1

2 **Running head: *Kale domestication***

3

4 The molecular basis of Kale domestication: Transcription profiling of leaves 5 and meristems provides new insights into the evolution of a *Brassica oleracea* 6 vegetative morphotype

7

8 Tatiana Arias[1,2,4*], Chad Niederhuth[1,3,4], Paula McSteen[1], J. Chris Pires[1]

9

10

11 [1] Division of Biological Sciences, University of Missouri, Columbia, MO, United States

12 [2] Current address: Tecnológico de Antioquia, Medellín, Colombia.

13 [3] Current address: Department of Plant Biology, Michigan State University, East Lansing MI,

14 United States

15 [4] Both authors contributed equally to this manuscript

16

17 *author for correspondence: tatiana.arias48@tdea.edu.co

18

19 Manuscript received _____; revision accepted _____.

20 Short title: Kale domestication

21

22

ABSTRACT [237 words - 250 max]

Morphotypes of *Brassica oleracea* are the result of a dynamic interaction between genes that regulate the transition between vegetative and reproductive stages and those that regulate leaf morphology and plant architecture. In kales ornate leaves, delayed flowering, and nutritional quality are some of the characters potentially selected by humans during domestication.

We used a combination of developmental studies and transcriptomics to understand the vegetative domestication syndrome of kale. To identify candidate genes that are responsible for the evolution of domestic kale we searched for transcriptome-wide differences among three vegetative *B. oleracea* morphotypes. RNAseq experiments were used to understand the global pattern of expressed genes during one single phase of development in kale, cabbage and the rapid cycling kale line TO1000.

We identified gene expression patterns that differ among morphotypes, and estimate the contribution of morphotype-specific gene expression that sets kale apart (3958 differentially expressed genes). Differentially expressed genes that regulate the vegetative to reproductive transition were abundant in all morphotypes. Genes involved in leaf morphology, plan architecture, defense and nutrition were differentially expressed in kale.

RNA-Seq experiments allow the discovery of novel candidate genes involved in the kale domestication syndrome. We identified candidate genes differentially expressed in kale that could be responsible for variation in flowering times, taste and herbivore defense, variation in leaf morphology, plant architecture, and nutritional value. Understanding candidate genes responsible for kale domestication is of importance to ultimately improve Cole crop production.

Keywords: *Brassica oleracea***,** domestication, kale, mustards, plant development, RNA-seq, transcriptomics

INTRODUCTION

Domestication is the process of selection used by humans to adapt wild plants to cultivation. During this process, a set of recurrent characters is acquired across a wide diversity of crops (known as the "domestication syndrome") (Hammer, 1984; Doebley, 1992). Recurrent traits observed in domesticated plants include: loss of seed shattering, changes in seed size, loss of photoperiod sensitivity, and changes in plant physiology and architecture (Doebley, 1992; Harlam, 1992; Allaby 2014). Here we explore the molecular basis behind the poorly defined vegetative domestication syndrome of kale, a *Brassica oleracea* morphotype, the leaves of which are consumed by humans and are also used as fodder. The domestication syndrome of kale includes: apical dominance, ornate leaf patterns, the capacity to delay flower formation and maintain a vegetative state producing a higher yield of the edible portion, nutritional value of leaves, and the ability to defend against herbivores that also give the leaves a unique pungent taste.

The Cole crops (*Brassica oleracea*) are perennial plants native to Europe and the Mediterranean and include a range of domesticated and wild varieties (Snogerup, 1980; Tsunoda, 1980; Purugganan et al., 2000; Allender et al., 2007; Maggioni et al. 2010). Each crop type is distinguished by domestication traits not commonly found among the wild populations

68 (Maggioni et al., 2010). They include dwarf plants with a main stem that has highly compressed
69 internodes (cabbages); plants with an elongated main stem in which the lateral branches are
70 highly compressed due to very short internodes (Brussels sprouts); plants with proliferation of
71 floral meristems (broccoli) or proliferation of aborted floral meristems (cauliflower); and plants
72 with swollen stems (kohlrabi, Marrow-stem kale) or ornate leaf patterns (kales) (Hodgkin, 1995;
73 Lan and Paterson, 2000). It is not so easy to define which domestication traits are common to all
74 of the above-mentioned types, as opposed to the wild species. At early stages in domestication,
75 plants must have been selected for less bitter-tasting, less fibrous, thicker stems, and more
76 succulent storage organs (Maggioni et al., 2010).
77
78 Research has provided evidence that polyploidy and genetic redundancy contribute
79 significantly to the phenotypic variation observed among these crops (Lukens et al., 2004;
80 Paterson, 2005). The genetic basis underlying the phenotypic diversity present in Cole crops has
81 been addressed only for some of these morphotypes. The cauliflower phenotype of *Arabidopsis*
82 is comparable to the inflorescence development in the cauliflower *B. oleracea* morphotype. It
83 has been shown that the gene *CAULIFLOWER* is responsible for the phenotype observed in both
84 species (Kempin et al., 1995; Carr & Irish, 1997; Purugganan et al., 2000; Smith and King,
85 2000). Also, a large QTL for the curd of cauliflower has been found (Lan and Paterson, 2000)
86 indicating that 86 genes are controlling eight curd-related traits, demonstrating this phenotype is
87 under complex genetic control. Other studies have found and confirmed *BoAP-a1* to be involved
88 in curd formation in cauliflower (Anthony et al., 1995; Carr and Irish, 1997; Gao et al., 2007)
89 and have shown that morphologies in this cultivar are under complex genetic control (Sebastian
90 et al., 2002). The molecular bases of broccoli have also been proposed as involving changes in

91 three codons in the *BoCAL* and *BoAP* genes (Carr and Irish, 1997; Lowman and Purugganan,
92 1999). A total of five QTLs have been detected for leaf traits in Brussel sprouts, but no
93 significant QTLs have been found for axillary buds (Sebastian et al., 2002). Head formation in
94 cabbage has been suggested as additively inherited (Tanaka et al., 2009). In kale, a recent study
95 suggest that the lobed-leaf trait is quantitatively inherited (Ren et al., 2019).
96
97 Apical dominance is manifest in kale varieties, involving selection for erect plants with
98 fewer side branches, and more compact plants, which allows more plants to fit into each unit of
99 cultivated soil (Deham et al., 2020). In kale, leaves are green to purple and do not form a
100 compact head like in cabbage. Crop varieties have surprising variation in leaf size, leaf type, and
101 margin curliness. Some also have an over-proliferation of leaf blade tissues. Kale leaf curliness,
102 blade texture, and tissue overgrowth are some of the more variable morphological characters
103 among varieties.
104
105 Leafy kales are considered the earliest cultivated brassicas, originally used for both
106 livestock feed and human consumption (Maggioni et al., 2010) due to leaves and floral buds with
107 great nutritional value (Velasco et al., 2011). Kale leaves also have antioxidative properties, and
108 have been shown to reduce cholesterol levels (Velasco et al., 2007; Soengas et al., 2012). Their
109 high content of beta-carotene, vitamin K, vitamin C, and calcium make it one of the most
110 nutritious vegetables available for human consumption (Stewarta and McDougalla, 2012). Kale
111 is a source of two carotenoids: lutein and zeaxanthin, and contains sulforaphane, a compound
112 suggested to have anti-cancer properties (Velasco et al., 2007). Kale leaves also have mustard
113 oils produced from glucosinolates. These natural chemicals most likely contribute to plant

114 defense against pests and diseases, and impart a pungent flavor property characteristic of all
115 cruciferous vegetables (Sonderby et al., 2010; Carmona et al., 2011). Definitions of
116 domestication syndrome in kale include traits affecting cultivation, harvesting, and cooking –
117 and its uses – including food, medicines, toxins and fibres (Deham et al., 2020).
118
119 Here we used a combination of developmental studies and transcriptomes to understand
120 the vegetative domestication syndrome of kale. Comparisons of kale morphotypes with the
121 "rapid-cycling" morphotype TO1000 and cabbage allow us to identified transcriptome-wide
122 differences that set kale apart. RNA-seq of meristems and immature leaves for the *B. oleracea*
123 vegetative morphotypes, cabbage, kale, and TO1000 reveal gene expression patterns unique to
124 kale and associated developmental and metabolic processes. This allowed us to identify a set of
125 candidate genes we suggest may be important in the kale domestication syndrome.
126
127
128 **MATERIALS AND METHODS**
129
130 ***Plant Material and Experimental Design***
131 Single accessions of cabbage (*B. oleracea* var. *sabauda*), kale (*B. oleracea* var. *viridis*)
132 and the rapid cycling TO1000 (*B. oleracea* var. *alboglabra*) were grown in an environmental
133 chamber under uniform conditions: 10hr day (light 7AM to 5PM, Dark 5:01PM to 6:59AM
134 daily) and watered every other day. Morphotypes were planted in a completely randomized block
135 design. Initially, two seeds per pot were planted, then only one was picked (32 plants per
136 morphotype total). A random number generator was used to move plants and flats around every

137 other day. Structural features of the meristems and developing leaves were observed using an
138 Environmental Scanning Electron Microscopy (ESEM) and light microscopy. Images were
139 processed with Adobe (San Jose, California, USA) Photoshop, version 7.0.
140
141 Tissues were collected 24 days after seeds were planted. The upper portion of the stems,
142 including immature leaves, and apical and lateral meristems were flash frozen in liquid nitrogen.
143 The assumption was made that genes involved in developing different morphologies should be
144 expressed in younger leaves, shoots, and meristems. Three biological replicates per morphotype
145 were collected, each replicate consisting of pooled tissue from eight plants.
146
147 ***RNA Extraction and cDNA Synthesis***
148 Lysis and homogenization of fresh soft tissue were performed before RNA extraction;
149 PureLink Micro-to-Midi Total RNA Purification System (Invitrogen, Cat. No 12183-018,
150 Carlsbad, California, USA) was used. RNA extraction was done using the Purelink RNA Mini
151 Kit (Ambion, Cat. No12183-018A, Carlsbad, California, USA). DNase was not used before
152 proceeding with cDNA synthesis as recommended in the protocol. For cDNA synthesis MINT
153 (Evrogen, Cat. No SK001, Moscow, Russia) was used, this kit enzymes synthesize full-length-
154 enriched double stranded cDNA from a total polyA^{+} RNA. The cDNA was then amplified by 16
155 -- 21 cycles of polymerase chain reaction and purified using a PCR Purification Kit (Invitrogen
156 K3100-01 Carlsbad, California, USA).
157
158 ***Library Construction and Illumina Sequencing***

159 Protocols previously published by Steele et al. (2012) were implemented here. End repair
160 was performed on 400µl with 3µg of normalized ds-cDNA prior to ligating barcoding adapters
161 for multiplexing, using NEB Prep kit E600L (New England Biolabs, Ipswich, Massachusetts,
162 USA). Shearing was done using a Bioruptor® sonication device (Diagenode, Denville New
163 Jersey, USA), at 4°C, continuously, for 15 min. total (7 mix) on high. Three replicates per
164 morphotype were tagged with different adaptors and sequenced on a single lane of an Illumina
165 GAIIx machine.

166 Oligonucleotides used as adapters were: Cabbage (1: ACGT, 2: GCTT, 3: TGCT), kale
167 (1: TACT, 2:ATGT, 3:GTAT), TO1000 (1:CTGT, 2:AGCT, 3:TCAT). To prepare adapters,
168 equal volumes were combined of each adapter at 100 µmol/L floating them in a beaker of boiling
169 water for 30 min and snap cooling them on ice. Ligation products were run on a 2% low-melt
170 agarose gel and size-selected for ~300 bp using an x-tracta Disposable Gel Extraction Tool
171 (Scientific, Ocala Florida, USA). Fragments were enriched using PCR in 50 µL volumes
172 containing 3 µL of ligation product, 20 µL of ddH2O, 25 µL master mix (from NEB kit), and 1
173 µL each of a 25 µmol/L solution of each forward and reverse primer (forward 5′-AAT GAT
174 ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC*
175 T-3′; reverse 5′-CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC
176 TGA ACC GCT CTT CCG ATC*-3′; both HPLC purified). Thermal cycle routine was as
177 follows: 98°C for 30 s, followed by 15 cycles of 98°C for 10 s, 65°C for 30 s, and 72°C for 30 s,
178 with a final extension step of 72°C for 5 min. Products were run on a 2% low-melt agarose gel
179 and the target product was excised and purified with the tools described above. DNA was
180 purified at each step during the end repair, adapter ligation, size selection and fragment
181 enrichment with either a Gel Extraction kit (Qiagen, Germantown Maryland, USA) or a DNA

182 QIAquick PCR Purification kit (Qiagen, Germantown Maryland, USA). The three replicates per
183 morphotype were ran on one third of a lane with single-end 98-bp reads on an Illumina GAIIx
184 Genome Analyzer using the Illumina Cluster Generation Kit v2-GA II, Cycle Sequencing Kit v3
185 (Illumina, San Diego, USA), and image analysis using Illumina RTA 1.4.15.0 (Illumina, San
186 Diego, USA) at University of Missouri DNA Core.
187
188 ***Data quality control and preprocessing, alignment and differential expression analysis***
189 Reads were first trimmed for the end quality using Cutadapt v2.4 (Martin et al., 2011)
190 and read quality assessed using FastQC (Andrews, 2010). These were then aligned against the
191 *Brassica oleracea* TO1000 genome (Parkin et al., 2014) downloaded from EnsemblPlants 44
192 (Kersey et al., 2018) with two passes of STAR v2.7.1a (Dobin et al., 2013). In the first pass,
193 reads for each sample were aligned for splice-junction discovery, in addition to those junctions
194 provided in the TO1000 annotations. Splice junctions from each sample were then provided as a
195 list for a second final alignment and read counts calculated for each gene using STAR. Mapping
196 statistics are provided in Supplementary Table 1. Raw sequencing data and unnormalized read
197 counts are publicly available through the Gene Expression Omnibus accession GSE149483.
198 Read counts were imported into R v3.6.3 (Team R) and differential expression was determined
199 using DESeq2 v1.26.0 (Love et al., 2014). Specifically, read counts were normalized, with the
200 TO1000 samples as reference, and fitted to a parametric model. We looked for differentially
201 expressed genes (DEGs) for each pairwise comparison (kale vs TO1000, kale vs cabbage, and
202 cabbage vs TO1000). Comparisons were done to ultimately separate unique kale-morphotype
203 DEGs. Since each pairwise comparison increases the potential number of false positives, these
204 were combined into a single table and the false discovery rate (FDR) (Benjamini and Hochberg,

205 1995) was adjusted for the entire table. Genes with an FDR < 0.05 and Log-fold change > 1 or <
206 -1 were considered as differentially expressed genes (DEGs). Plots were made using the R
207 packages ggplot2 (Wickham, 2009), pheatmap (Kolde, 2019), and a VennDiagram (Delhomme,
208 2018). Transcription-associated proteins (TAPs) were identified using the approach described for
209 the construction of PlnTFDB v3.0 (http://plntfdb.bio.uni-potsdam.de) (Riaño-Pachón et al.,
210 2007; Pérez-Rodríguez et al., 2010). All code for bioinformatic analysis and original figures pdfs
211 and tables are available on GitHub (https://github.com/niederhuth/The-molecular-basis-of-Kale-
212 domestication).
213
214 ***Gene annotation, GO, and KEGG analysis***
215 Gene ontology (GO) terms (Ashburner et al., 2000) are not currently available for the
216 TO1000 assembly. In order to annotate TO1000 gene models with GO terms, TO1000 protein
217 sequences were Blasted (blastp) against protein sequences from UniProtKB Swiss-Prot (release
218 2019_05) (Consortium 2018) using Diamond (Buchfink et al., 2014), with an e-value cutoff of
219 $1e^{-5}$ and a max of 1 hit per query. UniProt IDs were then used to extract GO terms and mapped to
220 the corresponding TO1000 gene model. In total, 35,422 genes (out of 59,225 total, ~59.8%) had
221 a GO term assigned. GO term enrichment was performed using topGO v 2.38.1 (Alexa and
222 Rahnenfuhrer, 2019) and GO.db v3.10.0 (Carlson, 2019) with the parent-child algorithm
223 (Grossman et al., 2007) and Fisher's Exact Test. A minimum node size of 5 was required, i.e.
224 each GO term required a minimum of 5 genes mapping to it in ordered to be considered. P-
225 values were adjusted for FDR and considered significant with an adjusted p-value < 0.05.
226
227 **RESULTS**

228 ***Kale leaf development***

229 Developmental transitions and morphological changes were observed in two kale
230 varieties (curly and non-curly kale) and the rapid cycling TO1000 for twenty-two days (Fig. 1-
231 2). A week after planting all morphotypes had germinated; at first only the cotyledons were
232 observed. Curly-kale grew faster than non-curly kale and TO1000. Using Environmental
233 Scanning Electron Microscopy (ESEM), tiny true leaves with lobed margins and abundant
234 trichomes were observed for curly-kale at five days (Fig. 2N), while non-curly kale and TO1000
235 had not developed lobes or leaves at this point (Fig. 2M, O). Eight-day-old curly-kale first
236 immature leaves were seen emerging from between the cotyledons and covered by trichomes
237 (Fig. 2B). Ten-days after germination curly-kale seedlings had abundant trichomes and highly
238 lobed leaf margins; leaf shape was elliptic to obovate and leaves had red venation and red
239 petioles (Fig. 2E). In comparison to the kale varieties, TO1000 seedlings had taller hypocotyls
240 and erect stems. Leaf shape was oblanceolate and leaf margins were sinuate (Fig. 2D). Twelve-
241 day-old curly-kale produce ectopic meristems in the leaf margins and more sporadically in the
242 leaf blade of mature leaves; microscopic scales surrounded these ectopic meristems (Fig. 2Q).
243 Ectopic meristems located at the tip of each lobe had a vascular bundle inserted to them and were
244 always subtended by a trichome. Nineteen-day-old curly-kale and TO1000 had four leaves, the
245 fourth one small; while non-curly kale has three leaves and the third one small. Twenty to
246 twenty-two day-old curly-kale leaves develop more lobes and less trichomes with time. TO1000
247 leaves were thinner with sinuate margins. In summary, we observed kale margins were more
248 lobed than TO1000 (Fig. 2).

249

250 ***RNA-seq and differential gene expression***

251 To identify genes that might be associated with domestication traits including leaf
252 development in curly-kale, RNA-seq was used to assess transcriptional differences between
253 curly-kale, cabbage, and TO1000. Three biological replicates each were sequenced, however,
254 one of the cabbage replicates was not used for assembly because very few reads of low quality
255 were obtained after sequencing. Between 9 to 25 million trimmed reads were obtained per
256 sample, with ~78-89% reads mapping uniquely to the TO1000 *B. oleracea* genome (Suppl. Table
257 1). After normalization with DESeq2, hierarchical clustering and principle component analysis
258 showed that samples were grouped by morphotype, with greater variance between morphotypes
259 than between replicates (Suppl. Fig. 1).
260
261 Differentially expressed genes (DEGs) were called for all pair-wise comparisons between
262 morphotypes (kale vs TO1000, kale vs cabbage, cabbage vs TO1000), with the aim of
263 identifying genes that were differentially expressed in kale compared to other morphotypes.
264 Genes were considered differentially expressed if they had a two-fold change in expression and a
265 p-value < 0.05 after correction for multiple testing. In total, 8854 DEGs were identified between
266 kale and TO1000 (4599 higher expressed in kale, 4255 lower expressed in kale), 8037 DEGs
267 between kale and cabbage (4704 higher expressed in kale, 3333 lower expressed in kale), and
268 8189 DEGs between cabbage and TO1000 (3406 higher expressed in cabbage, 4783 lower
269 expressed in cabbage) (Fig. 3A). There were a total of 3958 DEGs shared between the kale-
270 TO1000 and kale-cabbage comparisons, 2960 found only in the kale comparisons and 998 found
271 in all three comparisons (Fig. 3B).
272
273 ***Polyploidy and its contribution to the kale phenotype***

274 We explored how each of the three subgenomes resulting from an ancient whole genome
275 triplication contributed to differential gene expression between morphotypes. Syntenic genes
276 were significantly overrepresented amongst DEGs while non-syntenic genes were
277 underrepresented in the kale-cabbage and cabbage-TO1000 comparisons, but not the kale-
278 TO1000 comparisons (Fig. 4A). The gene content of the three sub-genomes of *B. oleracea* have
279 been previously reconstructed (Parking et al., 2014) and using this, the representation of each
280 sub-genome amongst DEGs was tested (Fig. 4C-D). The LF (least-fractionated) subgenome was
281 underrepresented in kale-TO1000 comparison (odds ratio = 0.9162346. FDR-corrected p-value =
282 4.116097e-03) and overrepresented in the cabbage-TO1000 comparison (odds ratio = 1.0816227.
283 FDR-corrected p-value = 9.948104e-03). The MF1 (more-fractionated 1) subgenome was
284 significantly overrepresented for DEGs in all three comparisons (kale-TO1000: odds ratio =
285 1.0912796 , FDR-corrected p-value < 1.103789e-02; kale-cabbage: odds ratio = 1.2091591,
286 FDR-corrected p-value < 5.909977e-08; cabbage-TO1000: odds ratio = 1.1776233, FDR-
287 corrected p-value = 2.913654e-06). The MF2 (more-fractionated 2) subgenome was
288 overrepresented in the cabbage-TO1000 compariosn (odds ratio = 1.1118552. FDR-corrected p-
289 value = 4.931485e-03).
290
291 ***Gene ontology and KEGG pathway enrichment analysis***
292 DEGs for each pair-wise comparison and the list of DEGs shared between the two kale
293 comparisons (3958 genes) were separated into lists of higher and lower-expressed genes and then
294 tested for enrichment of GO (Gene Ontology) terms (Suppl. Tables 5-12) and KEGG (Kyoto
295 Encyclopedia of Genes and Genomes) pathways (Suppl. Tables 13-20). We focus here
296 specifically on those kale comparisons to better understand the kale domestication syndrome, in

297 particular GO terms in the biological process (BP) category, as this includes developmental
298 terms. GO terms and KEGG pathways related to the cabbage vs TO1000 comparison are
299 provided as supplement (Suppl. Fig. 2).
300
301 *Kale vs. TO1000*- There were 91 (64 BP) and 17 (1 BP) GO terms enriched in the higher
302 (Suppl. Table 5) and lower (Suppl. Table 6) expressed DEGs, respectively. Higher expressed
303 genes were characterized by a number of metabolic processes: 'lipid metabolism', 'nitrile
304 metabolism,' 'sulfur compound metabolism', benzene-containing compound metabolism',
305 'organic hydroxy compound biosynthesis', and 'carbohydrate derivative catabolism'. There were
306 also multiple GO terms related to photosynthesis and plastid organization and environmental
307 responses to both biotic and abiotic stresses. The only developmental terms enriched were related
308 to senescence (Suppl. Fig. 2A). Lower expressed genes showed little enrichment, except for
309 genes involved in 'cellular nitrogen compound metabolism' (Suppl. Figure 2B). Six KEGG
310 pathways were enriched in higher expressed genes (Suppl. Table 13) and largely correspond to
311 metabolic pathways identified in the GO term analysis. These include processes related to
312 'photosynthesi-antenna proteins', 'fatty acid elongation', 'flavonoid biosynthesis', 'steroid
313 biosynthesis', 'glyoxylate and dicarboxylate metabolism', 'other glycan degradation', and
314 'cyanoamino acid metabolism'. Lower expressed genes were enriched in 'spliceosome' and
315 'circadian rhythm' pathways (Suppl. Table 14).
316
317 *Kale vs. cabbage*- Higher expressed genes had 21 (15 BP) enriched GO terms (Suppl.
318 Table 7), while lower expressed genes (Suppl. Table 8) had 26 (11 BP). Similar to the kale vs
319 TO1000 comparison, there was enrichment for processes involved in 'fatty acid derivative

320 metabolism', 'nitrile metabolism', and 'photosynthesis' (Suppl. Figure 2C). There was also
321 enrichment for environmental responses, specifically responses to herbivores. Additionally, there
322 was enrichment for processes in 'carbon fixation', 'benzoate metabolism, 'endosome
323 organization', and the 'negative regulation of ion transport. Lower expressed genes (Suppl. Fig.
324 2D) include 'glycosinolate metabolism', 'response to chitin', and 'gene expression'. 'Sulfur
325 compound metabolism' was enriched in both comparisons, but was enriched amongst lower-
326 expressed genes in the kale-cabbage comparison, while being enriched in higher-expressed genes
327 in the kale-TO1000 comparison. These are again reflected in KEGG pathways where higher
328 expressed genes were enriched for 'other glycan degradation' KEGG pathways (Suppl. Table
329 15). Lower expressed genes were enriched for 'glucosinolate biosynthesis' and 'glutathione
330 metabolism' KEGG pathways (Suppl. Table 16).
331
332 *Shared genes*- Higher expressed genes were enriched for 32 (22 BP) GO terms (Suppl.
333 Table 11 and Suppl. Fig. 2G). These again include multiple metabolic terms, especially related to
334 'fatty acid derivative metabolism', 'nitrile metabolism', 'carbohydrate derivative catabolism',
335 and 'benzoate metabolism' . Also similar to the individual comparisons, 'photosynthesis' and
336 terms related to abiotic and biotic stresses, especially herbivores, were enriched. Unique to this
337 subgroup was an enrichment for protein phosphorylation terms. Only four GO terms were
338 enriched in the lower expressed genes and these were all in the cellular component GO term
339 class (Suppl. Table 12). Higher expressed DEGs were enriched for 'other glycan degradation'
340 and 'Sesquiterpenoid and triterpenoid biosynthesis' KEGG pathways. The former being also
341 enriched in both of the individual comparisons. No KEGG pathways were enriched in the shared
342 lower expressed DEGs (Supp. Tables 19 and 20).

343

344 ***Candidate genes potentially involved in the domestication of kale-***

345 A set of candidate genes were identified based on their expression pattern (Fig. 5) and
346 functions related to the suite of kale domestication syndrome characteristics. We further
347 identified transcription factors (TF) families and found a total of 141 TF in a total of 3010
348 differentially expressed genes (Fig. 3B). Some of the most abundant included bHLH, ERF and
349 MYB TFs (Suppl. Table 21).

350

351 *Apical dominance and plant architecture* - Several genes known to be involved in shoot
352 grow inhibition were differentially expressed in kale tissues and upregulated, these include: an
353 auxin-resistance *AXR1* (Bo8g102450, Bo8g052710) (Stirnberg et al., 1999), *TCP18* (Teosinte
354 Branched 1-Like 1) (Bo3g071570), two copies of *REVOLUTA* (Bo9g130730, Bo7g021280)
355 (Poza-Carrión et al., 2007), and the Beta-carotene isomerase D27 chloroplastic (Protein *DWARF-*
356 *27* homolog) (Bo9g035190) (Waters et al., 2012). The More Axillary Branches gene *MAX1*
357 (Bo3g042090) (Bennett et al., 2006) was also differentially expressed and downregulated in kale
358 (Table 1; Fig. 5A).

359

360 *Leaf development* - Differentially expressed genes in kale include two upregulated copies
361 of *AXR1* (Bo8g052710; Bo8g102450). The downregulated LOB domain-containing protein *ASL6*
362 (ASYMMETRIC LEAVES 2-like protein 6) (Bo7g041000) (Semiarti et al., 2001). These genes
363 could have a major role in the highly lobed margin kale phenotype (Fig. 5A). Genes
364 differentially expressed in all three pair-wise comparisons that were potentially involved in the
365 kale domestication syndrome (Fig. 3) include: *PID* (Bo4g182790) that regulates auxin flower

366 and organ formation (Bennett et al., 1995), and *HAT5* (Bo5g155120) which is involved in leaf
367 formation and has been characterized as involved in leaf margin development (Aoyama et al.,
368 1995).

369 Identified transcription factors included: the AP2-like ethylene-responsive *ANT*
370 (Bo3g153310). This gene is a positive regulator of cell proliferation, expansion and regulation of
371 ectopic meristems (Shigyo et al., 2006), matching what was observed in the kale phenotypes
372 (Fig. 2Q). The growth-regulating factors 3, 4 (*GRF3* and *GRF4*; Bo4g186220 and Bo4g122960)
373 involved in the regulation of cell expansion in leaf and cell proliferation and the Basic helix-
374 loop-helix protein 64 *AtbHLH64* (Bo3g094060), an atypical *bHLH* transcription factor were also
375 identified here (Fig. 5A; Suppl. Table 21).

376 A significant amount of trichomes in kale leaves were observed during development (Fig.
377 2 N, H). More than 10 genes related to trichome formation (Table 1, Fig. 6A) were differentially
378 expressed in kale, some of those included: Two upregulated copies of an Myb-related proteins
379 *MYB106* (Bo5g156410 and Bo8g104210). These genes can act as both a positive or negative
380 regulators of cellular outgrowth and promote trichome morphogenesis and expansion (Gilding
381 and Marks, 2010). A Homeobox-leucine zipper protein GLABRA 2- Like *GL2* (Bo6g079990)
382 was upregulated in kale. This gene is required for correct morphological development and
383 maturation of trichomes, regulates the frequency of trichome initiation and determines trichome
384 spacing in *Arabidopsis* (Morohashi et al., 2007). Two expressed copies of the *WRKY*
385 transcription factor 44 (Bo3g031650 and Bo4g030720), and the transcription factor
386 TRIPTYCHON *TRY* (Bo1g051040) involved in trichome branching, were also differentially
387 expressed in kale, among others (Table1, Fig. 6A).
388

389 *Flowering* - Kale phenotypes display a significant delaying in flowering in comparison to
390 other *Brassica oleracea* morphotypes. We hypothesized this maintenance in a vegetative stage,
391 was an important character during domestication due mainly to the harvest of fresh leaves for
392 consumption. Our mature kale phenotypes did not display flowering during the course of our
393 experiments (24 days) (Fig. 1). A series of Agamous-like MADS-box proteins were exclusively
394 expressed in kale (Fig. 3B) including two downregulated genes *AGL12* (Bo6g094120) and
395 *AGL70* (Bo03856s010) and an upregulated *AGL27* (Bo3g100570*)*. MADS-box proteins such a
396 *SOC1* (Bo4g024850) was downregulated, and two copies of FLOWERING LOCUS C
397 (Bo9g173370, Bo3g024250) were upregulated (Table 1, Fig. 5A), both involved in the negative
398 regulation of flowering, together with the MYB-related protein *R1 MYB1R1* (Bo6g119010)
399 (Suppl. Table 21). Other genes differentially expressed in all three pair-wise comparisons that
400 were potentially involved in the negative regulation of flowering (Fig. 3) included: *SCO1*
401 (Bo4g024850, Bo3g038880), *SPL3* (Bo4g042390), *AP1* (Bo2g062650), *BLH3* (Bo6g119600),
402 *AGL68/MAF5* (Bo3g100540) among others (Table 1, Fig. 5A)
403
404 *Plant defense* - Differentially expressed genes involved in plant defense exclusively
405 expressed in curly kale leaves (Fig. 3b) include: Two copies of the upregulated Myb-related
406 protein *MYB96* (Bo2g082270) involved in the activation of cuticular wax biosynthesis that could
407 potentially deter herbivores. Five upregulated copies of a glucosinolate synthesis and regulation
408 enzyme Myrosinase *TGG1* (Bo00934s010, Bo2g155820, Bo2g155840, Bo2g155870,
409 Bo8g039420), involved in the degradation of glucosinolates to produce toxic degradation
410 products that can deter insect herbivores (Zheng et al. 2011). Last, two copies of the Indole
411 Glucosinolate O-Methyltransferase, one was upregulated *IGMT1* (Bo2g092580), an a second one

412 downregulated *IGMT1* (Bo2g092680) (Table 1, Fig. 5B), and *MYB29* (Bo9g175680) all involved
413 in plant defense against fungus (Suppl. Table 21). Refer to Table 1 and Fig. 5B more other
414 genes potentially involved in plant taste and defense.
415
416 *Nutritional value of leaves*- Kale is known to have a great nutritional value including
417 nutrients like Vitamin C, calcium and iron (Edelman & Colt, 2016). Most notable, were three
418 copies of the L-ascorbate oxidase (*AAO*) (Ascorbase) (EC 1.10.3.3)(Bo2g165650, Bo9g150050,
419 Bo7g118970) involved in Vitamin-C synthesis, all three of which were downregulated in kale
420 (Table 1; Fig. 5B) and the enzyme Ferritin-1 *LSC30* (Bo3g001360) proposed as a source of iron.
421 Refer to Table 1 and Fig. 5B more other genes potentially involved in nutritional value of kale
422 leaves.
423
424 **DISCUSSION**
425 The vegetative domestication syndrome of kale is characterized by apical dominance, leaf
426 morphology, maintenance of a vegetative stage for high leaf production, taste/defense, and
427 nutritional value. Differentially Expressed Genes (DEGs) were identified in all pair-wise
428 comparisons. In the cabbage comparisons these were enriched amongst syntenic genes, while
429 being depleted in non-syntenic genes (Fig. 4). This aligns with recent findings that genes derived
430 from the ancient *Brassica* polyploidy are more likely involved in domestication in *B. rapa* (Qi et
431 al., 2019). However, in the kale-TO1000 comparison, no such enrichment was found. As the
432 subgenomes of *B. oleracea* have been previously reconstructed, we further explored the relative
433 contribution of each. The LF (least-fractionated) subgenome contains the most genes and so
434 unsurprisingly also had the most DEGs and was over-represented in the cabbage-TO1000

435 comparions, however, it was underrepresented in the kale-TO1000 comparison. In contrast the
436 MF1 (more-fractionated 1) subgenome was consistently overrepresented. The MF2 (more-
437 fractionated 2) subgenome was overrepresented in the cabbage-TO1000 comparison. Given the
438 limited number of morphotypes in this study, we cannot conclude a general pattern regarding the
439 relative importance of syntenic vs non-syntenic genes and each subgenome, however, the over-
440 representation of the MF1 subgenome in each comparison should be explored further.

441 We focused our search on DEGs that were amongst comparisons with kale, reasoning
442 that these were more likely to underly the unique characteristics of kale. We tested for
443 enrichment in GO terms and KEGG pathways to gain a global view of the differences. Many
444 more terms and pathways were enriched amongst higher expressed genes, despite several
445 thousand DEGs being identified in both groups. This may suggest more coordination in the
446 regulation of kale higher-expressed genes or simply better annotation of genes in this list.
447 Enriched terms and pathways in all categories were typically involved in metabolism or defense,
448 while we found no clear enrichment of developmental GO terms. Changes in these functions
449 would affect the taste, texture, and nutrition of leaves and thus likely have been affected by
450 human selection. There was a relative absence of developmentally-related GO terms, with the
451 exception of senescence related terms in the kale-TO1000 comparison. This overall lack of
452 enrichment for developmentally-related GO terms could result from either insufficient annotation
453 of developmental genes or due to more subtle expression differences that would require tissue-
454 specific or single-cell approaches. Alternatively, the large developmental changes observed
455 could be caused by differential expression of only a small number of key genes. To address this,
456 we examined sets of genes, whose function based on homology and previous work, suggest an
457 involvement in the developmental traits examined in this study. This analysis found many

458 differentially expressed candidate genes shared amongst comparisons with kale, supporting their
459 potential role in the kale domestication syndrome.
460
461 We identified potential candidate genes for the morphological and chemical differences
462 among *Brassica* morphotypes. First, we found differentially expressed genes that set kale apart
463 in terms of its morphology. Apical dominance has been proposed as a main character involved in
464 crop domestication (Doebley et al., 1997). Suppression of axillary branches aiming to
465 concentrate resources in the main stem is a salient phenotypic character in most *Brassica* crops.
466 Apical dominance during domestication in kale can be explained by the indirect theory of apical
467 dominance as auxin-induced stem growth inhibits bud outgrowth by diverting sugars away from
468 buds since growing stems are a strong sink for sugars (Kebrom, 2017). Here genes involved in
469 apical dominance and branching suppression were differentially expressed (Table 1, Fig. 5).
470
471 Kale displays morphological diversity in leaves, including leaf shape, size and margins,
472 trichome development, colors among others which is likely the result of genetic variation,
473 selected by plant breeders. Environmental Scanning Electron Microscopy (ESEM) showed that
474 many of the characteristic leaf traits of kale are established early in leaf development. We
475 collected tissues during kale leaf initiation and primary morphogenesis (the upper portion of the
476 stem including immature leaves, and apical and lateral meristems) to capture candidate genes
477 related to the establishment of basic leaf structure such lamina initiation, specification of lamina
478 domains, and the formation of marginal structures such lobes or serrations. Leaf development
479 encompasses three continuous and overlapping phases. During leaf initiation, the leaf
480 primordium emerges from the flanks of the stem apical meristem (SAM) at positions determined

481 by specific phyllotactic patterns (Fig. 2 M-N). In the second phase, the leaf expands laterally and
482 primary morphogenesis events occur from specific meristematic regions at the leaf margin
483 (blastozones) (Fig. 2A-F). In the third phase of secondary morphogenesis, extensive cell
484 expansion and histogenesis occurs (Fig. 2 G-L, Q, R).
485
486 Three types differentially expressed genes related to leaf morphogenesis phenotype in
487 kale were found in our samples (Table 1; Fig. 3 and 5), genes: (1) involved in margin
488 development, (2) involved in cell proliferation, expansion and the development of ectopic
489 meristems and (3) involved in trichome development. The Lateral Organ Boundaries LOB
490 domain-containing protein 4 (Asymmetric Leaves 2-like protein 6) *ASL6* (Bo7g041000) was
491 identified as a potential important gene responsible of the curly kale leaf phenotype (Semiarti et
492 al., 2001) (Fig. 1, Fig. 2E, H, K). Loss-of-function mutations in the *AS2* homolog gene in
493 *Arabidopsis* result in asymmetric leaf serration, with generation of leaflet-like structures from
494 petioles and malformed entire veins and the ectopic expression of class 1 *KNOX* genes in the
495 leaves (Matsumura et al., 2009). *ASL6* was differentially expressed and upregulated in kale.
496 Asymmetric placement of auxin response at the distal leaf tip precedes visible asymmetric leaf
497 growth. A second set of candidates involved two copies of the BEL1-like homeodomain protein
498 1 *BEL1/BHL1* (Bo3g029830, Bo4g142080) that were found to be over-expressed in kale. This
499 gene has been involved in ovule development (Bellaoui et al., 2001), but since we did not collect
500 any reproductive tissue in our sampling, we suggest this might have some other developmental
501 role in kale.
502

503 In some species with simple leaves, *KNOXI* over-expression can lead to variable
504 phenotypes, which include knot-like structures on the leaves, curled or lobed leaves and ectopic
505 meristems on leaves (Tsiantis and Hay, 2003). Our differential expression analysis did not show
506 *KNOXI* over-expression, however, the differential expression of *BELL* genes, another class of
507 TALE HD proteins found in plants, was detected. Yeast two-hybrid studies have indicated that
508 *KNOX* proteins interact with BEL1-like (*BELL*) proteins. The mRNA expression patterns of
509 *KNOX* and *BELL* proteins overlap in meristems, suggesting that they may potentially interact in
510 vivo (Smith et al, 2000; Smith et al., 2002). Analogous *BLH–KNOX* interactions have also been
511 reported in other plants, such as potato (*Solanum tuberosum*) and barley (*Hordeum vulgare*),
512 indicating that these interactions are evolutionarily conserved and that the interaction is probably
513 required for their biological functions (Muller et al., 2001 ; Chen et al., 2003). A matter of debate
514 is whether *KNOX* and *BLH* can exert some of their functions independently of each other or
515 whether the formation of *KNOX/BLH* heterodimers is mandatory for TALEs to work. We did not
516 find differential expression of *KNOX* genes in this analysis perhaps for several reasons: (1)
517 *KNOX* do not participate in lobes formation in *Brassica oleracea*; (2) we did not target the
518 correct point in development to capture *KNOX* expression with relation to leaf lobation. Perhaps
519 these genes are expressed later during development in kale leaves, however we collected tissues
520 from early leaflets where lobes are starting to develop; (3) a change in *BELL* expression is
521 sufficient to affect *KNOX* and *BELL* interactions and alter leaf morphology without a
522 corresponding change in *KNOX* expression, if they were functionally redundant then *KNOX*
523 expression would suffice to hide any effects of *BELL* expression differences. In some species,
524 such as legumes dissected leaves do not express *KNOX* genes; on the contrary *KNOX* expression
525 has been observed in leaf primordia of species with un-lobed leaves, such as *Lepidium*

526 *oleraceum* (Piazza et al., 2010). We have also found several downregulated Growth-Regulating
527 Factors (*GRF3* and *GRF4*) that have been shown interact with KNOX genes (Kuijt et al., 2014).
528 A third gene involved in cell proliferation, expansion and the development of ectopic meristems
529 included the NGATHA *NGA1* (Bo3g039420) involved in cell proliferation (Lee et al., 2015),
530 however it has been reported as downregulated for cell over proliferation and in our copies was
531 upregulated (Table 1).
532
533 Epidermal cells are stimulated to differentiate into trichomes when a regulatory complex
534 including transcription factors are triggered (Yang and Ye, 2013). ESEM results showed
535 hairiness and individual trichomes are seen in kale leaf blades (Fig 2 N, Q). Differentially
536 expressed genes involved in trichome formation included two *MYB* transcription factors
537 (Higginson et al, 2003; Gilding and Marks, 2010; Pattanaik et al., 2014), the enhancer of *GL3*
538 *EGL3* (Bo9g035460), the Homeobox-leucine zipper proteins *GL1* (Bo7g090950) and *GL2*
539 (Bo6g079990) (Morohasi et al., 2007), the transcription factor *TRY* (Bo1g051040) (Schnittger et
540 al., 1999) and homeodomain GLABROUS 11, *HDG11* (Bo4g039530) (Khosla et a., 2014).
541 These five genes either interact or act redundantly to promote trichome differentiation and were
542 upregulated in kale (except *HDG11*) (Table 1; Fig. 5A). Trichome regulatory genes have been
543 studied in *Brassica villosa*, a wild C genome relative *B. oleracea* that is densely covered by
544 trichomes. Nayidu et al. (2014) found that *TRY* was upregulated in trichomes of *B. villosa* in
545 contrast to *Arabidospis* and other *Brassica* species where this gene has been proposed as a
546 negative regulator. Here we found *TRY* was upregulated in leaves and meristems of kale in
547 comparison to *ELG3*, *GL1*, *GL2* (Table 1; Fig. 5A).
548

549 Another set of candidate genes for the kale domestication syndrome we looked for, are
550 floral transition genes. Floral transition is a major development switch controlled by regulatory
551 pathways, integrating endogenous and environmental cues (Koornneef et al., 2004). We suggest
552 that during kale domestication, delaying of flowering was necessary to maintain leaf production
553 as main consumable. Four main regulatory pathways directly or indirectly involved in flowering
554 have been described in *Arabidopsis thaliana*: photoperiodic, autonomous, vernalization, and
555 gibberellin (Okazaki et al., 2007). The central integrator of flowering signals is the Flowering
556 Locus C (*FLC*) that encodes a transcription factor of the MADS box family. While *A. thaliana*
557 contains only one *FLC* gene, five copies have been described for *B. oleracea* (Okazaki et al.
558 2007). We found two copies of this gene being differentially expressed in kale. One was
559 upregulated (Bo9g173370) and together with Agamous-like MADS-box protein *AGL27/MAF1*
560 (Bo3g100570) have been proposed as floral repressors (Ratcliffe et al., 2001). A second copy of
561 the *FLC* (Bo3g024250) gene was found to be shared by the three morphotypes kale, cabbage and
562 TO1000 (Table 1, Fig. 5A).
563
564 Pathways to domestication in kale may involve selection of plants with the right balance
565 of phytochemicals concentrations. The direction of selection solely towards a reduction in
566 phytochemicals through time was not the case in kale since persistence of pungency and
567 bitterness could have reduced competition from mammals as well as loss to pests, during
568 cultivation and storage (Deham et al., 2020). In *Brassica* species, most glucosinolates are
569 biosynthesized from methionine. A series of copies of the Myrosinase enzyme *TGG1* were
570 differentially expressed and upregulated in kale (Bo00934s010, Bo2g155820, Bo2g155840,
571 Bo2g155870, Bo8g039420) (Zheng et al., 2011). This enzyme breaks down glucosinolates into

572 toxins (isothiocyanates, thiocynates, nitriles, and epithionitriles) upon leaf tissue damage.
573 Therefore, glucosinolates may act as a potent feeding deterrent for generalist insect species, as
574 their toxicity causes developmental and fitness damage (Santolamazza-Carbone 2014, Yi et al.,
575 2015). Gene ontology terms related to glucosinolates and response to herbivores were also found
576 enriched in our analysis specifically in the comparison between kale vs. cabbage terms.
577
578 Several genes were identified related to nutritional content in kale, for example
579 precursors and enzymes related to vitamin C, GDP-L-galactose phosphorylase 2 *VTC5* was
580 upregulated in kale (Bo2g041230) (Duan et al., 2016) and several copies of L-ascorbate oxidase
581 (*ASO*) (Ascorbase) (EC 1.10.3.3) *AAO* (Bo2g165650, Bo9g150050, Bo7g118970) (Fujiwara et
582 al., 2016) were downregulated. Precursors of the vitamin B1 Phosphomethylpyrimidine synthase
583 chloroplastic (EC 4.1.99.17) *THIC* (Bo4g061330) (Wachter et al., 2007). The RuBisCO large
584 subunit-binding protein subunit alpha chloroplastic (60 kDa chaperonin subunit alpha
585 (Bo04491s010) and large subunit-binding protein subunit beta chloroplastic (60 kDa chaperonin
586 subunit beta) (Bo6g034560) (Edelman & Colt, 2016). Differentially expressed genes found in the
587 three morphotypes included Ferredoxin-1 chloroplastic (AtFd1) *LSC30* (Bo8g059760) (Table 1,
588 Fig. 5B).
589
590 **CONCLUSIONS**
591 This study provided leaf developmental analysis and transcriptomics for *Brassica*
592 *oleracea* morphotypes kale, cabagge and TO1000, three economically important Cole crops with
593 distinct plant morphologies and domestication syndromes. RNA-Seq experiments allow the
594 discovery of novel candidate genes like those involved in domestication syndromes. We

595 identified gene expression patterns in *Brassica oleracea* that are shared among two vegetative
596 morphotypes cabbage and kale, and the "rapid-cycling" morphotype TO1000, and estimated the
597 contribution of morphotype-specific gene expression patterns that set kale apart. During kale
598 domestication farmers could have been selecting for different flowering times, taste and
599 herbivore defense from plants in the wild that have different vernalization times, different taste,
600 resistance to herbivores or that flower at different times during development and all variations in
601 forms could have arisen from that.
602

603 **Acknowledgements**

604 The authors thank G. Conant, J. Birchler, M. Liscum, C. Henriquez, M. Tang, A. Reyes and
605 reviewers for valuable comments on the manuscript. The authors acknowledge the following
606 research grants and granting institutions: National Science Foundation (DEB 1146603, DEB
607 1209137), Ministerio de Ciencia y Tecnologia de Colombia (64777) and The Newton Fund, UK-
608 Colombia mobility grant.
609

610 **Author Contributions**

611 TA: Co-developed questions and framework, obtained funding, made collections, performed
612 microscopy analysis, growth chamber experiments, lab work including library sequencing
613 preparation, performed analyses, wrote manuscript. CH: Performed analyses, performed lab
614 work, wrote and edited manuscript. PM: Co-developed questions and framework, provided
615 funding and mentored students. JCP: Co-developed questions and framework, provided funding
616 and mentored students.
617

618 **Data availability**

619 Raw reads and assemblies have been deposited in NCBI:

620 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149483

621 Additional data can be found in Figshare (https://figshare.com) (Suppl. Tables 2-21):

622 https://doi.org/10.6084/m9.figshare.13224383

623 Commands and their main arguments used for the main steps of analyses:

624 https://github.com/niederhuth/The-molecular-basis-of-Kale-domestication

625

626 **Literature cited**

627 Allaby, R. G. (2014). "Domestication Syndrome in Plants." Encyclopedia of Global

628 Archaeology: 2182-2184.

629 Allender, C. J., Allainguillaume, J., Lynn, J., and King, G. J. (2007). "Simple sequence repeats

630 reveal uneven distribution of genetic diversity in chloroplast genomes of *Brassica oleracea* L.

631 and (n = 9) wild relatives." Theoretical and Applied Genetics 114(4): 609-618.

632 Alexa, A. and J. Rahnenfuhrer (2019). "topGO: Enrichment Analysis for Gene Ontology." R

633 package. Available from: https://www.bioconductor.org/packages/release/bioc/html/topGO.html.

634 Andrews S. (2010). "FastQC a Quality Control Tool for High Throughput Sequence Data."

635 Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

636 Anthony, R. G., James, P. E., and Jordan, B. R. (1995). "The cDNA sequence of a cauliflower

637 apetala-1/squamosa homolog." Plant Physiology 108(1): 441-442.

638 Aoyama, T., Dong, C. H., Wu, Y., Carabelli, M., Sessa, G., Ruberti, I., Morelli, G. and Chua, N.
639 H. (1995). Ectopic expression of the *Arabidopsis* transcriptional activator *Athb-1* alters leaf cell
640 fate in tobacco. The Plant Cell, 7(11), 1773-1785.

641 Ashburner M., Ball C. A., Blake J. A., Botstein D., Butler H., Cherry J. M., Davis, A. P., et al.
642 (2000). "Gene ontology: tool for the unification of biology: The Gene Ontology Consortium".
643 Nature Genetics. 25(1):25-9. doi: 10.1038/75556.

644 Bellaoui, M., Pidkowich, M. S., Samach, A., Kushalappa, K., Kohalmi, S. E., Modrusan, Z.,
645 Crosby, W. L. et al. (2001). The Arabidopsis *BELL1* and *KNOX* TALE homeodomain proteins
646 interact through a domain conserved between plants and animals. The Plant Cell, 13(11), 2455-
647 2470.

648 Beltramino, M., Ercoli, M. F., Debernardi, J. M., Goldy, C., Rojas, A. M., Nota, F., Alvarez, M.
649 E., et al. (2018). Robust increase of leaf size by *Arabidopsis thaliana* GRF3-like transcription
650 factors under different growth conditions. *Scientific reports*, *8*(1), 1-13.

651 Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and
652 powerful approach to multiple testing. Journal of the Royal statistical society: series B
653 (Methodological), 57(1), 289-300.

654 Bennett, G. J. (1995). U.S. Patent No. 5,384,526. Washington, DC: U.S. Patent and Trademark
655 Office.

656 Bennett, T., Sieberer, T., Willett, B., Booker, J., Luschnig, C., and Leyser, O. (2006). The
657 Arabidopsis *MAX* pathway controls shoot branching by regulating auxin transport. Current
658 Biology, 16(6), 553-563.

659 Buchfink, B., C. Xie, and D.H. Huson (2014). “Fast and sensitive protein alignment using
660 DIAMOND.” Nature Methods 12: 59.

661 Cardon, G. H., Höhmann, S., Nettesheim, K., Saedler, H., and Huijser, P. (1997). Functional
662 analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral
663 transition. The Plant Journal, 12(2), 367-377.

664 Carmona, D., Lajeunesse, M. J., and Johnson, M. T. (2011). "Plant traits that predict resistance to
665 herbivores." Functional Ecology 25(2): 358-367.

666 Carr, S. M. and V. F. Irish (1997). "Floral homeotic gene expression defines developmental
667 arrest stages in *Brassica oleracea* L. vars. *botrytis* and *italica*." Planta 201(2): 179-188.

668 Carlson, M (2019). “GO.db: A set of annotation maps describing the entire Gene Ontology”. R
669 package. Available from:
670 https://www.bioconductor.org/packages/release/data/annotation/html/GO.db.html.

671 Chen, H., Rosin, F. M., Prat, S., and Hannapel, D. J. (2003). Interacting transcription factors
672 from the three-amino acid loop extension superclass regulate tuber formation. Plant Physiology,
673 132(3), 1391-1404.

674 Cho, H. T., and Cosgrove, D. J. (2000). Altered expression of expansin modulates leaf growth
675 and pedicel abscission in *Arabidopsis thaliana*. Proceedings of the National Academy of
676 Sciences, 97(17), 9783-9788.

677 Cole, M., Nolte, C., AND Werr, W. (2006). Nuclear import of the transcription factor SHOOT
678 MERISTEMLESS depends on heterodimerization with *BLH* proteins expressed in discrete sub-

679 domains of the shoot apical meristem of *Arabidopsis thaliana*. Nucleic Acids Research, 34(4),
680 1281-1292.

681 Consortium, T.U. (2018). "UniProt: a worldwide hub of protein knowledge*."* Nucleic Acids
682 Research 47(D1): D506-D515.

683 Denham, T., Barton, H., Castillo, C., Crowther, A., Dotte-Sarout, E., Florin, S. A., Pritchard, J. et
684 al. (2020). The domestication syndrome in vegetatively propagated field crops. Annals of
685 Botany,125(4), 581-597.

686 Delhomme, N., Padioleau, I., Furlong, E. E., and Steinmetz, L. M. (2012). "easyRNASeq: A
687 bioconductor package for processing RNA-Seq data." Bioinformatics 28(19): 2532-2533.

688 Doebley, J. (1992). "Molecular systematics and crop evolution." Molecular systematics of plants,
689 Springer: 202-222.

690 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. Batut, P., et al. (2013).
691 "STAR: ultrafast universal RNA-seq aligner." Bioinformatics 29(1): 15-21.

692 Dobney, S., Chiasson, D., Lam, P., Smith, S. P., & Snedden, W. A. (2009). The calmodulin-
693 related calcium sensor *CML42* plays a role in trichome branching. Journal of Biological
694 Chemistry, 284(46), 31647-31657.

695 Duan, W., Ren, J., Li, Y., Liu, T., Song, X., Chen, Z., Huang, Z., et al. (2016). Conservation and
696 expression patterns divergence of Ascorbic Acid d-mannose/l-galactose pathway genes in
697 *Brassica rapa*. Frontiers in Plant Science, 7, 778.

698 Edelman, M., and Colt, M. (2016). Nutrient value of leaf vs. seed. Frontiers in Chemistry, 4, 32.

699 Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of
700 plant transcription factors. Trends in Plant Science, 5(5), 199-206.

701 Fujiwara, A., Togawa, S., Hikawa, T., Matsuura, H., Masuta, C., and Inukai, T. (2016). Ascorbic
702 acid accumulates as a defense response to Turnip mosaic virus in resistant *Brassica rapa*
703 cultivars. Journal of Experimental Botany, 67(14), 4391-4402.

704 Gao, M., Li, G., Yang, B., Qiu, D., Farnham, M., and Quiros, C. (2007). "High-density *Brassica*
705 *oleracea* linkage map: Identification of useful new linkages." Theoretical and Applied Genetics
706 115(2): 277-287.

707 Gilding, E. K., and Marks, M. D. (2010). Analysis of purified glabra3-shapeshifter trichomes
708 reveals a role for NOECK in regulating early trichome morphogenic events. The Plant Journal,
709 64(2), 304-317.

710 Grossmann, S., S. Bauer, P. N. Robinson, and M. Vingron. 2007. 'Improved detection of
711 overrepresentation of Gene-Ontology annotations with parent child analysis', Bioinformatics, 23:
712 3024-31.

713 Hammer, K. (1984). "Das Domestikationssyndrom." Kulturpflanze 32: 11-34.

714 Harlan, J. (1992). "Origins and processes of domestication." Grass evolution and domestication
715 159: 175.

716 Higginson, T., Li, S. F., and Parish, R. W. (2003). *AtMYB103* regulates tapetum and trichome
717 development in *Arabidopsis thaliana*. The Plant Journal, 35(2), 177-192.

718 Hodgkin, T. (1995). Cabbages, Kales, etc. Evolution of crop plants. J. Small, N. W. Simmonds.
719 Harlow, Longman group: 76-82.

720 Hong, Z. H., Qing, T., Schubert, D., Kleinmanns, J. A., and Liu, J. X. (2019). BLISTER-
721 regulated vegetative growth is dependent on the protein kinase domain of ER stress modulator
722 *IRE1A* in *Arabidopsis thaliana*. *PLoS genetics*, *15*(12), e1008563.

723 Jang, H. J., Pih, K. T., Kang, S. G., Lim, J. H., Jin, J. B., Piao, H. L., and Hwang, I. (1998).
724 Molecular cloning of a novel Ca 2+-binding protein that is induced by NaCl stress. Plant
725 Molecular Biology, 37(5), 839-847.

726 Kebrom, T. H. (2017). A growing stem inhibits bud outgrowth–the overlooked theory of apical
727 dominance. Frontiers in Plant Science, 8, 1874.

728 Kempin, S. A., Savidge, B., and Yanofsky, M. F.(1995). "Molecular basis of the cauliflower
729 phenotype in *Arabidopsis*." Science 267(5197): 522-525.

730 Kersey, P.J., Julian P., Allen J. E., Allot A., Barba, M., Boddu, S., Bolt, B. J. et al. (2018).
731 "Ensembl Genomes 2018: "An integrated omics infrastructure for non-vertebrate species".
732 Nucleic Acids Research, 2017. 46(D1): p. D802-D808.

733 Kim, J. H., Choi, D., and Kende, H. (2003). The AtGRF family of putative transcription factors
734 is involved in leaf and cotyledon growth in *Arabidopsis*. *The Plant Journal*, *36*(1), 94-104.

735 Khosla, A., Boehler, A. P., Bradley, A. M., Neumann, T. R., and Schrick, K. (2014). HD-Zip
736 proteins GL2 and HDG11 have redundant functions in *Arabidopsis* trichomes, and *GL2* activates
737 a positive feedback loop via *MYB23.* The Plant Cell, 26(5), 2184-2200.

738 Kolde, R. (2019). “pheatmap: Pretty Heatmaps”. R package. Available from: https://CRAN.R-
739 project.org/package=pheatmap.

740 Koornneef, M., et al. (2004). "Naturally occurring genetic variation in *Arabidopsis thaliana*."
741 Annual Review Plant Biology 55: 141-172.

742 Kuijt, S. J., Greco, R., Agalou, A., Shao, J., CJ‘t Hoen, C., Övernäs, E., Osnato, M., et al. (2014).
743 Interaction between the growth-regulating factor and knotted1-like homeobox families of
744 transcription factors [W]. Plant Physiology, 164(4), 1952-1966.

745 Lan, T. H. and A. H. Paterson (2000). "Comparative mapping of quantitative trait loci sculpting
746 the curd of *Brassica oleracea*." Genetics 155(4): 1927-1954.

747 Larsson, A. S., Landberg, K., & Meeks-Wagner, D. R. (1998). The TERMINAL FLOWER2
748 (*TFL2*) gene controls the reproductive transition and meristem identity in *Arabidopsis thaliana*.
749 Genetics, 149(2), 597-605.

750 Ledger, S., Strayer, C., Ashton, F., Kay, S. A., and Putterill, J. (2001). Analysis of the function
751 of two circadian-regulated CONSTANS-LIKE genes. The Plant Journal, 26(1), 15-22.

752 Lee, B. H., Kwon, S. H., Lee, S. J., Park, S. K., Song, J. T., Lee, S., Hwang, Y. sic, et al. (2015).
753 The *Arabidopsis thaliana NGATHA* transcription factors negatively regulate cell proliferation of
754 lateral organs. Plant Molecular Biology, 89(4-5), 529-538.

755 Lian, X. P., Zeng, J., Zhang, H. C., Yang, X. H., Zhao, L., and Zhu, L. Q. (2018). Molecular
756 cloning and expression analysis of *CML27* in *Brassica oleracea* pollination process. Acta
757 Biologica Cracoviensia. Series Botanica, 60(2).

758 Lowman, A. C. and M. D. Purugganan (1999). "Duplication of the *Brassica oleracea*
759 *APETALA1* floral homeotic gene and the evolution of domesticated cauliflower." Journal of
760 Heredity 90(5): 514-520.

761 Love M. I., Huber W., and Anders S. (2014). “Moderated estimation of fold change and
762 dispersion for RNA-seq data with DESeq2”. Genome biology 15(12):550. doi: 10.1186/

763 Lukens, L. N., Quijada, P. A., Udall, J., Pires, J. C., Schranz, M. E., and Osborn, T. C. (2004).
764 "Genome redundancy and plasticity within ancient and recent *Brassica* crop species." Biological
765 Journal of the Linnean Society 82(4): 665-674.

766 Madrid, E., Chandler, J. W., and Coupland, G. (2020). Gene regulatory networks controlled by
767 FLOWERING LOCUS C that confer variation in seasonal flowering and life history. Journal of
768 Experimental Botany.

769 Maggioni, L., von Bothmer, R., Poulsen, G., and Branca, F. (2010). "Origin and domestication of
770 cole crops (*Brassica oleracea* L.): Linguistic and literary considerations." Economic Botany
771 64(2): 109-123.

772 Martin, J., Bruno, V. M., Fang, Z., Meng, X., Blow, M., Zhang, T., Sherlock, G., et al. (2010).
773 "Rnnotator: An automated de novo transcriptome assembly pipeline from stranded RNA-Seq
774 reads." BMC Genomics 11(1).

775 Matsumura, Y., Iwakawa, H., Machida, Y., and Machida, C. (2009). Characterization of genes in
776 the ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES (*AS2/LOB*) family in
777 Arabidopsis thaliana, and functional and molecular comparisons between *AS2* and other family
778 members. The Plant Journal, 58(3), 525-537.

779 Morohashi, K., Zhao, M., Yang, M., Read, B., Lloyd, A., Lamb, R., and Grotewold, E. (2007).
780 Participation of the *Arabidopsis* bHLH factor *GL3* in trichome initiation regulatory events. Plant
781 physiology, 145(3), 736-746

782 Müller, J., Wang, Y., Franzen, R., Santi, L., Salamini, F., and Rohde, W. (2001). In vitro
783 interactions between barley TALE homeodomain proteins suggest a role for protein–protein
784 associations in the regulation of Knox gene function. The Plant Journal, 27(1), 13-23.

785 Nasim, Z., Fahim, M., and Ahn, J. H. (2017). Possible role of MADS AFFECTING
786 FLOWERING 3 and B-BOX DOMAIN PROTEIN 19 in flowering time regulation of
787 *Arabidopsis* mutants with defects in nonsense-mediated mRNA decay. Frontiers in Plant
788 Science, 8, 191.

789 Nayidu, N. K., Tan, Y., Taheri, A., Li, X., Bjorndahl, T. C., Nowak, J., Wishart, D. S., et al.
790 (2014). *Brassica villosa*, a system for studying non-glandular trichomes and genes in the
791 Brassicas. Plant Molecular Biology, 85(4-5), 519-539.

792 Ojangu, E. L., Tanner, K., Pata, P., Järve, K., Holweg, C. L., Truve, E., and Paves, H. (2012).
793 Myosins *XI-K, XI-1*, and *XI-2* are required for development of pavement cells, trichomes, and
794 stigmatic papillae in *Arabidopsis*. BMC plant biology, 12(1), 1-16.

795 Okazaki, K., Sakamoto, K., Kikuchi, R., Saito, A., Togashi, E., Kuginuki, Y., Matsumoto, S., et
796 al.(2007). "Mapping and characterization of FLC homologs and QTL analysis of flowering time
797 in Brassica oleracea." Theoretical and Applied Genetics 114(4): 595-608.

798 Parkin, I.A., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E, Town, C. D., et
799 al.(2014). "Transcriptome and methylome profiling reveals relics of genome dominance in the
800 mesopolyploid *Brassica oleracea*". Genome Biology 15(6): p. R77.

801 Paterson, A. H. (2005). "Polyploidy, evolutionary opportunity, and crop adaptation." Genetica
802 123(1-2): 191-196.

803 Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). "An overview of the gene regulatory
804 network controlling trichome development in the model plant, *Arabidopsis*." Frontiers in Plant
805 Sciences 5: 259.

806 Pelaz, S., Gustafson-Brown, C., Kohalmi, S. E., Crosby, W. L., & Yanofsky, M. F. (2001).
807 APETALA1 and SEPALLATA3 interact to promote flower development. The Plant
808 Journal, 26(4), 385-394.

809 Pérez-Rodríguez, P., Riano-Pachon, D. M., Corrêa, L. G. G., Rensing, S. A., Kersten, B., and
810 Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant
811 transcription factor database. Nucleic Acids Research, 38(suppl_1), D822-D827.

812 Piazza, P., Bailey, C. D., Cartolano, M., Krieger, J., Cao, J., Ossowski, S., Schneeberger, K., et
813 al.(2010). "*Arabidopsis thaliana* leaf form evolved via loss of *KNOX* expression in leaves in
814 association with a selective sweep." Current Biology 20(24): 2223-2228

815 Poza-Carrión, C., Aguilar-Martínez, J. A., and Cubas, P. (2007). Role of TCP gene
816 BRANCHED1 in the control of shoot branching in *Arabidopsis*. Plant signaling &
817 behavior, 2(6), 551-552.

818 Purugganan, M. D., Boyles, A. L., and Suddith, J. I. (2000). "Variation and selection at the
819 CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated *Brassica*
820 *oleracea*." Genetics 155(2): 855-862.

821 Qi, X., An, H., Hall, T. E., Di, C., Blischak, P. D., Pires, J. C., and Barker, M. S. (2019). Genes
822 derived from ancient polyploidy have higher genetic diversity and are associated with
823 domestication in *Brassica rapa.* BioRxiv, 842351.

824 Ratcliffe, O. J., Nadzan, G. C., Reuber, T. L., and Riechmann, J. L. (2001). Regulation of
825 Flowering in *Arabidopsis* by an FLCHomologue. Plant Physiology, 126(1), 122-132.

826 Reed, J. W., Nagpal, P., Bastow, R. M., Solomon, K. S., Dowson-Day, M. J., Elumalai, R. P.,
827 and Millar, A. J. (2000). Independent action of *ELF3* and phyB to control hypocotyl elongation
828 and flowering time. Plant Physiology, 122(4), 1149-1160.

829 Ren, J., Liu, Z., Du, J., Fu, W., Hou, A., and Feng, H. (2019). Fine-mapping of a gene for the
830 lobed leaf, BoLl, in ornamental kale (*Brassica oleracea* L. var. *acephala*). Molecular
831 Breeding, *39*(3), 40.

832 Riaño-Pachón, D. M., Ruzicic, S., Dreyer, I., and Mueller-Roeber, B. (2007). PlnTFDB: an
833 integrative plant transcription factor database. BMC bioinformatics, 8(1), 42.

834 Santolamazza-Carbone, S., et al. (2014). "Bottom-up and top-down herbivore regulation
835 mediated by glucosinolates in *Brassica oleracea* var. *acephala*." Oecologia 174(3): 893-907.

836 Schnittger, A., Folkers, U., Schwab, B., Jürgens, G., and Hülskamp, M. (1999). Generation of a
837 spacing pattern: the role of TRIPTYCHON in trichome patterning in Arabidopsis. The Plant
838 Cell, 11(6), 1105-1116.

839 Schönrock, N., Bouveret, R., Leroy, O., Borghi, L., Köhler, C., Gruissem, W., and Hennig, L.
840 (2006). Polycomb-group proteins repressthe floral activator AGL19 in the FLC-independent
841 vernalization pathway. Genes & development, 20(12), 1667-1678.

842 Sebastian, R. L., Velasco, P., Soengas, P., and Cartea, M. E. (2002). "Identification of
843 quantitative trait loci controlling developmental characteristics of *Brassica oleracea* L."
844 Theoretical and Applied Genetics 104(4): 601-609.

845 Semiarti, E., Ueno, Y., Tsukaya, H., Iwakawa, H., Machida, C., and Machida, Y. (2001). The
846 ASYMMETRIC LEAVES2 gene of *Arabidopsis thaliana* regulates formation of a symmetric
847 lamina, establishment of venation and repression of meristem-related homeobox genes in
848 leaves. Development, 128(10), 1771-1783.

849 Schilmiller, A. L., Koo, A. J., and Howe, G. A. (2007). Functional diversification of acyl-
850 coenzyme A oxidases in jasmonic acid biosynthesis and action. Plant Physiology, 143(2), 812-
851 824.

852 Sivitz, A. B., Hermand, V., Curie, C., and Vert, G. (2012). *Arabidopsis bHLH100* and *bHLH101*
853 control iron homeostasis via a FIT-independent pathway. PloS one, 7(9), e44843.

854 Shigyo, M., Hasebe, M., and Ito, M. (2006). Molecular evolution of the AP2 subfamily. Gene,
855 366(2), 256-265.

856 Smith, L. B. and King, G. J. (2000). "The distribution of BoCAL-a alleles in *Brassica oleracea* is
857 consistent with a genetic model for curd development and domestication of the cauliflower."
858 Molecular Breeding 6(6): 603-613.

859 Smith, H. M., Boschke, I., and Hake, S. (2002). "Selective interaction of plant homeodomain
860 proteins mediates high DNA-binding affinity." Proceedings of the National Academy of
861 Sciences U S A 99(14): 9579-9584.

862 Snogerup, S. (1980). The wild forms of the *Brassica oleraceae* group (2n=18) and their possible
863 relations to the cultivated ones. Brassica crops and wild allies, biology and breeding. S. Tsunoda,
864 K. Hinata, C. Gomez-Campo. Tokyo, Japan Scientific Societies Press: 121-132.

865 Soengas, P., Cartea, M. E., Francisco, M., Sotelo, T., and Velasco, P. (2012). "New insights into
866 antioxidant activity of *Brassica* crops." Food chemistry 134(2): 725-733.

867 Sønderby, I. E., Geu-Flores, F., and Halkier, B. A (2010). "Biosynthesis of glucosinolates–gene
868 discovery and beyond." Trends in Plant Science 15(5): 283-290.

869 Steele, R. P., Geu-Flores, F., and Halkier, B. A. (2012). "Quality and quantity of data recovered
870 from massively parallel sequencing: Examples in asparagales and Poaceae." American Journal of
871 Botany 99(2): 330-348.

872 Stewarta, D. and McDougalla, G. (2012). The Brassicas – An Undervalued Nutritional and
873 Health Beneficial Plant Family. The James Hutton Institute, Invergowrie, and the School of Life
874 Sciences, Heriot-Watt University, Edinburgh, EH14 4AS

875 Stirnberg, P., Chatfield, S. P., and Leyser, H. O. (1999). *AXR1* acts after lateral bud formation to
876 inhibit lateral bud growth in *Arabidopsis.* Plant Physiology, 121(3), 839-847.

877 Tanaka, N., Niikura, S., and Takeda, K. (2009). "Inheritance of cabbage head formation in
878 crosses of cabbage x ornamental cabbage and cabbage x kale." Plant Breeding 128(5): 471-477.

879 Tapia-López, R., García-Ponce, B., Dubrovsky, J. G., Garay-Arroyo, A., Pérez-Ruíz, R. V., Kim,
880 S. H., Acevedo, F., et al. (2008). An AGAMOUS-related MADS-box gene, *XAL1* (*AGL12*),
881 regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. Plant
882 physiology, 146(3), 1182-1192.

883 Team, R. D. C. (2008). "R: A language and environment for statistical computing. R Foundation
884 for Statistical Computing." Available from: http://www.R-project.org/.

885 Tsiantis, M. and Hay, A. (2003). "Comparative plant development: The time of the leaf?" Nature
886 Reviews Genetics 4(3): 169-180.

887 Tsunoda, S. (1980). Eco-physiology of wild and cultivated forms in Brassica and allied genera.
888 Brassica crops and wild allies, biology and breeding. S. Tsunoda, K. Hinata, C. Gomez-Campo.
889 Tokyo, Japan Scientific Societies Press: 109-116.

890 Velasco, P., Cartea, M. E., González, C., Vilar, M., and Ordás, A. (2007). "Factors affecting the
891 glucosinolate content of kale (*Brassica oleracea* acephala group)." Journal of agricultural and
892 food chemistry 55(3): 955-962.

893 Velasco ,P., Francisco, M., Moreno, D. A., Ferreres F., García-Viguera, C., Cartea, M.E. (2011)

894 Phytochemical fingerprinting of vegetable *Brassica oleracea* and *Brassica napus* by

895 simultaneous identification of glucosinolates and phenolics. Phytochemical Analysis

896 Mar-Apr;22(2):144-52.

897 Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., and Breaker, R. R.
898 (2007). Riboswitch control of gene expression in plants by splicing and alternative 3′ end
899 processing of mRNAs. The Plant Cell, 19(11), 3437-3450.

900 Wang, H., Wu, J., Sun, S., Liu, B., Cheng, F., Sun, R., & Wang, X. (2011). Glucosinolate
901 biosynthetic genes in *Brassica rapa*. Gene, 487(2), 135-142.

902 Waters, M. T., Brewer, P. B., Bussell, J. D., Smith, S. M., and Beveridge, C. A. (2012). The
903 Arabidopsis ortholog of rice *DWARF27* acts upstream of *MAX1* in the control of plant
904 development by strigolactones. Plant physiology, 159(3), 1073-1085.

905 Wickham H. (2009). “Eggplot2: Elegant Graphics for Data Analysis”: Springer-Verlag New
906 York; 2009.

907 Yang, C. and Ye, Z. (2013). "Trichomes as models for studying plant cell differentiation." Cell
908 Molecular Life Sciences 70(11): 1937-1948.

909 Yi, G.-E., Robin, A. H. K., Yang, K., Park, J. I., Kang, J. G., Yang, T. J., and Nou, I. S.(2015).
910 "Identification and Expression Analysis of Glucosinolate Biosynthetic Genes and Estimation of
911 Glucosinolate Contents in Edible Organs of *Brassica oleracea* Subspecies." Molecules 20(7):
912 13089-13111.

913 Yi, G. E., Robin, A. H. K., Yang, K., Park, J. I., Hwang, B. H., and Nou, I. S. (2016). Exogenous
914 methyl jasmonate and salicylic acid induce subspecies-specific patterns of glucosinolate
915 accumulation and gene expression in *Brassica oleracea* L. Molecules, 21(10), 1417.

916 Zheng, S. J., Zhang, P. J., van Loon, J. J., and Dicke, M. (2011). Silencing defense pathways in
917 Arabidopsis by heterologous gene sequences from *Brassica oleracea* enhances the performance
918 of a specialist and a generalist herbivorous insect. Journal of Chemical Ecology, 37(8), 818.

919

920

921

922

923

924

925

Brassica oleracea ID	**Gene ID**	**Possible function**	**Sources**
Plant development			
1. Aplical meristem dominance			
Genes differentially expressed in Kale			
Bo3g071570	*TCP18*	Regulation of secondary shoot formation in meritem buds acts downstream of MAX genes	Poza-Carrión et al., 2007
Bo9g130730 Bo7g021280	*REVOLUTA*	Epistatic to TCP18	Poza-Carrión et al., 2007

Bo3g042090	*CYP711A1/MAX1*	Synthesis and perception of mobile carotenoid signal that promotes bud arrest	Bennett et al., 2006
Bo9g035190	*DWARF27*	Acts Upstream of MAX1 in the Control of Plant Development by Strigolactones	Waters et al., 2012
Bo8g102450 Bo8g052710	*AXR1*	Controls shoot branching among other activities	Stirnberg et al., 1999
Differentially expressed genes shared by all morphotypes			
Bo8g115290	*AXR1*	Controls shoot branching among other activities	Stirnberg et al., 1999
Bo4g182790	*PID*	Regulates auxin flow and organ formation	Bennett et al., 1995
2. Leaf development			
Genes differentially expressed only in Kale			
Bo7g041000	*ASL6*	Development of normal leaf shape	Semiarti et al., 2001
Bo3g039420	*NGA1*	Leaf development and regulation of leaf morphogenesis	Lee et al., 2015
Bo3g029830 Bo4g142080	*BEL1/BLH1*	Ovule development, unknown function in vegetative tissues in B. oleracea	Bellaoui et al., 2001
Bo4g186220	*GRF3*	Leaf development, negative regulation of cell proliferation	Beltramino et al., 2018
Bo3g082080	*BLI*	Leaf morphogenesis, positive regulation of cell division, trichome branching and flower development	Hong et al., 2019
Bo4g122960	*GRF4*	Leaf development	Kim et al., 2003
Differentially expressed genes shared by all morphotypes			
Bo4g182790	*PID*	Regulates auxin flow and organ formation	Bennett et al., 1995
Bo5g155120	*HAT5*	Leaf margin development	Aoyama et al., 1995

3. Trichome development			
Bo5g156410 Bo8g104210	*MYB106*	Promote conical cell outgrowth, and in some cases, trichome initiation in diverse plant species	Gilding and Marks 2010
Bo8g091080	*MYB80*	Epidermal cell fate specification, jasmonic acid mediated signaling pathway	Higginson et al., 2003
Bo9g035460	*EGL1*	Epidermal cell fate specification, jasmonic acid mediated signaling pathway	Morohashi et al., 2007
Bo7g090950	*GL1*	Trichome development	Morohashi et al., 2007
Bo6g079990	*GL2*	Probable transcription factor required for correct morphological development and maturation of trichomes. Regulates the frequency of trichome initiation and determines trichome spacing	Morohashi et al., 2007
Bo3g031650 Bo4g030720	*WRKY44*	Regulates trichome development, epidermal cell fate specification	Eulgem et al., 2000
Bo1g051040	*TRY*	Generation of spacing patterns among trichomes	Schnittger et al., 1999
Bo4g039530	*HDG11*	Trichome branching	Khosla et a., 2014
Bo3g162280	*CML42*	Involved in the regulation of trichome branching and herbivore defense	Dobney et al., 2009
Bo5g027850	*XIK*	Trichome branching and morphogenesis	Ojangu et al., 2012
Flowering delayed			
Genes differentially expressed only in Kale			
Bo4g042390	*SPL3*	Adult plants extremely dwarfed with curled-up leaves, vegetative to reproductive phase transition of meristem	Cardon et al., 1997
Bo2g062650	*AP1*	Promotes flower meristem identity	Pelaz et al., 2001

Bo6g119600	*BLH3*	Regulation of timing of transition from vegetative to reproductive phase	Cole et al., 2006
Bo3g100570	*AGL27/MAF1*	Repressor of flowering acting with FLC	Ratcliffe et al., 2001
Bo03856s010	*AGL70/MAF3*	Negative regulation of flowering acting with FLC	Ratcliffe et al., 2001
Bo3g100540	*AGL68/MAF5*	Negative regulator of flower development	Ratcliffe et al., 2001
Bo7g108360	*AGL19/*	Regulation of timing of transition from vegetative to reproductive phase	Schönrock et al., 2006
Bo6g094120	*AGL12/XAL1*	Controls flowering	Tapia-Lopez et al., 2020
Bo9g173370 Bo3g024250	*FLC*	May inhibit flowering and inflorescence growth via a pathway involving GAI and by enhancing FLC expression and repressing FT and LFY	Madrid et al., 2020
Bo9g163720	*COL1*	Regulation of flower development	Ledger et al., 2001
Bo4g024850	*SOC1*	Negative regulation of flower development	Nasim et al., 2017
Bo4g159930	*ELF3*	Interaction between light and the circadian clock ultimately regulating flowering	Reed et al., 2000
Bo5g097000	*FRL4A*	Redundantly with MYR2 as a repressor of flowering and organ elongation under decreased light intensity	https://www.uniprot.org/uniprot/Q9LUV4-1
Differentially expressed genes shared by all morphotypes			
Bo2g013840	*LHP1/TFL2*	Negative regulation of flowering acting with FLC	Larsson et al., 1998
Bo01054s030	*AGL68/MAF5*	Negative regulator of flower development	Ratcliffe et al., 2001
Bo3g038880	*SOC1*	Negative regulation of flower development	Nasim et al., 2017

Defense and taste			
Genes differentially expressed only in Kale			
Bo3g052110	*AOP3*	Controls the composition and amount of aliphatic glucosinolates	Wang et al., 2011
Bo9g175680	*MYB29*	Aliphatic glucosionale production	Kim et al., 2013
Bo2g092580 Bo2g092680	*IGMT1*	Indole glucosinolate metabolic process	Yi et al., 2016
Bo00934s010 Bo2g155820 Bo2g155840 Bo2g155870 Bo8g039420	*TGG1*	Degradation of glucosinolates into toxins that can deter insect herbivores	Zheng et al., 2011
Bo6g040050	*MBP2*	Myrosinase binding protein	Zheng et al., 2011
Bo3g028490	*ACX5*	Defense response to insects	Schilmiller et al., 2007
Differentially expressed genes shared by all morphotypes			
Bo6g095790 Bo8g039530 Bo2g155810 Bo8g063740	*TGG1*	Degradation of glucosinolates into toxins that can deter insect herbivores	Zheng et al., 2011
Bo9g094320	*IGMT4*	Indole glucosinolate metabolic process	Yi et al., 2016
Nutrients			
Genes differentially expressed only in Kale			
Bo2g041230	*VTC5*	Catalyzes a reaction of the Smirnoff-Wheeler pathway, the major route to ascorbate biosynthesis in plants	Duan et al., 2016
Bo7g118970 Bo2g165650 Bo9g150050	*AAO*	Redox system involving ascorbic acid	Fujiwara et al., 2016
Bo4g061330	*THIC*	Thiamin biosynthetic gene source of Vitamin B1	Wachter et al., 2007
Bo6g117890	*CML26*	Calcium as nutrient source	Lian et al., 2018
Bo01178s020	*CML42*	Calcium as nutrient source	Lian et al., 2018

Bo9g009850	*CP1*	Calcium as nutrient source	Jang et al., 1998
Bo4g013080	*bHLH100*	Likely regulates genes involved in the distribution of iron within the plant	Sivitz et al., 2012
Bo04491s010 Bo6g034560	RuBisCO	Plant protein source	Edelman & Colt, 2016
Differentially expressed genes shared by all morphotypes			
Bo3g001360	*LSC30*	Iron source	Edelman & Colt, 2016
Bo2g156090	*FER3*	Iron source	Edelman & Colt, 2016

926

927 **Table 1**. Candidate genes that are differentially expressed in curly kale (red winter kale) and

928 shared among the three morphotypes. Genes involved in plant development, flowering, plant

929 defence responses and taste, nutrition.

930

931

932

933

934

935

936

937

938 **Figures**

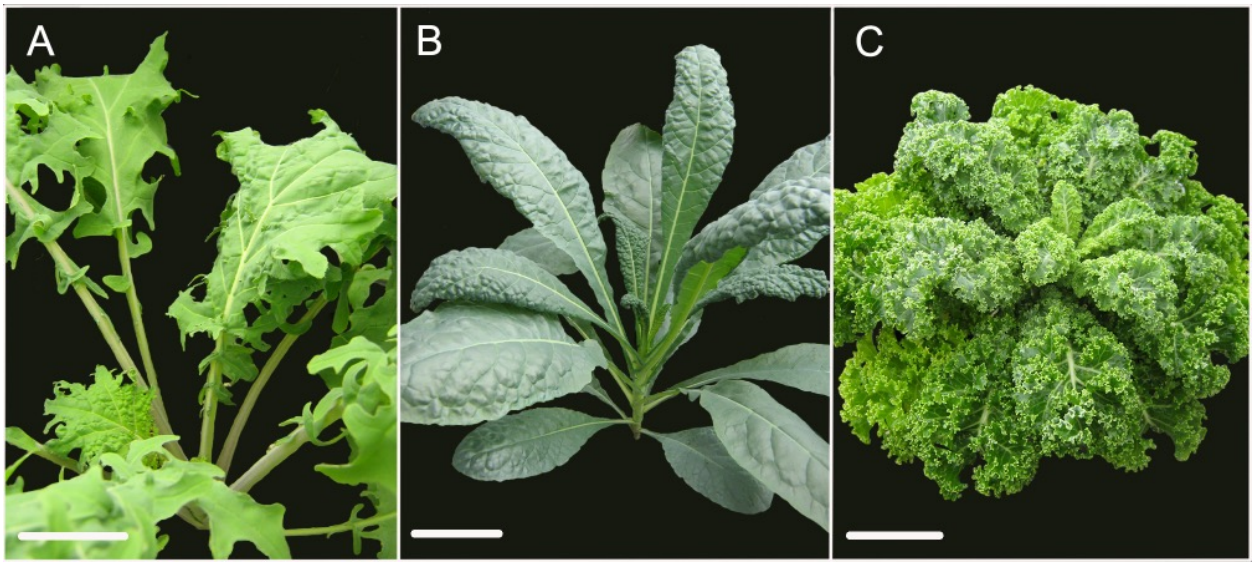
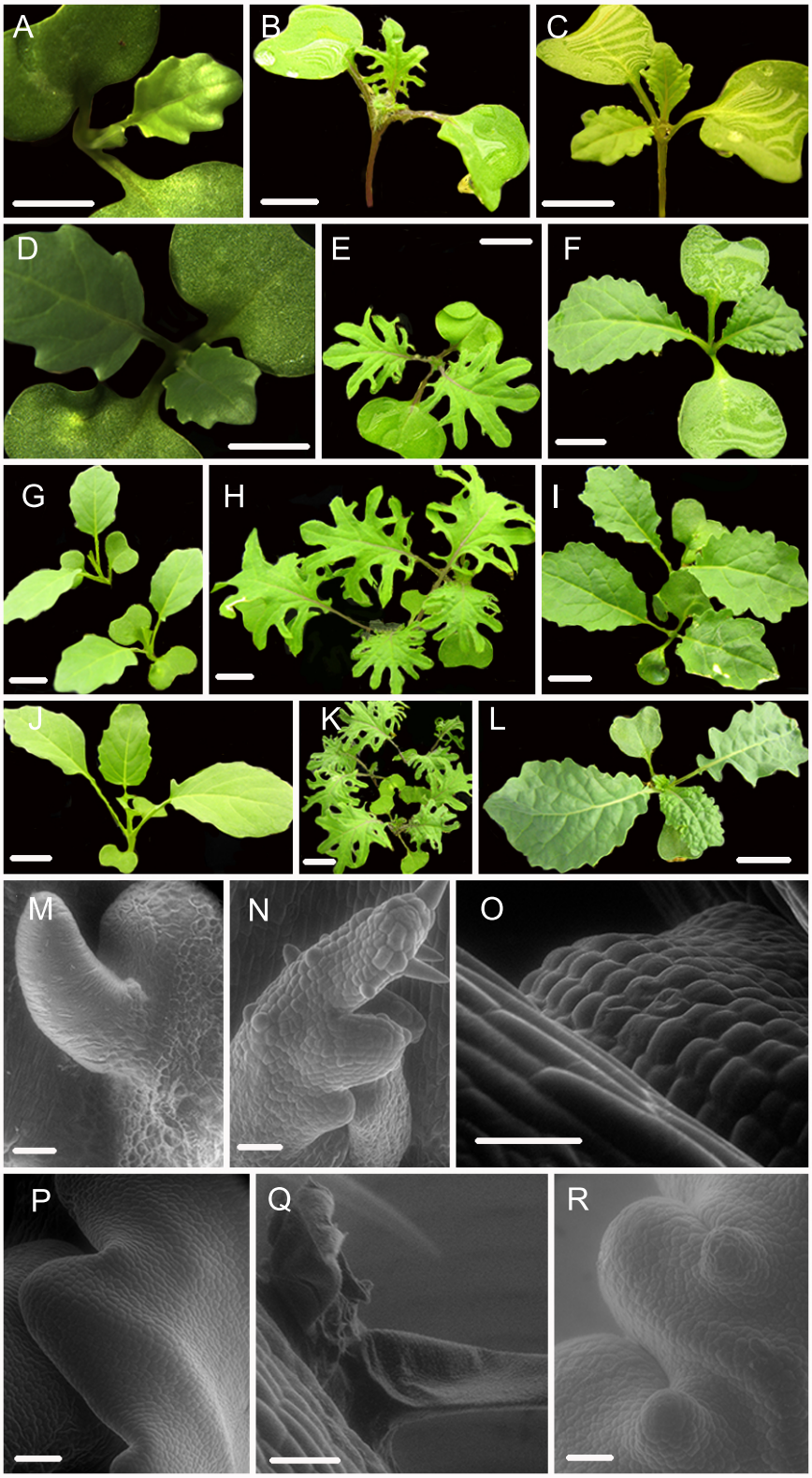


Figure 1. Comparison of mature leaf morphologies within kale varieties. (A) red winter kale (curly kale) with lobed leaf margins, (B) Lacinato Nero Toscana kale (non-curly kale) with crenulate leaf margins, (C) parsley kale with crisped leaf margins. Scale bar = 10 cm



945

946

947 **Figure 2.** Leaf shape development in different Kale varieties. A-L: progressive development of
948 the different Kale varieties and the rapid cycling TO1000 control (stereoscope and regular pics),
949 scale bar = 1 cm. M-R: Scanning Electron Microscopy (SEM) of apical meristem, primordial
950 leaves and details of lobes in different kale varieties and TO1000 control, scale bar = 5 microns.
951 A, D, G, J: leaf development series for TO1000. B, E, H, K: leaf development series for curly
952 kale (red winter kale). C, F, I, L: leaf development series for non-curly kale. M, P: SEM of
953 TO1000 leaf primordia and margin formation detail. N: SEM of curly kale leaf primordial.
954 O: SEM of non-curly kale (Lacinato Nero Toscana kale) leaf primordia. Q: SEM of ectopic
955 growing of leaf blade in mature leaves subtended by a trichome. R: Non-curly kale (Lacinato
956 Nero Toscana kale) margin formation detail.
957
958
959
960
961
962
963
964
965
966
967
968
969

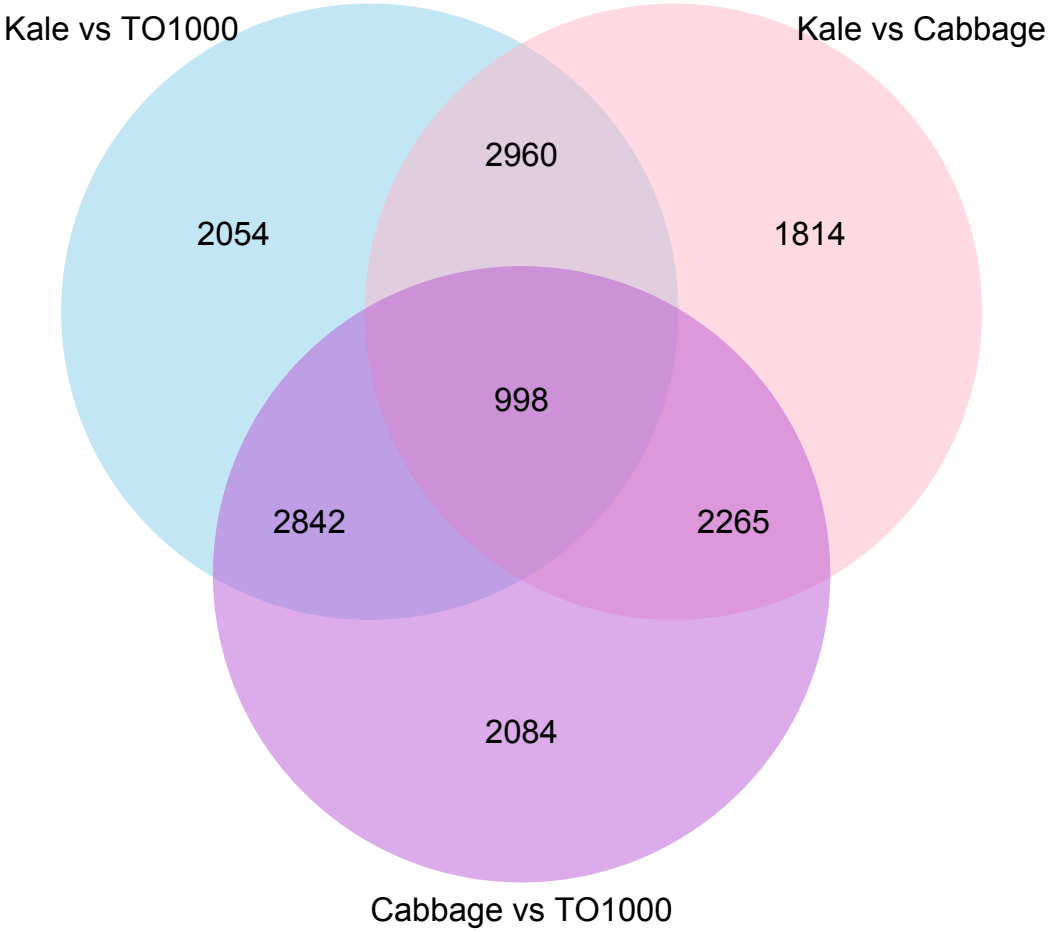
970

971

A. **Differentially Expressed Genes**

Comparison	Higher Expressed DEGs	Lower Expressed DEGs	Total DEGs
Kale vs T01000	4599	4255	8854
Kale vs Cabbage	4704	3333	8037
Cabbage vs TO1000	3406	4783	8189

B.



972

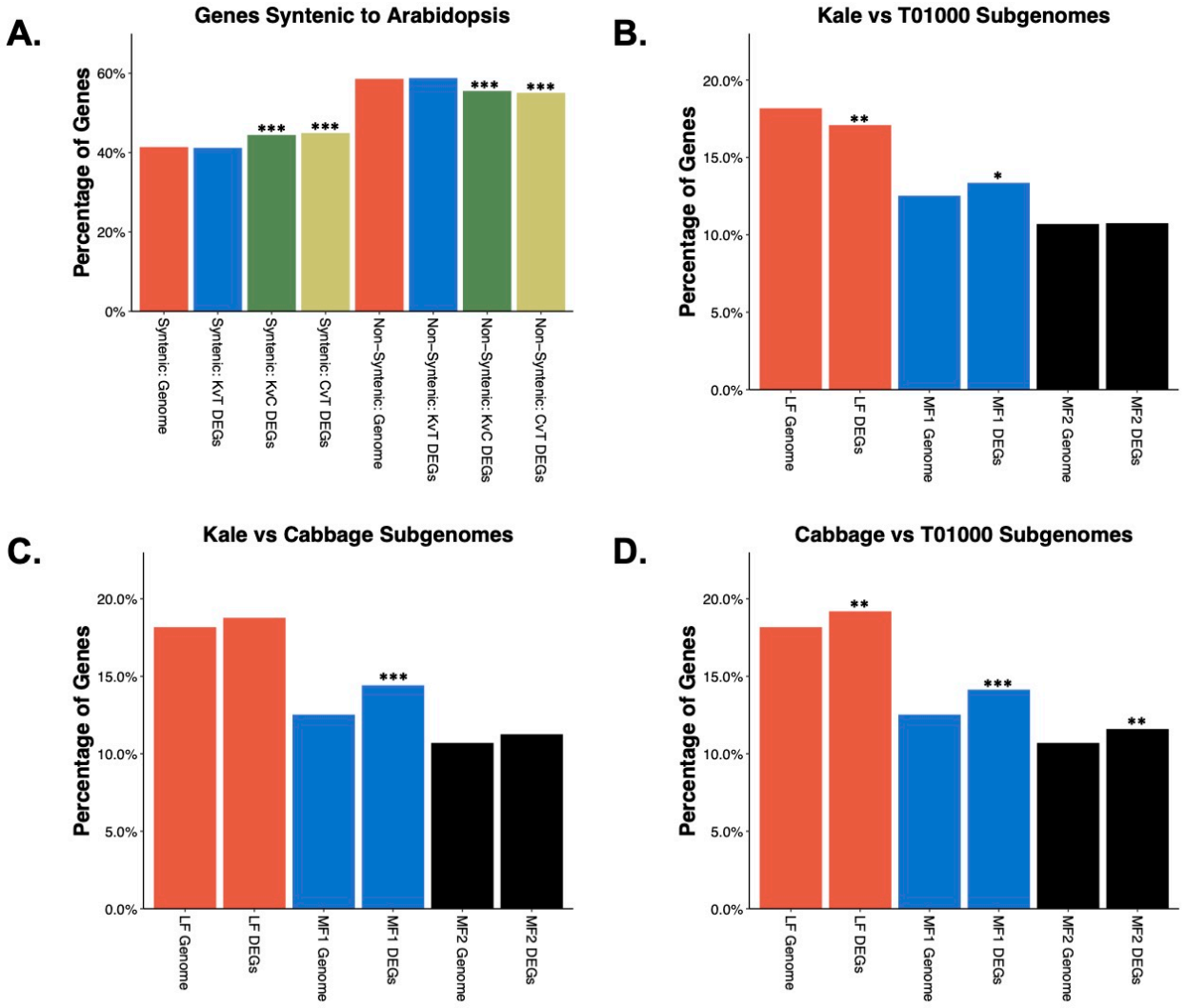
973 **Figure 3.** Number of genes differentially expressed in this study after comparing three different

974 *Brassica oleracea* morphotypes (cabbage, kale and TO1000). A. Number of genes differentially

975 expressed. B. Venn Diagram showing shared genes after doing all possible comparisons between

976 three different *B. oleracea* morphotypes

977



978

979

980

981 **Figure 4.** Syntenic genes from the three subgenomes are enriched in Differentially Expressed

982 Genes (DEGs).

983

984

985

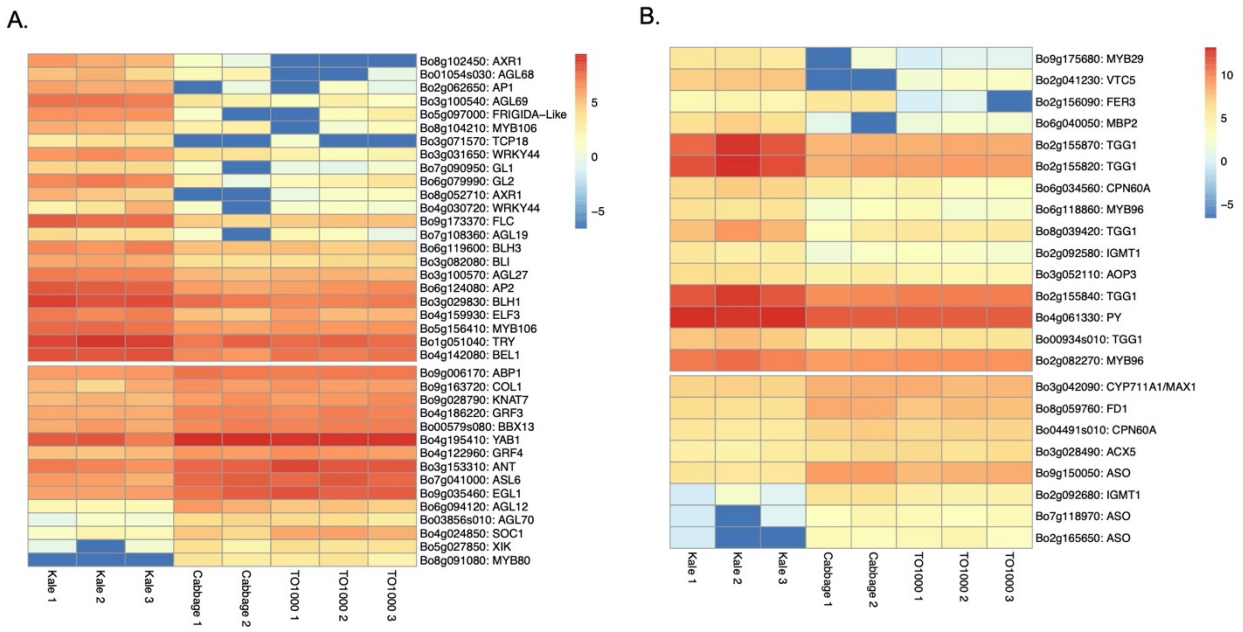


Figure 5. Heatmap of differentially expressed genes for curly kale (red winter kale). Up and down-regulated genes involved in A. plant and leaf development, and flowering B. plant defence responses and taste, nutrition.

Brassica oleracea ID	Gene ID	Possible function	Sources
Plant development			
1. Aplical meristem dominance			
Genes differentially expressed in Kale			
Bo3g071570	*TCP18*	Regulation of secondary shoot formation in meritem buds acts downstream of MAX genes	Poza-Carrión et al., 2007
Bo9g130730 Bo7g021280	*REVOLUTA*	Epistatic to TCP18	Poza-Carrión et al., 2007
Bo3g042090	*CYP711A1/MAX1*	Synthesis and perception of mobile carotenoid signal that promotes bud arrest	Bennett et al., 2006
Bo9g035190	*DWARF27*	Acts Upstream of MAX1 in the Control of Plant Development by Strigolactones	Waters et al., 2012
Bo8g102450 Bo8g052710	*AXR1*	Controls shoot branching among other activities	Stirnberg et al., 1999
Differentially expressed genes shared by all morphotypes			
Bo8g115290	*AXR1*	Controls shoot branching among other activities	Stirnberg et al., 1999
Bo4g182790	*PID*	Regulates auxin flow and organ formation	Bennett et al., 1995
2. Leaf development			
Genes differentially expressed only in Kale			
Bo7g041000	*ASL6*	Development of normal leaf shape	Semiarti et al., 2001
Bo3g039420	*NGA1*	Leaf development and regulation of leaf morphogenesis	Lee et al., 2015
Bo3g029830 Bo4g142080	*BEL1/BLH1*	Ovule development, unknown function in vegetative tissues in B. oleracea	Bellaoui et al., 2001
Bo4g186220	*GRF3*	Leaf development, negative regulation of cell proliferation	Beltramino et al., 2018

Bo3g082080	*BLI*	Leaf morphogenesis, positive regulation of cell division, trichome branching and flower development	Hong et al., 2019
Bo4g122960	*GRF4*	Leaf development	Kim et al., 2003
Differentially expressed genes shared by all morphotypes			
Bo4g182790	*PID*	Regulates auxin flow and organ formation	Bennett et al., 1995
Bo5g155120	*HAT5*	Leaf margin development	Aoyama et al., 1995
3. Trichome development			
Bo5g156410 Bo8g104210	*MYB106*	Promote conical cell outgrowth, and in some cases, trichome initiation in diverse plant species	Gilding and Marks 2010
Bo8g091080	*MYB80*	Epidermal cell fate specification, jasmonic acid mediated signaling pathway	Higginson et al., 2003
Bo9g035460	*EGL1*	Epidermal cell fate specification, jasmonic acid mediated signaling pathway	Morohashi et al., 2007
Bo7g090950	*GL1*	Trichome development	Morohashi et al., 2007
Bo6g079990	*GL2*	Probable transcription factor required for correct morphological development and maturation of trichomes. Regulates the frequency of trichome initiation and determines trichome spacing	Morohashi et al., 2007
Bo3g031650 Bo4g030720	*WRKY44*	Regulates trichome development, epidermal cell fate specification	Eulgem et al., 2000
Bo1g051040	*TRY*	Generation of spacing patterns among trichomes	Schnittger et al., 1999
Bo4g039530	*HDG11*	Trichome branching	Khosla et a., 2014

Bo3g162280	*CML42*	Involved in the regulation of trichome branching and herbivore defense	Dobney et al., 2009
Bo5g027850	*XIK*	Trichome branching and morphogenesis	Ojangu et al., 2012
Flowering delayed			
Genes differentially expressed only in Kale			
Bo4g042390	*SPL3*	Adult plants extremely dwarfed with curled-up leaves, vegetative to reproductive phase transition of meristem	Cardon et al., 1997
Bo2g062650	*AP1*	Promotes flower meristem identity	Pelaz et al., 2001
Bo6g119600	*BLH3*	Regulation of timing of transition from vegetative to reproductive phase	Cole et al., 2006
Bo3g100570	*AGL27/MAF1*	Repressor of flowering acting with FLC	Ratcliffe et al., 2001
Bo03856s010	*AGL70/MAF3*	Negative regulation of flowering acting with FLC	Ratcliffe et al., 2001
Bo3g100540	*AGL68/MAF5*	Negative regulator of flower development	Ratcliffe et al., 2001
Bo7g108360	*AGL19/*	Regulation of timing of transition from vegetative to reproductive phase	Schönrock et al., 2006
Bo6g094120	*AGL12/XAL1*	Controls flowering	Tapia-Lopez et al., 2020
Bo9g173370 Bo3g024250	*FLC*	May inhibit flowering and inflorescence growth via a pathway involving GAI and by enhancing FLC expression and repressing FT and LFY	Madrid et al., 2020
Bo9g163720	*COL1*	Regulation of flower development	Ledger et al., 2001
Bo4g024850	*SOC1*	Negative regulation of flower development	Nasim et al., 2017

Bo4g159930	*ELF3*	Interaction between light and the circadian clock ultimately regulating flowering	Reed et al., 2000
Bo5g097000	*FRL4A*	Redundantly with MYR2 as a repressor of flowering and organ elongation under decreased light intensity	https://www.uniprot.org/uniprot/Q9LUV4-1
Differentially expressed genes shared by all morphotypes			
Bo2g013840	*LHP1/TFL2*	Negative regulation of flowering acting with FLC	Larsson et al., 1998
Bo01054s030	*AGL68/MAF5*	Negative regulator of flower development	Ratcliffe et al., 2001
Bo3g038880	*SOC1*	Negative regulation of flower development	Nasim et al., 2017
Defense and taste			
Genes differentially expressed only in Kale			
Bo3g052110	*AOP3*	Controls the composition and amount of aliphatic glucosinolates	Wang et al., 2011
Bo9g175680	*MYB29*	Aliphatic glucosionale production	Kim et al., 2013
Bo2g092580 Bo2g092680	*IGMT1*	Indole glucosinolate metabolic process	Yi et al., 2016
Bo00934s010 Bo2g155820 Bo2g155840 Bo2g155870 Bo8g039420	*TGG1*	Degradation of glucosinolates into toxins that can deter insect herbivores	Zheng et al., 2011
Bo6g040050	*MBP2*	Myrosinase binding protein	Zheng et al., 2011
Bo3g028490	*ACX5*	Defense response to insects	Schilmiller et al., 2007
Differentially expressed genes shared by all morphotypes			
Bo6g095790 Bo8g039530 Bo2g155810 Bo8g063740	*TGG1*	Degradation of glucosinolates into toxins that can deter insect herbivores	Zheng et al., 2011
Bo9g094320	*IGMT4*	Indole glucosinolate metabolic process	Yi et al., 2016

Nutrients			
Genes differentially expressed only in Kale			
Bo2g041230	*VTC5*	Catalyzes a reaction of the Smirnoff-Wheeler pathway, the major route to ascorbate biosynthesis in plants	Duan et al., 2016
Bo7g118970 Bo2g165650 Bo9g150050	*AAO*	Redox system involving ascorbic acid	Fujiwara et al., 2016
Bo4g061330	*THIC*	Thiamin biosynthetic gene source of Vitamin B1	Wachter et al., 2007
Bo6g117890	*CML26*	Calcium as nutrient source	Lian et al., 2018
Bo01178s020	*CML42*	Calcium as nutrient source	Lian et al., 2018
Bo9g009850	*CP1*	Calcium as nutrient source	Jang et al., 1998
Bo4g013080	*bHLH100*	Likely regulates genes involved in the distribution of iron within the plant	Sivitz et al., 2012
Bo04491s010 Bo6g034560	RuBisCO	Plant protein source	Edelman & Colt, 2016
Differentially expressed genes shared by all morphotypes			
Bo3g001360	*LSC30*	Iron source	Edelman & Colt, 2016
Bo2g156090	*FER3*	Iron source	Edelman & Colt, 2016