To be submitted to: *Am. J. Bot.* [3.03 IF] (*Research Article*)

Running head: ***Kale domestication***

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

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

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

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

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

[4] Both authors contributed equally to this manuscript

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

Manuscript received _____; revision accepted _____.

Short title: Kale transcriptomics

ABSTRACT [237 words - 250 max]

Premise of the study: 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.

Methods: We used a combination of developmental studies and RNAseq experiments 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.

Key Results: We identified gene expression patterns that differ among morphotypes, and estimate the contribution of morphotype-specific gene expression that sets kale apart (2378 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.

Conclusions: 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 (Maggioni et al., 2010). They include dwarf plants with a main stem that has highly compressed internodes (cabbages); plants with an elongated main stem in which the lateral branches are highly compressed due to very short internodes (Brussels sprouts); plants with proliferation of floral meristems (broccoli) or proliferation of aborted floral meristems (cauliflower); and plants with swollen stems (kohlrabi, Marrow-stem kale) or ornate leaf patterns (kales) (Hodgkin, 1995; Lan and Paterson, 2000). It is not so easy to define which domestication traits are common to all of the above-mentioned types, as opposed to the wild species. At early stages in domestication, plants must have been selected for less bitter-tasting, less fibrous, thicker stems, and more succulent storage organs (Maggioni et al., 2010).

Research has provided evidence that polyploidy and genetic redundancy contribute significantly to the phenotypic variation observed among these crops (Lukens et al., 2004; Paterson, 2005). The genetic basis underlying the phenotypic diversity present in Cole crops has been addressed only for some of these morphotypes. The cauliflower phenotype of *Arabidopsis* is comparable to the inflorescence development in the cauliflower *B. oleracea* morphotype. It has been shown that the gene *CAULIFLOWER* is responsible for the phenotype observed in both species (Kempin et al., 1995; Carr & Irish, 1997; Purugganan et al., 2000; Smith and King, 2000). Also, a large QTL for the curd of cauliflower has been found (Lan and Paterson, 2000) indicating that 86 genes are controlling eight curd-related traits, demonstrating this phenotype is

under complex genetic control. Other studies have found and confirmed *BoAP-a1* to be involved in curd formation in cauliflower (Anthony et al., 1995; Carr and Irish, 1997; Gao et al., 2007) and have shown that morphologies in this cultivar are under complex genetic control (Sebastian et al., 2002). The molecular bases of broccoli have also been proposed as involving changes in three codons in the *BoCAL* and *BoAP* genes (Carr and Irish, 1997; Lowman and Purugganan, 1999). A total of five QTLs have been detected for leaf traits in Brussel sprouts, but no significant QTLs have been found for axillary buds (Sebastian et al., 2002). Head formation in cabbage has been suggested as additively inherited (Tanaka et al., 2009). In kale, a recent study suggest that the lobed-leaf trait is quantitatively inherited (Ren et al., 2019).

Apical dominance is manifest in kale varieties, involving selection for erect plants with fewer side branches, and more compact plants, which allows more plants to fit into each unit of cultivated soil (Deham et al., 2020). In kale, leaves are green to purple and do not form a compact head like in cabbage. Crop varieties have surprising variation in leaf size, leaf type, and margin curliness. Some also have an over-proliferation of leaf blade tissues. Kale leaf curliness, blade texture, and tissue overgrowth are some of the more variable morphological characters among varieties.

Leafy kales are considered the earliest cultivated brassicas, originally used for both livestock feed and human consumption (Maggioni et al., 2010) due to leaves and floral buds with great nutritional value (Velasco et al., 2011). Kale leaves also have antioxidative properties, and have been shown to reduce cholesterol levels (Velasco et al., 2007; Soengas et al., 2012). Their high content of beta-carotene, vitamin K, vitamin C, and calcium make it one of the most

nutritious vegetables available for human consumption (Stewarta and McDougalla, 2012). Kale is a source of two carotenoids: lutein and zeaxanthin, and contains sulforaphane, a compound suggested to have anti-cancer properties (Velasco et al., 2007). Kale leaves also have mustard oils produced from glucosinolates. These natural chemicals most likely contribute to plant defense against pests and diseases, and impart a pungent flavor property characteristic of all cruciferous vegetables (Sonderby et al., 2010; Carmona et al., 2011). Definitions of domestication syndrome in kale include traits affecting cultivation, harvesting, and cooking – and its uses – including food, medicines, toxins and fibres (Deham et al., 2020).

Here we used a combination of developmental studies and transcriptomes to understand the vegetative domestication syndrome of kale. Comparisons of kale morphotypes with the "rapid-cycling" morphotype TO1000 and cabbage allow us to identified transcriptome-wide differences that set kale apart. RNA-seq of meristems and immature leaves for the *B. oleracea* vegetative morphotypes, cabbage, kale, and TO1000 reveal gene expression patterns unique to kale and associated developmental and metabolic processes. This allowed us to identify a set of candidate genes we suggest may be important in the kale domestication syndrome.

MATERIALS AND METHODS

Plant Material and Experimental Design

Single accessions of cabbage (*B. oleracea* var. *sabauda*), kale (*B. oleracea* var. *viridis*) and the rapid cycling TO1000 (*B. oleracea* var. *alboglabra*) were grown in an environmental

chamber under uniform conditions: 10hr day (light 7AM to 5PM, Dark 5:01PM to 6:59AM daily) and watered every other day. Morphotypes were planted in a completely randomized block design. Initially, two seeds per pot were planted, then only one was picked (32 plants per morphotype total). A random number generator was used to move plants and flats around every other day. Structural features of the meristems and developing leaves were observed using an Environmental Scanning Electron Microscopy (ESEM) and light microscopy. Images were processed with Adobe (San Jose, California, USA) Photoshop, version 7.0.

Tissues were collected 24 days after seeds were planted. The upper portion of the stems, including immature leaves, and apical and lateral meristems were flash frozen in liquid nitrogen. The assumption was made that genes involved in developing different morphologies should be expressed in younger leaves, shoots, and meristems. Three biological replicates per morphotype were collected, each replicate consisting of pooled tissue from eight plants.

RNA Extraction and cDNA Synthesis

Lysis and homogenization of fresh soft tissue were performed before RNA extraction; PureLink Micro-to-Midi Total RNA Purification System (Invitrogen, Cat. No 12183-018, Carlsbad, California, USA) was used. RNA extraction was done using the Purelink RNA Mini Kit (Ambion, Cat. No12183-018A, Carlsbad, California, USA). DNase was not used before proceeding with cDNA synthesis as recommended in the protocol. For cDNA synthesis MINT (Evrogen, Cat. No SK001, Moscow, Russia) was used, this kit enzymes synthesize full-length-enriched double stranded cDNA from a total polyA$^+$ RNA. The cDNA was then amplified by 16

-- 21 cycles of polymerase chain reaction and purified using a PCR Purification Kit (Invitrogen K3100-01 Carlsbad, California, USA).

Library Construction and Illumina Sequencing

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

Oligonucleotides used as adapters were: Cabbage (1: ACGT, 2: GCTT, 3: TGCT), kale (1: TACT, 2:ATGT, 3:GTAT), TO1000 (1:CTGT, 2:AGCT, 3:TCAT). To prepare adapters, equal volumes were combined of each adapter at 100 μmol/L floating them in a beaker of boiling water for 30 min and snap cooling them on ice. Ligation products were run on a 2% low-melt agarose gel and size-selected for ~300 bp using an x-tracta Disposable Gel Extraction Tool (Scientific, Ocala Florida, USA). Fragments were enriched using PCR in 50 μL volumes containing 3 μL of ligation product, 20 μL of ddH2O, 25 μL master mix (from NEB kit), and 1 μL each of a 25 μmol/L solution of each forward and reverse primer (forward 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC* T-3′; reverse 5′-CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC*-3′; both HPLC purified). Thermal cycle routine was as 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,

with a final extension step of 72°C for 5 min. Products were run on a 2% low-melt agarose gel and the target product was excised and purified with the tools described above. DNA was purified at each step during the end repair, adapter ligation, size selection and fragment enrichment with either a Gel Extraction kit (Qiagen, Germantown Maryland, USA) or a DNA QIAquick PCR Purification kit (Qiagen, Germantown Maryland, USA). The three replicates per morphotype were ran on one third of a lane with single-end 98-bp reads on an Illumina GAIIx Genome Analyzer using the Illumina Cluster Generation Kit v2-GA II, Cycle Sequencing Kit v3 (Illumina, San Diego, USA), and image analysis using Illumina RTA 1.4.15.0 (Illumina, San Diego, USA) at University of Missouri DNA Core.

Data quality control and preprocessing, alignment and differential expression analysis

Reads were first trimmed for the end quality using Cutadapt v2.4 (Martin et al., 2011) and read quality assessed using FastQC (Andrews, 2010). These were then aligned against the *Brassica oleracea* TO1000 genome (Parkin et al., 2014) downloaded from EnsemblPlants 44 (Kersey et al., 2018) with two passes of STAR v2.7.1a (Dobin et al., 2013). In the first pass, reads for each sample were aligned for splice-junction discovery, in addition to those junctions provided in the TO1000 annotations. Splice junctions from each sample were then provided as a list for a second final alignment and read counts calculated for each gene using STAR. Mapping statistics are provided in Supplementary Table 1. Raw sequencing data and unnormalized read counts are publicly available through the Gene Expression Omnibus accession GSE149483. Read counts were imported into R v3.6.3 (Team R) and differential expression was determined using DESeq2 v1.26.0 (Love et al., 2014). Specifically, read counts were normalized, with the TO1000 samples as reference, and fitted to a parametric model. We looked for differentially

expressed genes (DEGs) for each pairwise comparison (kale vs TO1000, kale vs cabbage, and cabbage vs TO1000). Comparisons were done to ultimately separate unique kale-morphotype DEGs. Since each pairwise comparison increases the potential number of false positives, these were combined into a single table and the false discovery rate (FDR) (Benjamini and Hochberg, 1995) was adjusted for the entire table. Genes with an FDR < 0.05 and Log-fold change > 1 or < -1 were considered as differentially expressed genes (DEGs). Plots were made using the R packages ggplot2 (Wickham, 2009), pheatmap (Kolde, 2019), and a VennDiagram (Delhomme, 2018). Transcription-associated proteins (TAPs) were identified using the approach described for the construction of PlnTFDB v3.0 (http://plntfdb.bio.uni-potsdam.de) (Riaño-Pachón et al., 2007; Pérez-Rodríguez et al., 2010). All code for bioinformatic analysis and original figures pdfs and tables are available on GitHub (https://github.com/niederhuth/The-molecular-basis-of-Kale-domestication).

Gene annotation, GO, and KEGG analysis

Gene ontology (GO) terms (Ashburner et al., 2000) are not currently available for the TO1000 assembly. In order to annotate TO1000 gene models with GO terms, TO1000 protein sequences were Blasted (blastp) against protein sequences from UniProtKB Swiss-Prot (release 2019_05) (Consortium 2018) using Diamond (Buchfink et al., 2014), with an e-value cutoff of $1e^{-5}$ and a max of 1 hit per query. UniProt IDs were then used to extract GO terms and mapped to the corresponding TO1000 gene model. In total, 35,422 genes (out of 59,225 total, ~59.8%) had a GO term assigned. GO term enrichment was performed using topGO v 2.38.1 (Alexa and Rahnenfuhrer, 2019) and GO.db v3.10.0 (Carlson, 2019) with the parent-child algorithm (Grossman et al., 2007) and Fisher's Exact Test. A minimum node size of 5 was required, i.e.

each GO term required a minimum of 5 genes mapping to it in ordered to be considered. P-values were adjusted for FDR and considered significant with an adjusted p-value < 0.05.

RESULTS

Kale leaf development

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

leaves were thinner with sinuate margins. In summary, we observed kale margins were more lobed than TO1000 (Fig. 2).

RNA-seq and differential gene expression

To identify genes that might be associated with domestication traits including leaf development in curly-kale, RNA-seq was used to assess transcriptional differences between curly-kale, cabbage, and TO1000. Three biological replicates each were sequenced, however, one of the cabbage replicates was not used for assembly because very few reads of low quality were obtained after sequencing. Between 9 to 25 million trimmed reads were obtained per sample, with ~75-89% reads mapping uniquely to the TO1000 *B. oleracea* genome (Suppl. Table 1). After normalization with DESeq2, hierarchical clustering and principle component analysis showed that samples were grouped by morphotype, with greater variance between morphotypes than between replicates (Suppl. Fig. 1).

Differentially expressed genes (DEGs) were called for all pair-wise comparisons between morphotypes (kale vs TO1000, kale vs cabbage, cabbage vs TO1000), with the aim of identifying genes that were differentially expressed in kale compared to other morphotypes. Genes were considered differentially expressed if they had a two-fold change in expression and a p-value < 0.05 after correction for multiple testing. In total, 7274 DEGs were identified between kale and TO1000 (3817 higher expressed in kale, 3457 lower expressed in kale), 6322 DEGs between kale and cabbage (3600 higher expressed in kale, 2722 lower expressed in kale), and 6316 DEGs between cabbage and TO1000 (2775 higher expressed in cabbage, 3541 lower expressed in cabbage) (Fig. 3A). There were a total of 3010 DEGs shared between the kale-

TO1000 and kale-cabbage comparisons, 2378 found only in the kale comparisons and 632 found in all three comparisons (Fig. 3B). A word cloud analysis for differentially expressed genes unique to kale among leaves and meristematic tissues, emphasized the predominant role of chloroplast and mitochondrial genes and several types of enzymes (Fig. 4)

Polyploidy and its contribution to the kale phenotype

We explored how each of the three subgenomes resulting from an ancient whole genome triplication contributed to differential gene expression between morphotypes. Syntenic genes were significantly overrepresented amongst DEGs in all three comparisons, while non-syntenic genes were underrepresented (Fig. 5A). The gene content of the three sub-genomes of *B. oleracea* have been previously reconstructed (Parking et al., 2014) and using this, the representation of each sub-genome amongst DEGs was tested. For the kale vs TO1000 comparison, both the MF1 and MF2 subgenomes were significantly overrepresented in DEGs (MF1: Odds Ratio = ~1.14, p-value < 0.01; MF2: Odds Ratio = ~1.10, p-value < 0.01), while the LF subgenome was not (LF: Odds Ratio = ~0.98, p-value = ~0.72) (Fig. 5B). All three subgenomes were significantly overrepresented in DEGs in both the kale vs cabbage (LF: Odds Ratio = ~1.15, p-value < 0.01; MF1: Odds Ratio = ~1.30, p-value < 0.01; MF2: Odds Ratio = ~1.15, p-value < 0.01) and cabbage vs TO1000 (LF: : Odds Ratio = ~1.26, p-value < 0.01; MF1: Odds Ratio = ~1.26, p-value < 0.01; MF2: Odds Ratio = ~1.29, p-value < 0.01) comparison (Fig. 5 C,D).

Gene ontology and KEGG pathway enrichment analysis

DEGs for each pair-wise comparison and the list of DEGs shared between the two kale comparisons (3010 genes) were separated into lists of higher and lower-expressed genes and then tested for enrichment of GO (Gene Ontology) terms (Suppl. Tables 5-12) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Suppl. Tables 13-20). We focus here specifically on those kale comparisons to better understand the kale domestication syndrome, in particular GO terms in the biological process (BP) category, as this includes developmental terms. GO terms and KEGG pathways related to the cabbage vs TO1000 comparison are provided as supplement.

Kale vs. TO1000- There were 104 (65 BP) and 34 (10 BP) GO terms enriched in the higher (Suppl. Table 5) and lower (Suppl. Table 6) expressed DEGs, respectively. Higher expressed genes were characterized by a number of metabolic processes, especially fatty acid derivative and nitrile biosynthesis. There were also multiple GO terms related to photosynthesis and plastid organization and environmental responses to both biotic and abiotic stresses (Suppl. Fig. 2A). Lower expressed genes were characterized by processes involved in ‘nucleoside metabolism’, ‘ribonucleoprotein complex biogenesis’, and ‘carbohydrate derivative metabolism’ (Suppl. Figure 2B). Seven KEGG pathways were enriched in higher expressed genes (Suppl. Table 13) and largely match metabolic pathways identified in the GO term analysis. These include processes related to ‘photosynthesis’, ‘cutin, suberine, and wax biosynthesis’, ‘fatty acid elongation’, ‘flavonoid biosynthesis’, ‘glyoxylate and dicarboxylate metabolism’, and ‘other glycan degradation’. Lower expressed genes were enriched in the spliceosome pathway (Suppl. Table 14).

Kale vs. cabbage- Higher expressed genes had 33 (22 BP) enriched GO terms (Suppl. Table 7), while lower expressed genes (Suppl. Table 8) had 17 (4 BP). Similar to the kale vs TO1000 comparison, there was enrichment for processes involved in 'fatty acid metabolism', 'nitrile metabolism', and 'photosynthesis' (Suppl. Figure 2C). There was also enrichment for environmental responses, specifically responses to herbivores. Additionally, there was enrichment for processes in 'carbon fixation and carbohydrate metabolism', 'amylopectin biosynthesis', 'vacuole organization', and the 'negative regulation of gene expression'. Lower expressed genes (Suppl. Fig. 2D) had many fewer enriched terms. Unique enriched terms include responses to 'sulfur metabolism and starvation', 'glycosinolate metabolism'. Term shared with the kale vs TO1000 comparison included 'carbohydrate derivative metabolism'. These are again reflected in KEGG pathways where higher expressed genes were enriched for 'other glycan degradation' and 'starch and sucrose metabolism' KEGG pathways (Suppl. Table 15). Lower expressed genes were enriched for 'glucosinolate biosynthesis' and 'glutathione metabolism' KEGG pathways (Suppl. Table 16).

Shared genes- Fewer terms and pathways were significantly enriched amongst genes differentially expressed in both kale comparisons, in contrast to the individual comparisons. Higher expressed genes were enriched for 20 (12 BP) GO terms (Suppl. Table 11 and Suppl. Fig. 2G). These again include terms related to 'nitrile metabolism' and 'response to other organisms'. Fatty acid and lipid metabolism terms were also enriched and largely related to 'isoprenoid biosynthesis.' Additionally, terms related to' plastid organization', 'polysaccharide catabolism', and the' negative regulation of ion transport' were enriched. Only three GO terms were enriched in the lower expressed genes and these were all in the cellular component GO term class (Suppl.

Table 12). 'Other glycan degradation' was the only KEGG pathway enriched amongst higher expressed genes. Which was also enriched in both of the individual comparisons. No KEGG pathways were enriched in the shared lower expressed DEGs (Supp. Tables 19 and 20).

Candidate genes potentially involved in the domestication of kale-

A set of candidate genes were identified based on their expression pattern (Fig. 6) and functions related to the suite of kale domestication syndrome characteristics. We further identified transcription factors (TF) families and found a total of 141 TF in a total of 3010 differentially expressed genes (Fig. 3B). Some of the most abundant included bHLH, ERF and MYB TFs (Suppl. Table 21).

Apical dominance and plant architecture - Several genes known to be involved in shoot grow inhibition were differentially expressed in kale tissues and upregulated, these include: an auxin-resistance *AXR1* (Bo8g102450, Bo8g052710) (Stirnberg et al., 1999), *TCP18* (Teosinte Branched 1-Like 1) (Bo3g071570), two copies of *REVOLUTA* (Bo9g130730, Bo7g021280) (Poza-Carrión et al., 2007), and the Beta-carotene isomerase D27 chloroplastic (Protein *DWARF-27* homolog) (Bo9g035190) (Waters et al., 2012). The More Axillary Branches gene *MAX1* (Bo3g042090) (Bennett et al., 2006) was also differentially expressed and downregulated in kale (Table 1; Fig. 6A).

Leaf development - Differentially expressed genes in kale include two upregulated copies of *AXR1* (Bo8g052710; Bo8g102450). The downregulated LOB domain-containing protein *ASL6* (ASYMMETRIC LEAVES 2-like protein 6) (Bo7g041000) (Semiarti et al., 2001). These genes

could have a major role in the highly lobed margin kale phenotype (Fig. 6A). Genes differentially expressed in all three pair-wise comparisons that were potentially involved in the kale domestication syndrome (Fig. 3) include: *PID* (Bo4g182790) that regulates auxin flower and organ formation (Bennett et al., 1995), and *HAT5* (Bo5g155120) which is involved in leaf formation and has been characterized as involved in leaf margin development (Aoyama et al. 1995).

Identified transcription factors included: the AP2-like ethylene-responsive *ANT* (Bo3g153310). This gene is a positive regulator of cell proliferation, expansion and regulation of ectopic meristems (Shigyo et al., 2006), matching what was observed in the kale phenotypes (Fig. 2Q). The growth-regulating factor 7 (*GRF7*; Bo2g044580) involved in the regulation of cell expansion in leaf and cell proliferation and the Basic helix-loop-helix protein 64 *AtbHLH64* (Bo3g094060), an atypical *bHLH* transcription factor that acts as positive regulator of cell elongation were also identified here (Suppl. Table 21).

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

TRIPTYCHON *TRY* (Bo1g051040) involved in trichome branching, were also differentially expressed in kale, among others (Table1, Fig. 6A).

Flowering - Kale phenotypes display a significant delaying in flowering in comparison to other *Brassica oleracea* morphotypes. We hypothesized this maintenance in a vegetative stage, was an important character during domestication due mainly to the harvest of fresh leaves for consumption. Our mature kale phenotypes did not display flowering during the course of our experiments (24 days) (Fig. 1). A series of Agamous-like MADS-box proteins were exclusively expressed in kale (Fig. 3B) including two downregulated genes *AGL12* (Bo6g094120) and *AGL70* (Bo03856s010) and an upregulated *AGL27* (Bo3g100570*)*. MADS-box proteins such a *SOC1* (Bo4g024850) was downregulated, and two copies of FLOWERING LOCUS C (Bo9g173370, Bo3g024250) were upregulated (Table 1, Fig. 6A), both involved in the negative regulation of flowering, together with the MYB-related protein *R1 MYB1R1* (Bo6g119010) (Suppl. Table 21). Other genes differentially expressed in all three pair-wise comparisons that were potentially involved in the negative regulation of flowering (Fig. 3) included: *SCO1* (Bo4g024850, Bo3g038880), *SPL3* (Bo4g042390), *AP1* (Bo2g062650), *BLH3* (Bo6g119600), *AGL68/MAF5* (Bo3g100540) among others (Table 1, Fig. 6A)

Plant defense - Differentially expressed genes involved in plant defense exclusively expressed in curly kale leaves (Fig. 3b) include: Two copies of the upregulated Myb-related protein *MYB96* (Bo2g082270) involved in the activation of cuticular wax biosynthesis that could potentially deter herbivores. Five upregulated copies of a glucosinolate synthesis and regulation enzyme Myrosinase *TGG1* (Bo00934s010, Bo2g155820, Bo2g155840, Bo2g155870,

Bo8g039420), involved in the degradation of glucosinolates to produce toxic degradation products that can deter insect herbivores (Zheng et al. 2011). Last, two copies of the Indole Glucosinolate O-Methyltransferase, one was upregulated *IGMT1* (Bo2g092580), an a second one downregulated *IGMT1* (Bo2g092680) (Table 1, Fig. 6B), and *MYB29* (Bo9g175680) all involved in plant defense against fungus (Suppl. Table 21). Refer to Table 1 and Fig. 6B more other genes potentially involved in plant taste and defense.

Nutritional value of leaves- Kale is known to have a great nutritional value including nutrients like Vitamin C, calcium and iron (Edelman & Colt, 2016). Most notable, were three copies of the L-ascorbate oxidase (*AAO*) (Ascorbase) (EC 1.10.3.3)(Bo2g165650, Bo9g150050, Bo7g118970) involved in Vitamin-C synthesis, all three of which were downregulated in kale (Table 1; Fig. 6B) and the enzyme Ferritin-1 *LSC30* (Bo3g001360) proposed as a source of iron.

DISCUSSION

The vegetative domestication syndrome of kale is characterized by apical dominance, leaf morphology, maintenance of a vegetative stage for high leaf production, taste/defense, and nutritional value. Differentially Expressed Genes (DEGs) were identified in all pair-wise comparisons and were enriched amongst syntenic genes, while being depleted in non-syntenic genes (Fig. 5). This aligns with recent findings that genes derived from the ancient *Brassica* polyploidy are more likely involved in domestication in *B. rapa* (Qi et al., 2019). We focused our search on DEGs that were amongst comparisons with kale, reasoning that these were more likely to underly the unique characteristics of kale. We tested for enrichment in GO terms and KEGG pathways to gain a global view of the differences. Many more terms and pathways were

enriched amongst higher expressed genes, despite several thousand DEGs being identified in both groups. This may suggest more coordination in the regulation of kale higher-expressed genes or simply better annotation of genes in this list. Enriched terms and pathways in all categories were typically involved in metabolism or defense, while we found no clear enrichment of developmental GO terms. Changes in these functions would affect the taste, texture, and nutrition of leaves and thus likely have been affected by human selection. The absence of enrichment for developmentally-related GO terms could result from either insufficient annotation of developmental genes or due to more subtle expression differences that would require tissue-specific or single-cell approaches. Alternatively, the large developmental changes observed could be caused by differential expression of only a small number of key genes. To address this, we examined sets of genes, whose function based on homology and previous work, suggest an involvement in the developmental traits examined in this study. This analysis found many differentially expressed candidate genes shared amongst comparisons with kale, supporting their potential role in the kale domestication syndrome.

We identified potential candidate genes for the morphological and chemical differences among *Brassica* morphotypes. First, we found differentially expressed genes that set kale apart in terms of its morphology. Apical dominance has been proposed as a main character involved in crop domestication (Doebley et al., 1997). Suppression of axillary branches aiming to concentrate resources in the main stem is a salient phenotypic character in most *Brassica* crops. Apical dominance during domestication in kale can be explained by the indirect theory of apical dominance as auxin-induced stem growth inhibits bud outgrowth by diverting sugars away from

buds since growing stems are a strong sink for sugars (Kebrom, 2017). Here genes involved in apical dominance and branching suppression were differentially expressed (Table 1, Fig. 6).

Kale displays morphological diversity in leaves, including leaf shape, size and margins, trichome development, colors among others which is likely the result of genetic variation, selected by plant breeders. Environmental Scanning Electron Microscopy (ESEM) showed that many of the characteristic leaf traits of kale are established early in leaf development. We collected tissues during kale leaf initiation and primary morphogenesis (the upper portion of the stem including immature leaves, and apical and lateral meristems) to capture candidate genes related to the establishment of basic leaf structure such lamina initiation, specification of lamina domains, and the formation of marginal structures such lobes or serrations. Leaf development encompasses three continuous and overlapping phases. During leaf initiation, the leaf primordium emerges from the flanks of the stem apical meristem (SAM) at positions determined by specific phyllotactic patterns (Fig. 2 M-N). In the second phase, the leaf expands laterally and primary morphogenesis events occur from specific meristematic regions at the leaf margin (blastozones) (Fig. 2A-F). In the third phase of secondary morphogenesis, extensive cell expansion and histogenesis occurs (Fig. 2 G-L, Q, R).

Three types differentially expressed genes related to leaf morphogenesis phenotype in kale were found in our samples (Table 1; Fig. 3 and 6), genes: (1) involved in margin development, (2) involved in cell proliferation, expansion and the development of ectopic meristems and (3) involved in trichome development. The Lateral Organ Boundaries LOB domain-containing protein 4 (Asymmetric Leaves 2-like protein 6) *ASL6* (Bo7g041000) was

identified as a potential important gene responsible of the curly kale leaf phenotype (Semiarti et al., 2001) (Fig. 1, Fig. 2E, H, K). Loss-of-function mutations in the *AS2* homolog gene in *Arabidopsis* result in asymmetric leaf serration, with generation of leaflet-like structures from petioles and malformed entire veins and the ectopic expression of class 1 *KNOX* genes in the leaves (Matsumura et al., 2009). *ASL6* was differentially expressed and upregulated in kale. Asymmetric placement of auxin response at the distal leaf tip precedes visible asymmetric leaf growth. A second set of candidates involved two copies of the BEL1-like homeodomain protein 1 *BEL1/BHL1* (Bo3g029830, Bo4g142080) that were found to be over-expressed in kale. This gene has been involved in ovule development (Bellaoui et al., 2001), but since we did not collect any reproductive tissue in our sampling, we suggest this might have some other developmental role in kale.

In some species with simple leaves, *KNOXI* over-expression can lead to variable phenotypes, which include knot-like structures on the leaves, curled or lobed leaves and ectopic meristems on leaves (Tsiantis and Hay, 2003). Our differential expression analysis did not show *KNOXI* over-expression, however, the differential expression of *BELL* genes, another class of TALE HD proteins found in plants, was detected. Yeast two-hybrid studies have indicated that *KNOX* proteins interact with BEL1-like (*BELL*) proteins. The mRNA expression patterns of *KNOX* and *BELL* proteins overlap in meristems, suggesting that they may potentially interact in vivo (Smith et al, 2000; Smith et al., 2002). Analogous *BLH–KNOX* interactions have also been reported in other plants, such as potato (*Solanum tuberosum*) and barley (*Hordeum vulgare*), indicating that these interactions are evolutionarily conserved and that the interaction is probably required for their biological functions (Muller et al., 2001 ; Chen et al., 2003). A matter of debate

is whether *KNOX* and *BLH* can exert some of their functions independently of each other or whether the formation of *KNOX/BLH* heterodimers is mandatory for TALEs to work. We did not find differential expression of *KNOX* genes in this analysis perhaps for several reasons: (1) *KNOX* do not participate in lobes formation in *Brassica oleracea*; (2) we did not target the correct point in development to capture *KNOX* expression with relation to leaf lobation. Perhaps these genes are expressed later during development in kale leaves, however we collected tissues from early leaflets where lobes are starting to develop; (3) a change in *BELL* expression is sufficient to affect *KNOX* and *BELL* interactions and alter leaf morphology without a corresponding change in *KNOX* expression, if they were functionally redundant then *KNOX* expression would suffice to hide any effects of *BELL* expression differences. In some species, such as legumes dissected leaves do not express *KNOX* genes; on the contrary *KNOX* expression has been observed in leaf primordia of species with un-lobed leaves, such as *Lepidium oleraceum* (Piazza et al., 2010). A third gene involved in cell proliferation, expansion and the development of ectopic meristems included the NGATHA *NGA1* (Bo3g039420) involved in cell proliferation (Lee et al., 2015), however it has been reported as downregulated for cell over proliferation and in our copies was upregulated (Table 1).

Epidermal cells are stimulated to differentiate into trichomes when a regulatory complex including transcription factors are triggered (Yang and Ye, 2013). ESEM results showed hairiness and individual trichomes are seen in kale leaf blades (Fig 2 N, Q). Differentially expressed genes involved in trichome formation included two *MYB* transcription factors (Higginson et al, 2003; Gilding and Marks 2010; Pattanaik et al., 2014), the enhancer of *GL3 EGL3* (Bo9g035460), the Homeobox-leucine zipper proteins *GL1* (Bo7g090950) and *GL2*

(Bo6g079990) (Morohasi et al., 2007), the transcription factor *TRY* (Bo1g051040) (Schnittger et al., 1999) and homeodomain GLABROUS 11, *HDG11* (Bo4g039530) (Khosla et a., 2014). These five genes either interact or act redundantly to promote trichome differentiation and were upregulated in kale (except HDG11) (Table 1; Fig. 6A). Trichome regulatory genes have been studied in *Brassica villosa*, a wild C genome relative *B. oleracea* that is densely covered by trichomes. Nayidu et al. (2014) found that *TRY* was upregulated in trichomes of *B. villosa* in contrast to *Arabidospis* and other *Brassica* species where this gene has been proposed as a negative regulator. Here we found *TRY* was upregulated in leaves and meristems of kale in comparison to *ELG3*, *GL1*, *GL2* (Table 1; Fig. 6A).

Another set of candidate genes for the kale domestication syndrome we looked for, are floral transition genes. Floral transition is a major development switch controlled by regulatory pathways, integrating endogenous and environmental cues (Koornneef et al., 2004). We suggest that during kale domestication, delaying of flowering was necessary to maintain leaf production as main consumable. Four main regulatory pathways directly or indirectly involved in flowering have been described in *Arabidopsis thaliana*: photoperiodic, autonomous, vernalization, and gibberellin (Okazaki et al., 2007). The central integrator of flowering signals is the Flowering Locus C (*FLC*) that encodes a transcription factor of the MADS box family. While *A. thaliana* contains only one *FLC* gene, five copies have been described for *B. oleracea* (Okazaki et al. 2007). We found two copies of this gene being differentially expressed in kale. One was upregulated (Bo9g173370) and together with Agamous-like MADS-box protein *AGL27/MAF1* (Bo3g100570) have been proposed as floral repressors (Ratcliffe et al., 2001). A second copy of

the *FLC* (Bo3g024250) gene was found to be shared by the three morphotypes kale, cabbage and TO1000 (Table 1, Fig. 6A).

Pathways to domestication in kale may involve selection of plants with the right balance of phytochemicals concentrations. The direction of selection solely towards a reduction in phytochemicals through time was not the case in kale since persistence of pungency and bitterness could have reduced competition from mammals as well as loss to pests, during cultivation and storage (Deham et al., 2020). In *Brassica* species, most glucosinolates are biosynthesized from methionine. The enzymes *CYP79B*1 (Bo7g118840) and *CYP83A1* (Bo4g191120) catalyze important reactions in the synthesis of glucosinolates precursors. *CYP79B1* is involved in the conversion of amino acids to aldoxime precursors for glucosinolate biosynthesis (Bak et al., 1998) and *CYP83A1* catalyzes the biosynthesis of indole glucosinolates indole-3-acetaldoxime into an indole-3-S-alkyl-thiohydroximate (Bak & Feyereisen, 2001). A series of copies of the Myrosinase enzyme *TGG1* were differentially expressed and upregulated in kale (Bo00934s010, Bo2g155820, Bo2g155840, Bo2g155870, Bo8g039420) (Zheng et al., 2011). This enzyme breaks down glucosinolates into toxins (isothiocyanates, thiocynates, nitriles, and epithionitriles) upon leaf tissue damage. Therefore, glucosinolates may act as a potent feeding deterrent for generalist insect species, as their toxicity causes developmental and fitness damage (Santolamazza-Carbone 2014, Yi et al., 2015). Gene ontology terms related to glucosinolates and response to herbivores were also found enriched in our analysis specifically in the comparison between kale vs. cabbage terms.

Several genes were identified related to nutritional content in kale, for example precursors and enzymes related to vitamin C, GDP-L-galactose phosphorylase 2 *VTC5* was upregulated in kale (Bo2g041230) (Duan et al., 2016) and several copies of L-ascorbate oxidase (*ASO*) (Ascorbase) (EC 1.10.3.3) *AAO* (Bo2g165650, Bo9g150050, Bo7g118970) (Fujiwara et al., 2016) were downregulated. Precursors of the vitamin B1 Phosphomethylpyrimidine synthase chloroplastic (EC 4.1.99.17) *THIC* (Bo4g061330) (Wachter et al., 2007). The RuBisCO large subunit-binding protein subunit alpha chloroplastic (60 kDa chaperonin subunit alpha (Bo04491s010) and large subunit-binding protein subunit beta chloroplastic (60 kDa chaperonin subunit beta) (Bo6g034560) (Edelman & Colt, 2016). Differentially expressed genes found in the three morphotypes included Ferredoxin-1 chloroplastic (AtFd1) *LSC30* (Bo8g059760) (Table 1, Fig. 6B).

CONCLUSIONS

This study provided leaf developmental analysis and RNAseq experiments for *Brassica oleracea* morphotypes kale, cabagge and TO1000, three economically important Cole crops with distinct plant morphologies and domestication syndromes. RNA-Seq experiments allow the discovery of novel candidate genes like those involved in domestication syndromes. We identified gene expression patterns in *Brassica oleracea* that are shared among two vegetative morphotypes cabbage and kale, and the “rapid-cycling” morphotype TO1000, and estimated the contribution of morphotype-specific gene expression patterns that set kale apart. During kale domestication farmers could have been selecting for different flowering times, taste and herbivore defense from plants in the wild that have different vernalization times, different taste,

resistance to herbivores or that flower at different times during development and all variations in forms could have arisen from that.

Acknowledgements

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

Author Contributions

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

Data availability

Raw reads and assemblies have been deposited in NCBI:

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

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

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

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

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

Literature cited

Allaby, R. G. (2014). "Domestication Syndrome in Plants." Encyclopedia of Global Archaeology: 2182-2184.

Allender, C. J., Allainguillaume, J., Lynn, J., and King, G. J. (2007). "Simple sequence repeats reveal uneven distribution of genetic diversity in chloroplast genomes of *Brassica oleracea* L. and (n = 9) wild relatives." Theoretical and Applied Genetics 114(4): 609-618.

Alexa, A. and J. Rahnenfuhrer (2019). “topGO: Enrichment Analysis for Gene Ontology.” R package. Available from: https://www.bioconductor.org/packages/release/bioc/html/topGO.html.

Andrews S. (2010). “FastQC a Quality Control Tool for High Throughput Sequence Data.” Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anthony, R. G., James, P. E., and Jordan, B. R. (1995). "The cDNA sequence of a cauliflower apetala-1/squamosa homolog." Plant Physiology 108(1): 441-442.

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

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

Bak, S., Nielsen, H. L., and Halkier, B. A. (1998). The presence of *CYP79* homologues in glucosinolate-producing plants shows evolutionary conservation of the enzymes in the conversion of amino acid to aldoxime in the biosynthesis of cyanogenic glucosides and glucosinolates. Plant Molecular Biology, 38(5), 725-734.

Bak, S., and Feyereisen, R. (2001). The involvement of two P450 enzymes, *CYP83B1* and *CYP83A1*, in auxin homeostasis and glucosinolate biosynthesis. Plant Physiology, 127(1), 108-118.

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

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

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

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

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

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

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

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

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

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

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

Cole, M., Nolte, C., AND Werr, W. (2006). Nuclear import of the transcription factor SHOOT MERISTEMLESS depends on heterodimerization with *BLH* proteins expressed in discrete sub-domains of the shoot apical meristem of *Arabidopsis thaliana*. Nucleic Acids Research, 34(4), 1281-1292.

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

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

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

Doebley, J. (1992). “Molecular systematics and crop evolution.” Molecular systematics of plants, Springer: 202-222.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kim, Y. B., Li, X., Kim, S. J., Kim, H. H., Lee, J., Kim, H., and Park, S. U. (2013). MYB transcription factors regulate glucosinolate biosynthesis in different organs of Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Molecules, 18(7), 8682-8695.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Nasim, Z., Fahim, M., and Ahn, J. H. (2017). Possible role of MADS AFFECTING FLOWERING 3 and B-BOX DOMAIN PROTEIN 19 in flowering time regulation of

Arabidopsis mutants with defects in nonsense-mediated mRNA decay. Frontiers in Plant Ccience, 8, 191.

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

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

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

Pastore, J. J., Limpuangthip, A., Yamaguchi, N., Wu, M. F., Sang, Y., Han, S. K., Malaspina, N., et al. (2011). LATE MERISTEM IDENTITY2 acts together with LEAFY to activate APETALA1. Development, 138(15), 3189-3198.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Velasco ,P., Francisco, M., Moreno, D. A., Ferreres F., García-Viguera, C., Cartea, M.E. (2011) Phytochemical fingerprinting of vegetable *Brassica oleracea* and *Brassica napus* by simultaneous identification of glucosinolates and phenolics. Phytochemical Analysis Mar-Apr;22(2):144-52.

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

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

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

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

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

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

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

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

911

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
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 formation			
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
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	Pelaz et al., 2001

		meristem identity	
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
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
Bo4g024850	*SOC1*	Negative regulation of flower development	Nasim et al., 2017
Bo6g094120	*AGL12/XAL1*	Controls flowering	Tapia-Lopez et al., 2020
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
Bo3g038880	*SOC1*	Negative regulation of flower development	Nasim et al., 2017
Bo8g095790	*MYB17*	Regulation of flower development, targets LEAFY	Pastore et al., 2011
Defense and taste			
Genes differentially expressed only in Kale			
Bo7g118840	*CYP79B1*	Catalyzes the conversion of amino acids to aldoxime	Bak et al., 1998

		precursors for glucosinolate biosynthesis	
Bo4g191120	*CYP83A1*	Catalyzes the biosynthesis of indole glucosinolates indole-3-acetaldoxime into an indole-3-S-alkyl-thiohydroximate	Bak & Feyereisen, 2001
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
Differentially expressed genes shared by all morphotypes			
Bo2g161180	*MYB34*	Indole glucosinolate biosinthesis	Kim et al., 2013
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	Wachter et al., 2007

		B1	
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

Table 1. Candidate genes that are differentially expressed in curly kale (red winter kale) and shared among the three morphotypes. Genes involved in plant development, flowering, plant defence responses and taste, nutrition.

Figures

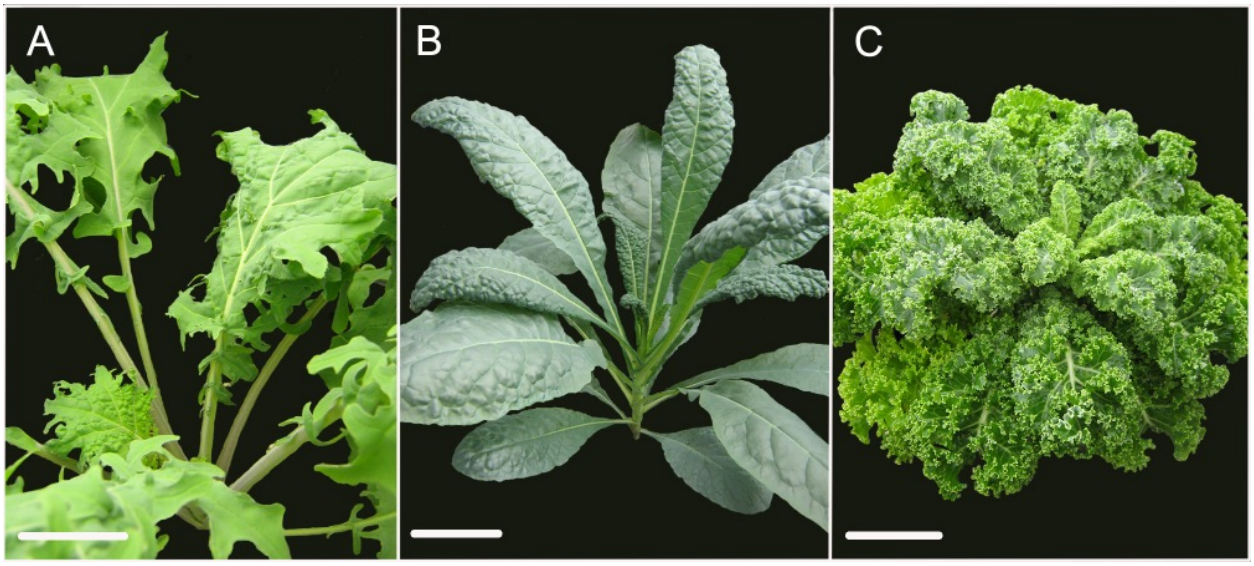

Figure 1. Comparison of mature leaf morphologies within kale varieties. (A) red winter kale (curly kale) with lobed leaf margins, (B) Lacinato Nero Toscana kale (non-curly kale) with crenulate leaf margins, (C) parsley kale with crisped leaf margins. Scale bar = 10 cm

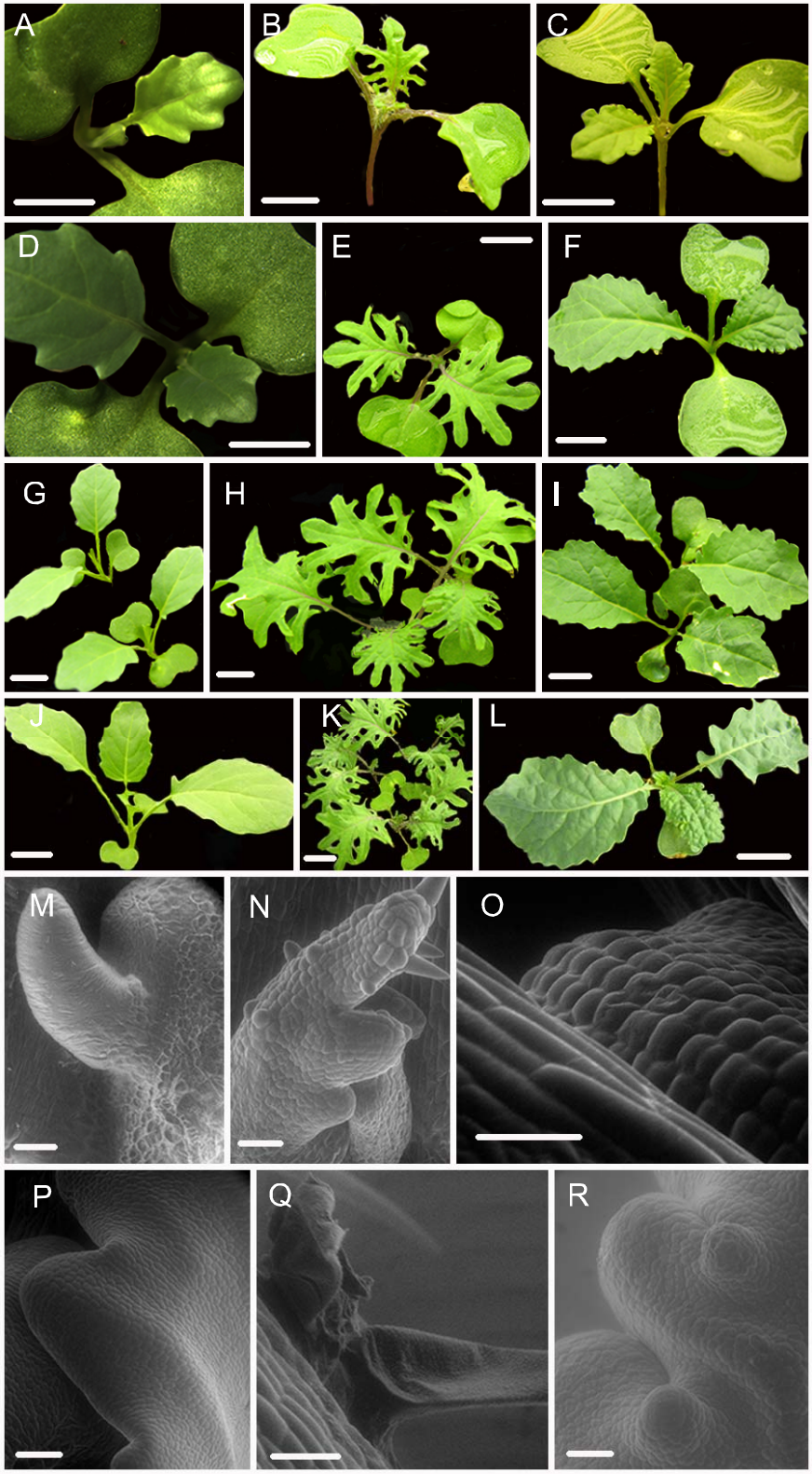
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R

Figure 2. Leaf shape development in different Kale varieties. A-L: progressive development of the different Kale varieties and the rapid cycling TO1000 control (stereoscope and regular pics), scale bar = 1 cm. M-R: Scanning Electron Microscopy (SEM) of apical meristem, primordial leaves and details of lobes in different kale varieties and TO1000 control, scale bar = 5 microns. A, D, G, J: leaf development series for TO1000. B, E, H, K: leaf development series for curly kale (red winter kale). C, F, I, L: leaf development series for non-curly kale. M, P: SEM of TO1000 leaf primordia and margin formation detail. N: SEM of curly kale leaf primordial. O: SEM of non-curly kale (Lacinato Nero Toscana kale) leaf primordia. Q: SEM of ectopic growing of leaf blade in mature leaves subtended by a trichome. R: Non-curly kale (Lacinato Nero Toscana kale) margin formation detail.

A. **Differentially Expressed Genes**

Comparison	Higher Expressed DEGs	Lower Expressed DEGs	Total DEGs
Kale vs T01000	3817	3457	7274
Kale vs Cabbage	3600	2722	6322
Cabbage vs TO1000	2775	3541	6316

B.

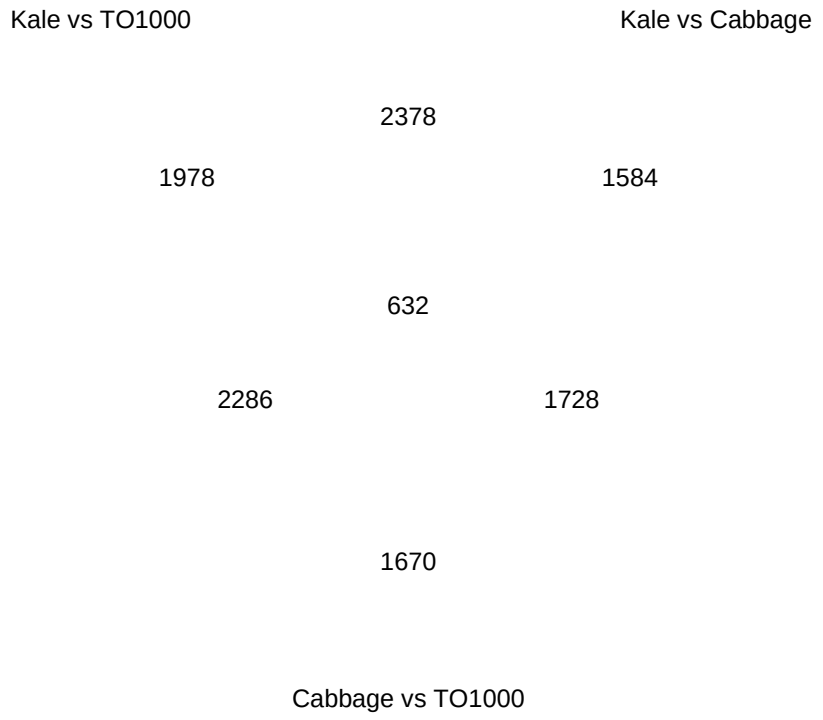

Figure 3. Number of genes differentially expressed in this study after comparing three different *Brassica oleracea* morphotypes (cabbage, kale and TO1000). A. Number of genes differentially expressed. B. Venn Diagram showing shared genes after doing all possible comparisons between three different *B. oleracea* morphotypes

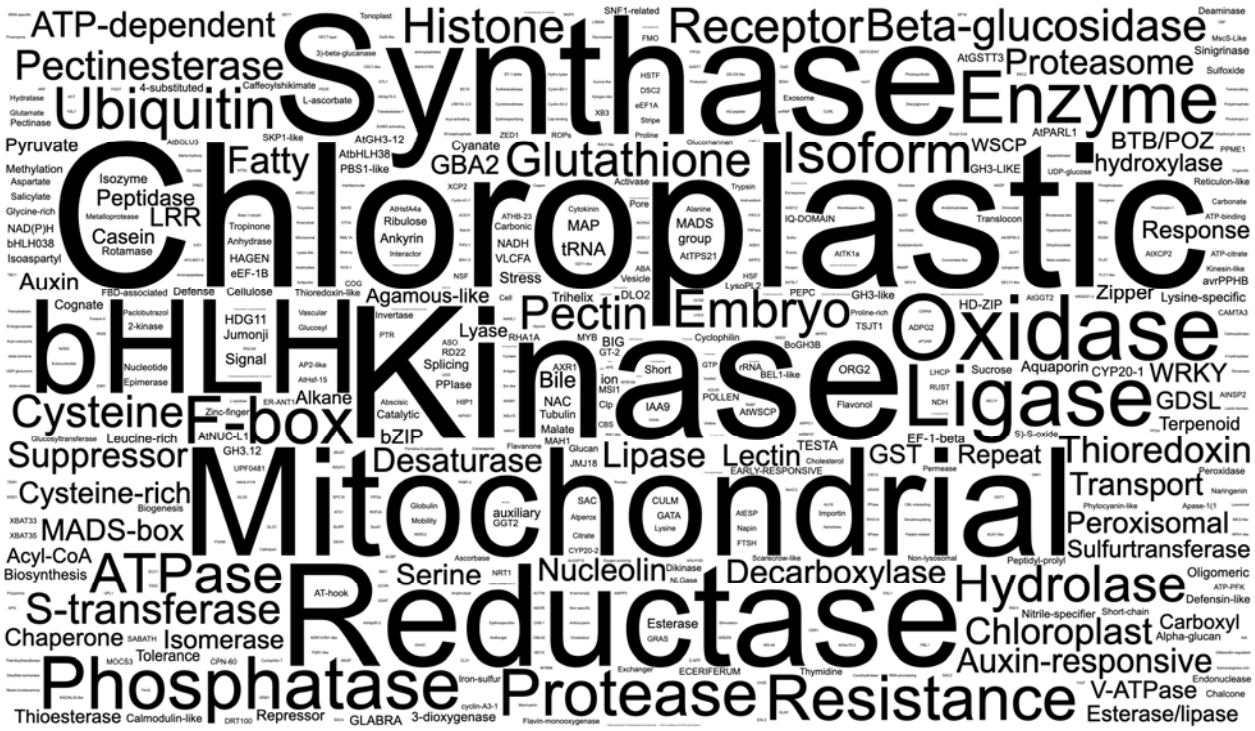

Figure 4. Word cloud for differentially expressed genes unique to kale among leaf and meristematic tissues of *Brassica oleracea* morphotypes, emphasizing the predominant role of chloroplast and mitochondrial genes, Kinases, Reductases and Synthases.

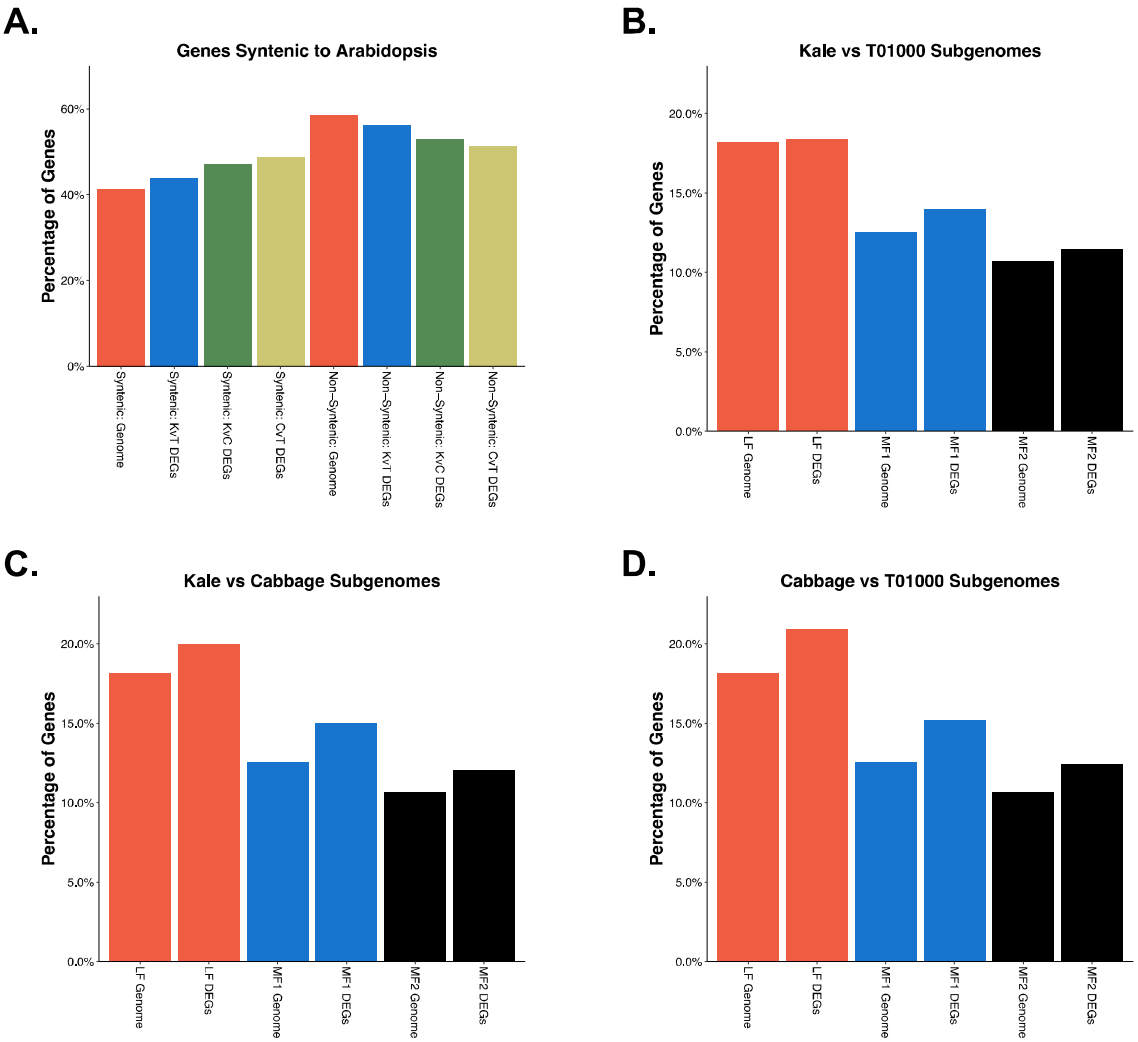

Figure 5. Syntenic genes from the three subgenomes are enriched in Differentially Expressed Genes (DEGs).

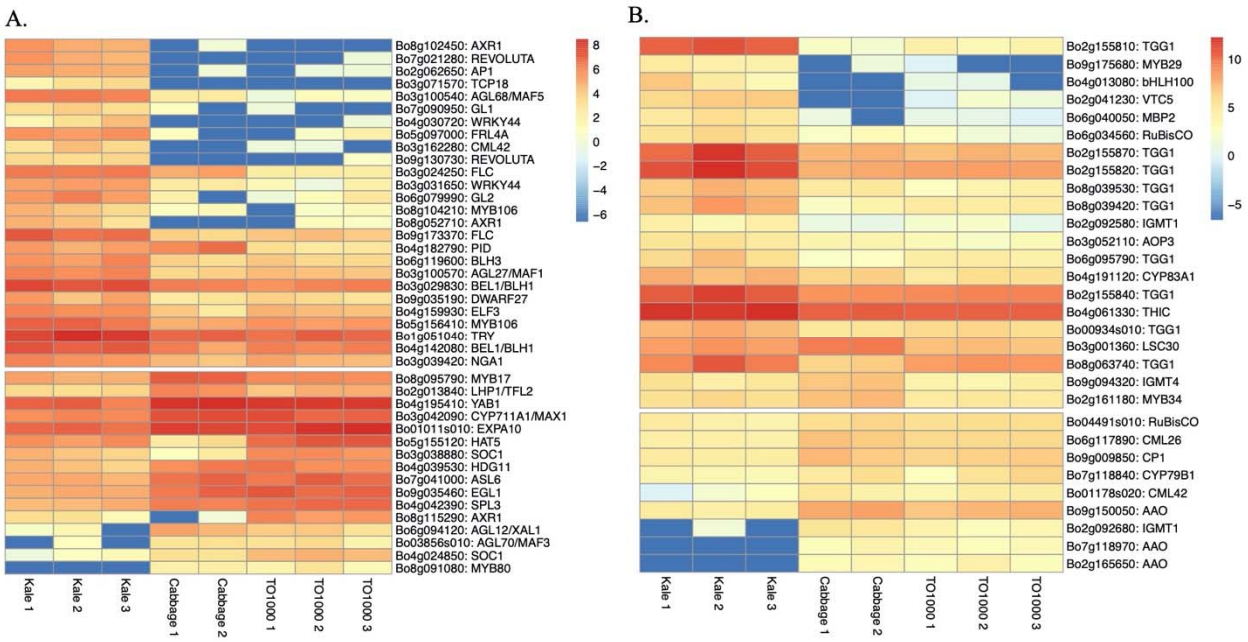

Figure 6. Heatmap of differentially expressed genes for curly kale (red winter kale). Up and down-regulated genes involved in A. plant and leaf development, and flowering B. plant defence responses and taste, nutrition.