation in
1035 Polyploid Bioenergy Switchgrass." *Nature* 590 (7846): 438–44.
- 1036 Löytynoja, Ari. 2014. "Phylogeny-Aware Alignment with PRANK." *Methods in Molecular Biology*
1037 1079: 155–70.
- 1038 Mandel, Jennifer R., Rebecca B. Dikow, Vicki A. Funk, Rishi R. Masalia, S. Evan Staton, Alex Kozik,
1039 Richard W. Michelmore, Loren H. Rieseberg, and John M. Burke. 2014. "A Target
1040 Enrichment Method for Gathering Phylogenetic Information from Hundreds of Loci: An
1041 Example from the Compositae." *Applications in Plant Sciences* 2 (2).
1042 <https://doi.org/10.3732/apps.1300085>.
- 1043 Mandel, Jennifer R., Rebecca B. Dikow, Carolina M. Siniscalchi, Ramhari Thapa, Linda E. Watson,
1044 and Vicki A. Funk. 2019. "A Fully Resolved Backbone Phylogeny Reveals Numerous
1045 Dispersals and Explosive Diversifications throughout the History of Asteraceae."
1046 *Proceedings of the National Academy of Sciences of the United States of America* 116 (28):
1047 14083–88.
- 1048 Marçais, Guillaume, and Carl Kingsford. 2011. "A Fast, Lock-Free Approach for Efficient Parallel
1049 Counting of Occurrences of K-Mers." *Bioinformatics* 27 (6): 764–70.
- 1050 Markakis, Marios Nektarios, Tinne De Cnodder, Michal Lewandowski, Damien Simon, Agnieszka
1051 Boron, Daria Balcerowicz, Thanaa Doubbo, et al. 2012. "Identification of Genes Involved in
1052 the ACC-Mediated Control of Root Cell Elongation in Arabidopsis Thaliana." *BMC Plant*
1053 *Biology*. <https://doi.org/10.1186/1471-2229-12-208>.
- 1054 Mayr, E. 1942. "Systematics and the Origin of Species from the Viewpoint of a Zoologist,
1055 Columbia Uni." *Press, New York*.
- 1056 Ma, Zhaoxue, Wenjuan Wu, Weihua Huang, and Jirong Huang. 2015. "Down-Regulation of
1057 Specific Plastid Ribosomal Proteins Suppresses thf1 Leaf Variegation, Implying a Role of
1058 THF1 in Plastid Gene Expression." *Photosynthesis Research* 126 (2-3): 301–10.
- 1059 Mendes, Fábio K., Dan Vanderpool, Ben Fulton, and Matthew W. Hahn. 2020. "CAFE 5 Models
1060 Variation in Evolutionary Rates among Gene Families." *Bioinformatics*, December.
1061 <https://doi.org/10.1093/bioinformatics/btaa1022>.
- 1062 Meudt, Heidi M., Dirk C. Albach, Andrew J. Tanentzap, Javier Igea, Sophie C. Newmarch, Angela J.
1063 Brandt, William G. Lee, and Jennifer A. Tate. 2021. "Polyploidy on Islands: Its Emergence
1064 and Importance for Diversification." *Frontiers in Plant Science* 12 (March): 637214.
- 1065 Meurer, J., C. Greveling, P. Westhoff, and B. Reiss. 1998. "The PAC Protein Affects the
1066 Maturation of Specific Chloroplast mRNAs in Arabidopsis Thaliana." *Molecular and General*
1067 *Genetics MGG*. <https://doi.org/10.1007/s004380050740>.
- 1068 Meurer, Jörg, Lisa-Marie Schmid, Rhea Stoppel, Dario Leister, Andreas Brachmann, and Nikolay
1069 Manavski. 2017. "PALE CRESS Binds to Plastid RNAs and Facilitates the Biogenesis of the

- 1070 50S Ribosomal Subunit." *The Plant Journal*. <https://doi.org/10.1111/tpj.13662>.
- 1071 Mitros, Therese, Adam M. Session, Brandon T. James, Guohong Albert Wu, Mohammad B.
- 1072 Belaffif, Lindsay V. Clark, Shengqiang Shu, et al. 2020. "Genome Biology of the
- 1073 Paleotetraploid Perennial Biomass Crop *Miscanthus*." *Nature Communications* 11 (1):
- 1074 5442.
- 1075 Mitsui, Yuki, and Hiroaki Setoguchi. 2012. "Recent Origin and Adaptive Diversification of
- 1076 *Ainsliaea* (Asteraceae) in the Ryukyu Islands: Molecular Phylogenetic Inference Using
- 1077 Nuclear Microsatellite Markers." *Osterreichische Botanische Zeitschrift* 298 (5): 985–96.
- 1078 Moore, B., C. Holt, A. S. Alvarado, and M. Yandell. 2008. "MAKER: An Easy-to-Use Annotation
- 1079 Pipeline Designed for Emerging Model Organism Genomes." *Genome / National Research*
- 1080 *Council Canada = Genome / Conseil National de Recherches Canada*.
- 1081 <https://genome.cshlp.org/content/18/1/188.short>.
- 1082 Noh, Seong Woo, Ri-Ra Seo, Hee Jin Park, and Ho Won Jung. 2021. "Two Arabidopsis Homologs
- 1083 of Human Lysine-Specific Demethylase Function in Epigenetic Regulation of Plant Defense
- 1084 Responses." *Frontiers in Plant Science* 12 (June): 688003.
- 1085 Noutoshi, Yoshiteru, Takashi Kuromori, Takuji Wada, Takashi Hirayama, Asako Kamiya, Yuko
- 1086 Imura, Michiko Yasuda, Hideo Nakashita, Ken Shirasu, and Kazuo Shinozaki. 2006. "Loss of
- 1087 Necrotic Spotted Lesions 1 Associates with Cell Death and Defense Responses in
- 1088 Arabidopsis *Thaliana*." *Plant Molecular Biology* 62 (1-2): 29–42.
- 1089 Ono, Mikio. 1967. "Chromosome Number of *Scalesia* (Compositae), an Endemic Genus of the
- 1090 Galapagos Islands." *Journal of Japanese Botany* 42 (12): 353–60.
- 1091 Oravecz, Attila, Alexander Baumann, Zoltán Máté, Agnieszka Brzezinska, Jean Molinier, Edward
- 1092 J. Oakeley, Eva Adám, Eberhard Schäfer, Ferenc Nagy, and Roman Ulm. 2006.
- 1093 "CONSTITUTIVELY PHOTOMORPHOGENIC1 Is Required for the UV-B Response in
- 1094 Arabidopsis." *The Plant Cell* 18 (8): 1975–90.
- 1095 Ószi, Erika, Csaba Papdi, Binish Mohammed, Aladár Petkó-Szandtner, Tünde Leviczky, Eszter
- 1096 Molnár, Carlos Galvan-Ampudia, et al. 2020. "E2FB Interacts with RETINOBLASTOMA
- 1097 RELATED and Regulates Cell Proliferation during Leaf Development." *Plant Physiology* 182
- 1098 (1): 518–33.
- 1099 Parra, Genis, Keith Bradnam, and Ian Korf. 2007. "CEGMA: A Pipeline to Accurately Annotate
- 1100 Core Genes in Eukaryotic Genomes." *Bioinformatics* 23 (9): 1061–67.
- 1101 Peona, Valentina, Matthias H. Weissensteiner, and Alexander Suh. 2018. "How Complete Are
- 1102 'complete' Genome Assemblies?-An Avian Perspective." *Molecular Ecology Resources* 18
- 1103 (6): 1188–95.
- 1104 Pond, Sergei L. Kosakovsky, Simon D. W. Frost, and Spencer V. Muse. 2005. "HyPhy: Hypothesis
- 1105 Testing Using Phylogenies." *Bioinformatics* 21 (5): 676–79.
- 1106 Putnam, Nicholas H., Brendan L. O'Connell, Jonathan C. Stites, Brandon J. Rice, Marco Blanchette,
- 1107 Robert Calef, Christopher J. Troll, et al. 2016. "Chromosome-Scale Shotgun Assembly Using
- 1108 an in Vitro Method for Long-Range Linkage." *Genome Research* 26 (3): 342–50.
- 1109 Renny-Byfield, Simon, Lei Gong, Joseph P. Gallagher, and Jonathan F. Wendel. 2015. "Persistence
- 1110 of Subgenomes in Paleopolyploid Cotton after 60 My of Evolution." *Molecular Biology and*
- 1111 *Evolution* 32 (4): 1063–71.
- 1112 Reyes-Chin-Wo, Sebastian, Zhiwen Wang, Xinhua Yang, Alexander Kozik, Siwaret Arikrit, Chi
- 1113 Song, Liangfeng Xia, et al. 2017. "Genome Assembly with in Vitro Proximity Ligation Data
- 1114 and Whole-Genome Triplication in Lettuce." *Nature Communications* 8 (April): 14953.
- 1115 Ruan, Jue, and Heng Li. 2020. "Fast and Accurate Long-Read Assembly with wtdbg2." *Nature*
- 1116 *Methods* 17 (2): 155–58.
- 1117 Sanderson, Michael J. 2003. "r8s: Inferring Absolute Rates of Molecular Evolution and
- 1118 Divergence Times in the Absence of a Molecular Clock." *Bioinformatics* 19 (2): 301–2.
- 1119 Sasaki-Sekimoto, Yuko, Yusuke Jikumaru, Takeshi Obayashi, Hikaru Saito, Shinji Masuda, Yuji
- 1120 Kamiya, Hiroyuki Ohta, and Ken Shirasu. 2013. "Basic Helix-Loop-Helix Transcription
- 1121 Factors JASMONATE-ASSOCIATED MYC2-LIKE1 (JAM1), JAM2, and JAM3 Are Negative
- 1122 Regulators of Jasmonate Responses in Arabidopsis." *Plant Physiology*.
- 1123 <https://doi.org/10.1104/pp.113.220129>.

- 1124 Schilling, Edward E., José L. Panero, and Uno H. Eliasson. 1994. "Evidence from Chloroplast DNA
1125 Restriction Site Analysis on the Relationships of *Scalesia* (Asteraceae: Heliantheae)." *American Journal of Botany* 81 (2): 248–54.
- 1126
1127 Schubert, Mikkel, Stinus Lindgreen, and Ludovic Orlando. 2016. "AdapterRemoval v2: Rapid
1128 Adapter Trimming, Identification, and Read Merging." *BMC Research Notes* 9 (February):
1129 88.
- 1130 Seppey, Mathieu, Mosè Mani, and Evgeny M. Zdobnov. 2019. "BUSCO: Assessing Genome
1131 Assembly and Annotation Completeness." *Methods in Molecular Biology* 1962: 227–45.
- 1132 Session, Adam M., Yoshinobu Uno, Taejoon Kwon, Jarrod A. Chapman, Atsushi Toyoda, Shuji
1133 Takahashi, Akimasa Fukui, et al. 2016. "Genome Evolution in the Allotetraploid Frog
1134 *Xenopus laevis*." *Nature* 538 (7625): 336–43.
- 1135 Shen, Lisha, Zhonghui Thong, Ximing Gong, Qing Shen, Yinbo Gan, and Hao Yu. 2014. "The
1136 Putative PRC1 RING-Finger Protein ATRING1A Regulates Flowering through Repressing
1137 MADS AFFECTING FLOWERING Genes in *Arabidopsis*." *Development* 141 (6): 1303–12.
- 1138 Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and
1139 Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation
1140 Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12.
- 1141 Singh, Surjit, Sakthivel Kailasam, Jing-chi Lo, and Kuo-chen Yeh. 2021. "Histone H3 lysine4
1142 Trimethylation-regulated GRF11 Expression Is Essential for the Iron-deficiency Response
1143 in *Arabidopsis thaliana*." *New Phytologist*. <https://doi.org/10.1111/nph.17130>.
- 1144 Smeekens, Sjf. 2006. "Faculty Opinions Recommendation of Large-Scale Analysis of mRNA
1145 Translation States during Sucrose Starvation in *Arabidopsis* Cells Identifies Cell
1146 Proliferation and Chromatin Structure as Targets of Translational Control." *Faculty*
1147 *Opinions – Post-Publication Peer Review of the Biomedical Literature*.
1148 <https://doi.org/10.3410/f.1032260.373846>.
- 1149 Smit, A., R. Hubley, and P. Green. 2013. "RepeatMasker 4.0." *Seattle, WA: Institute for Systems*
1150 *Biology*.
- 1151 Smith, Martin D., Joel O. Wertheim, Steven Weaver, Ben Murrell, Konrad Scheffler, and Sergei L.
1152 Kosakovsky Pond. 2015. "Less Is More: An Adaptive Branch-Site Random Effects Model for
1153 Efficient Detection of Episodic Diversifying Selection." *Molecular Biology and Evolution* 32
1154 (5): 1342–53.
- 1155 Spring, Otmar, Norbert Heil, and Bernhard Vogler. 1997. "Sesquiterpene Lactones and
1156 Flavanones in *Scalesia* Species." *Phytochemistry* 46 (8): 1369–73.
- 1157 Steinbiss, Sascha, Ute Willhoeft, Gordon Gremme, and Stefan Kurtz. 2009. "Fine-Grained
1158 Annotation and Classification of de Novo Predicted LTR Retrotransposons." *Nucleic Acids*
1159 *Research* 37 (21): 7002–13.
- 1160 Stöcklin, Jürg. 2009. "Darwin and the Plants of the Galápagos-Islands." *Bauhinia*, 33–48.
- 1161 Suda, Jan, Anna Krahulcová, Pavel Trávníček, Radka Rosenbaumová, Tomás Peckert, and
1162 Frantisek Krahulec. 2007. "Genome Size Variation and Species Relationships in Hieracium
1163 Sub-Genus Pilosella (Asteraceae) as Inferred by Flow Cytometry." *Annals of Botany* 100
1164 (6): 1323–35.
- 1165 Suda, J., T. Kyncl, and V. Jarolímová. 2005. "Genome Size Variation in Macaronesian
1166 Angiosperms: Forty Percent of the Canarian Endemic Flora Completed." *Plant Systematics*
1167 *and Evolution = Entwicklungsgeschichte Und Systematik Der Pflanzen* 252 (3): 215–38.
- 1168 Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "REVIGO Summarizes and
1169 Visualizes Long Lists of Gene Ontology Terms." *PloS One* 6 (7): e21800.
- 1170 Swain, Swadhin, Nidhi Singh, and Ashis Kumar Nandi. 2015. "Identification of Plant Defence
1171 Regulators through Transcriptional Profiling of *Arabidopsis thaliana* cdd1 Mutant." *Journal*
1172 *of Biosciences* 40 (1): 137–46.
- 1173 Szklarczyk, Damian, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller,
1174 Jaime Huerta-Cepas, Milan Simonovic, et al. 2015. "STRING v10: Protein-Protein Interaction
1175 Networks, Integrated over the Tree of Life." *Nucleic Acids Research* 43 (Database issue):
1176 D447–52.
- 1177 Talavera, Gerard, and Jose Castresana. 2007. "Improvement of Phylogenies after Removing

- 1178 Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments." *Systematic*
1179 *Biology* 56 (4): 564–77.
- 1180 Tan, Yew-Foon, Nicholas O'Toole, Nicolas L. Taylor, and A. Harvey Millar. 2010. "Divalent Metal
1181 Ions in Plant Mitochondria and Their Role in Interactions with Proteins and Oxidative
1182 Stress-Induced Damage to Respiratory Function." *Plant Physiology* 152 (2): 747–61.
- 1183 Tardaguila, Manuel, Lorena de la Fuente, Cristina Marti, Cécile Pereira, Francisco Jose Pardo-
1184 Palacios, Hector Del Risco, Marc Ferrell, et al. 2018. "Corrigendum: SQANTI: Extensive
1185 Characterization of Long-Read Transcript Sequences for Quality Control in Full-Length
1186 Transcriptome Identification and Quantification." *Genome Research* 28 (7): 1096.
- 1187 Valverde, Federico, Andrew Groover, and José M. Romero. 2018. *Evolution of Gene Regulatory*
1188 *Networks in Plant Development*. Frontiers Media SA.
- 1189 Vitales, Daniel, Teresa Garnatje, Jaume Pellicer, Joan Vallès, Arnaldo Santos-Guerra, and Isabel
1190 Sanmartín. 2014. "The Explosive Radiation of Cheirolophus (Asteraceae, Cardueae) in
1191 Macaronesia." *BMC Evolutionary Biology* 14 (June): 118.
- 1192 Wallace, Alfred Russel. 1962. *The Malay Archipelago: The Land of the Orang-Utan and the Bird*
1193 *of Paradise; a Narrative of Travel, with Studies of Man and Nature*. Courier Corporation.
- 1194 Wang, Pengwei, Roman Pleskot, Jingze Zang, Joanna Winkler, Jie Wang, Klaas Yperman, Tong
1195 Zhang, et al. 2019. "Plant AtEH/Pan1 Proteins Drive Autophagosome Formation at ER-PM
1196 Contact Sites with Actin and Endocytic Machinery." *Nature Communications* 10 (1): 5132.
- 1197 Wang, Zhen, Fuxing Wang, Yechun Hong, Jirong Huang, Huazhong Shi, and Jian-Kang Zhu. 2016.
1198 "Two Chloroplast Proteins Suppress Drought Resistance by Affecting ROS Production in
1199 Guard Cells." *Plant Physiology* 172 (4): 2491–2503.
- 1200 Waterhouse, Robert M., Mathieu Seppely, Felipe A. Simão, Mosè Mani, Panagiotis Ioannidis,
1201 Guennadi Klioutchnikov, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2018. "BUSCO
1202 Applications from Quality Assessments to Gene Prediction and Phylogenomics." *Molecular*
1203 *Biology and Evolution* 35 (3): 543–48.
- 1204 White, Oliver W., Alfredo Reyes-Betancort, Mark A. Chapman, and Mark A. Carine. 2018.
1205 "Independent Homoploid Hybrid Speciation Events in the Macaronesian Endemic Genus
1206 *Argyranthemum*." *Molecular Ecology* 27 (23): 4856–74.
- 1207 White, Oliver W., J. Alfredo Reyes-Betancort, Mark A. Chapman, and Mark A. Carine. 2020.
1208 "Geographical Isolation, Habitat Shifts and Hybridisation in the Diversification of the
1209 Macaronesian Endemic Genus *Argyranthemum* (Asteraceae)." *The New Phytologist* 228
1210 (6): 1953–71.
- 1211 Whitewoods, Chris D., Joseph Cammarata, Zoe Nemeč Venza, Stephanie Sang, Ashley D. Crook,
1212 Tsuyoshi Aoyama, Xiao Y. Wang, et al. 2020. "CLAVATA Was a Genetic Novelty for the
1213 Morphological Innovation of 3D Growth in Land Plants." *Current Biology: CB* 30 (13):
1214 2645–48.
- 1215 Whitkus, Richard. 1998. "Genetics of Adaptive Radiation in Hawaiian and Cook Islands Species
1216 of *Tetramolopium* (Asteraceae). II. Genetic Linkage Map and Its Implications for
1217 Interspecific Breeding Barriers." *Genetics*. <https://doi.org/10.1093/genetics/150.3.1209>.
- 1218 Whittaker, Robert J., School of Geography Robert J Whittaker, and José Maria Fernandez-
1219 Palacios. 2007. *Island Biogeography: Ecology, Evolution, and Conservation*. OUP Oxford.
- 1220 Wolfe, K. H. 2001. "Yesterday's Polyploids and the Mystery of Diploidization." *Nature Reviews.*
1221 *Genetics* 2 (5): 333–41.
- 1222 Wu, Martin, Sourav Chatterji, and Jonathan A. Eisen. 2012. "Accounting for Alignment
1223 Uncertainty in Phylogenomics." *PloS One* 7 (1): e30288.
- 1224 Xiong, Xue, Deyang Xu, Zhongnan Yang, Hai Huang, and Xiaofeng Cui. 2013. "A Single Amino-
1225 Acid Substitution at Lysine 40 of an Arabidopsis Thaliana α -Tubulin Causes Extensive Cell
1226 Proliferation and Expansion Defects." *Journal of Integrative Plant Biology* 55 (3): 209–20.
- 1227 Zhang, Bangyue, Jianheng Jia, Min Yang, Chunxia Yan, and Yuzhen Han. 2012. "Overexpression of
1228 a LAM Domain Containing RNA-Binding Protein LARP1c Induces Precocious Leaf
1229 Senescence in Arabidopsis." *Molecules and Cells* 34 (4): 367–74.