



**HAL**  
open science

# Identification and functional prediction of anthocyanin biosynthesis regulatory long non-coding RNAs (lncRNAs) in carrot

Constanza Chialva, Thomas Blein, Martin Crespi, Diego Lijavetzky

## ► To cite this version:

Constanza Chialva, Thomas Blein, Martin Crespi, Diego Lijavetzky. Identification and functional prediction of anthocyanin biosynthesis regulatory long non-coding RNAs (lncRNAs) in carrot. 2020. hal-03016329

**HAL Id: hal-03016329**

**<https://hal.science/hal-03016329v1>**

Preprint submitted on 20 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Identification and functional prediction of anthocyanin biosynthesis**  
2 **regulatory long non-coding RNAs (lncRNAs) in carrot**

3

4

5 **Constanza Chialva<sup>1#</sup>, Thomas Blein<sup>2#</sup>, Martín Crespi<sup>2\*</sup> & Diego Lijavetzky<sup>1\*</sup>**

6

7 <sup>1</sup>Instituto de Biología Agrícola de Mendoza (IBAM), UNCuyo, CONICET, Facultad de  
8 Ciencias Agrarias, Almirante Brown 500, M5528AHB, Chacras de Coria, Mendoza,  
9 Argentina. <sup>2</sup>Institute of Plant Sciences Paris-Saclay (IPS2), CNRS, INRA, University  
10 Paris-Saclay and University of Paris Batiment 630, Gif Sur Yvette, France.

11

12 <sup>#</sup>Constanza Chialva and Thomas Blein contributed equally to this work.

13 <sup>\*</sup>Correspondence and requests for materials should be addressed to D.L. (email:  
14 [dlijavetzky@conicet.gov.ar](mailto:dlijavetzky@conicet.gov.ar)) or M.C. (email: [martin.crespi@ips2.universite-paris-  
15 saclay.fr](mailto:martin.crespi@ips2.universite-paris-<br/>15 saclay.fr))

16

17 **ABSTRACT**

18

19 Carrot (*Daucus carota* L.) is one of the most cultivated vegetable in the world and of  
20 great importance in the human diet. Its storage organs can accumulate large quantities of  
21 anthocyanins, metabolites that confer the purple pigmentation to carrot tissues and  
22 whose biosynthesis is well characterized. Long non-coding RNAs (lncRNAs) play  
23 critical roles in regulating gene expression of various biological processes in plants. In  
24 this study, we used a high throughput stranded RNA-seq to identify and analyze the  
25 expression profiles of lncRNAs in phloem and xylem root samples using two genotypes  
26 with a strong difference in anthocyanin production. We identified 639 differentially  
27 expressed lncRNAs between genotypes, and certain were specifically associated with a  
28 particular tissue. We then established regulatory correlations between lncRNAs and  
29 anthocyanin biosynthesis genes in order to identify a molecular framework for the  
30 differential expression of the pathway between genotypes. A specific natural antisense  
31 transcript (NAT) linked to the *DcMYB7* key anthocyanin biosynthetic transcription  
32 factor suggested how the regulation of this pathway may have evolved between  
33 genotypes.

34

35

36 **INTRODUCTION**

37

38 Anthocyanins are flavonoids, a class of phenolic compounds synthesized via the  
39 phenylpropanoid pathway, a late branch of the shikimic acid pathway<sup>1</sup>. They are  
40 secondary metabolites that confer purple, red, and blue pigmentation to several organs  
41 and tissues of many plant species<sup>2</sup>. These water-soluble pigments serve in various roles

42 in the plant, including attracting pollinators to flowers and seed dispersers to fruits,  
43 protection against UV radiation, amelioration of different abiotic and biotic stresses,  
44 such as drought, wounding, cold temperatures, and pathogen attacks<sup>3,4</sup>, as well as  
45 participation in physiological processes such as leaf senescence<sup>5,6</sup>. As dietary  
46 components, anthocyanins possess various health-promoting effects, mainly due to their  
47 antioxidant and anti-inflammatory properties, including protection against cancer,  
48 strokes and other chronic human disorders<sup>7</sup>.

49 Carrot (*Daucus carota* subsp. *carota* L.;  $2n = 2x = 18$ ) is a globally important root crop  
50 with yellow and purple as the first documented colors for domesticated carrot in Central  
51 Asia approximately 1,100 years ago<sup>8</sup>. Orange carrots were not reliably reported until the  
52 sixteenth century in Europe<sup>9,10</sup>, where its popularity was fortuitous for modern  
53 consumers because the orange pigmentation results from high quantities of  $\alpha$ - and  $\beta$ -  
54 carotene, making carrots the richest source of provitamin A in the US diet<sup>11</sup>.  
55 Additionally, with its great nutrition and economic value, carrot has been well known as  
56 a nice model plant for genetic and molecular studies<sup>11</sup>. Carrot is one of the crops that  
57 can accumulate large quantities of anthocyanins in its storage roots (up to 17–18  
58 mg/100 g fresh weight)<sup>12</sup>. Purple carrots accumulate almost exclusively derivatives of  
59 cyanidin glycosides with five cyanidin pigments reported in most studies<sup>13,14</sup>. The root  
60 content of these five anthocyanin pigments vary across carrot genetic backgrounds<sup>12,15</sup>.  
61 In addition, anthocyanin pigmentation also varies between root tissues, ranging from  
62 fully pigmented roots (i.e., purple color in the root phloem and xylem) to pigmentation  
63 only in the outer-most layer of the phloem<sup>16,17</sup>.

64 Regardless of the plant species, at least two classes of genes are involved in anthocyanin  
65 biosynthesis: structural genes encoding the enzymes that directly catalyze the  
66 production of anthocyanins, and regulatory genes that control the transcription of

67 structural genes<sup>18,19</sup>. In most cases, the anthocyanin biosynthetic structural genes are  
68 regulated by transcription factors (TFs) belonging to the R2R3-MYB, basic helix-loop-  
69 helix (bHLH) and WD-repeat protein families, in the form of the ‘MBW’ complex<sup>19,20</sup>.  
70 Recent reports pointed out that gene regulation by TFs may play a key role controlling  
71 anthocyanin pigmentation in purple carrots<sup>17,21,22</sup>. Moreover, the broad variation  
72 observed among purple carrot root genotypes, regarding both anthocyanin concentration  
73 and pigment distribution in the phloem and xylem tissues, suggests independent genetic  
74 regulation in these two root tissues<sup>23</sup>. In this sense, Xu et al.<sup>16</sup> found that the expression  
75 pattern of a R2R3-MYB TF, *DcMYB6*, is correlated with anthocyanins production in  
76 carrot roots and that the overexpression of this gene in *Arabidopsis thaliana* enhanced  
77 anthocyanins accumulation in vegetative and reproductive tissues in this heterologous  
78 system. Similarly, Kodama et al.<sup>24</sup> found that a total of 10 *MYB*, *bHLH* and *WD40* genes  
79 were consistently up- or downregulated in a purple color-specific manner, including  
80 *DcMYB6*. Iorizzo et al.<sup>25</sup> identified a cluster of MYB TFs, with *DcMYB7* as a candidate  
81 gene for root and petiole pigmentation, and *DcMYB11* as a candidate gene for petiole  
82 pigmentation. Bannoud et al.<sup>23</sup> showed that *DcMYB7* and *DcMYB6* participate in the  
83 regulation of phloem pigmentation in purple-rooted samples. Finally, Xu et al.<sup>26</sup>, by  
84 means of lost- and gain-of-function mutation experiments, demonstrated that *DcMYB7*  
85 is the main determinant that controls purple pigmentation in carrot roots.

86 Non-coding RNAs with a length higher than 200 nucleotides are defined as long  
87 noncoding RNAs (lncRNAs). They were originally considered to be transcriptional  
88 byproducts, or transcriptional ‘noise’, and were often dismissed in transcriptome  
89 analyses due to their low expression and low sequence conservation compared with  
90 protein-coding mRNAs. However, specific lncRNAs were shown to be involved in  
91 chromatin modification, epigenetic regulation, genomic imprinting, transcriptional

92 control as well as pre- and post-translational mRNA processing in diverse biological  
93 processes in plants<sup>27–30</sup>. Certain lncRNAs can be precursors of small interfering RNA  
94 (siRNA) or microRNA (miRNAs), triggering the repression of protein-coding genes at  
95 the transcription level (transcriptional gene silencing or TGS) or at post-transcriptional  
96 level (PTGS)<sup>27,31</sup>. Additionally, other lncRNAs can act as endogenous target mimics of  
97 miRNAs, to fine-tune the miRNA-dependent regulation of target genes<sup>32,33</sup>. It has been  
98 suggested that lncRNAs can regulate gene expression in both the *cis*- and *trans*-acting  
99 mode<sup>34</sup>. The *cis*-acting lncRNAs can be classified by their relative position to annotated  
100 genes<sup>27,35,36</sup> and notably include long noncoding natural antisense (lncNATs)  
101 transcribed in opposite strand of a coding gene, overlapping with at least one of its  
102 exons<sup>37,38</sup>. Other so-called intronic lncRNAs are transcribed within introns of a protein-  
103 coding gene<sup>39</sup> whereas long intergenic ncRNAs (lincRNAs) are transcripts located  
104 farther than 1 kb from protein-coding genes<sup>27,35,36</sup>. Among these *cis*-lncRNAs, NATs  
105 are of special interest as they have been shown to provide a mechanism for locally  
106 regulating the transcription or translation of the target gene on the other strand,  
107 providing novel mechanisms involved in the regulation of key biological processes<sup>40</sup>  
108 and environmentally dependent gene expression<sup>37,38</sup>.

109 As mentioned above, several differential expression analyses have been performed  
110 between purple and non-purple carrot roots allowing the identification of the main  
111 structural genes and TFs involved in anthocyanin biosynthesis in whole roots and/or  
112 phloem tissues<sup>16,21,23–26</sup>. However, the identification and functional prediction of  
113 lncRNA in carrot or putatively involved in carrot anthocyanin biosynthesis regulation  
114 has not yet been reported. In the present study, we combined a high throughput stranded  
115 RNA-Seq based approach with a dedicated bioinformatic pipeline, to annotate lncRNAs  
116 and analyze the expression profiles of lncNATs putatively associated to the carrot root

117 anthocyanin biosynthesis regulation. In addition, we individually analyzed the gene  
118 expression patterns in phloem and xylem root of purple and orange *D. carota*  
119 genotypes. Our findings point for a role of antisense transcription in the anthocyanin  
120 biosynthesis regulation in the carrot root at a tissue-specific level.

121

122

## 123 **RESULTS**

124

### 125 **RNA-seq data mining, identification and annotation of anthocyanin-related** 126 **lncRNAs**

127 In order to thoroughly identify and annotate lncRNAs related to anthocyanin  
128 biosynthesis regulation in carrot roots, we performed a whole transcriptome RNA-seq  
129 analysis of the carrot genotypes ‘Nightbird’ (purple phloem and xylem) and ‘Musica’  
130 (orange phloem and xylem) that had been studied in a tissue specific manner  
131 (Supplementary Figure S1). We generated an average of 51.4 million of reads per  
132 sample from the 12 carrot root samples (i.e., two phenotypes x two tissues x three  
133 biological replicates), ranging from 43.5 million to 60.3 million. The average GC  
134 content (%) was 44.8% and the average ratio of bases that have phred<sup>41</sup> quality score of  
135 over 30 (Q30) was 94.1%. The average mapping rate to the carrot genome was 90.9%  
136 (Supplementary Table S1). We identified and annotated 8484 new transcripts, including  
137 2095 new protein-coding transcripts and 6373 noncoding transcripts (1521 lncNATs,  
138 4852 lincRNAs and 16 structural transcripts) (Supplementary Table S2 and  
139 Supplementary File S1). Those were added to the 34263 known carrot transcripts<sup>42</sup> to  
140 conform the final set of 42747 transcripts used for this work. The set contains 34204  
141 coding transcripts and 7288 noncoding transcripts (1521 lncNATs, 5767 lincRNAs) and

142 1255 structural transcripts (Fig 1A and Supplementary Table S3). Most noncoding  
143 transcripts presented less than 1000 bp long, being 400-800 bp the most frequent length  
144 class. Coding transcripts between 500-1000 bp long were the most frequent, while most  
145 structural transcripts presented less than 200 bp (Fig 1B). Noncoding transcripts  
146 predominantly presented one exon and unexpectedly<sup>43</sup>, only one exon was also the most  
147 frequent class for coding transcripts (Fig 1C). Finally, we found no particular bias for  
148 the distribution of the noncoding transcripts along the nine carrot chromosomes (Fig  
149 1D). As expected, the average expression level of the ncRNAs, measured in Transcripts  
150 Per Kilobase Million (TPMs), was lower than that of the mRNAs (Fig. 1E).

151

152 **Variation in coding and noncoding expression was mainly explained by the**  
153 **anthocyanin-pigmentation phenotype difference between orange and purple**  
154 **carrots**

155 We sampled orange and purple carrot genotypes in two different tissues, phloem and  
156 xylem (Supplementary Figure S1). Considering the global gene variation of the 12  
157 evaluated libraries (i.e., three for each phenotype/tissue combination), the color  
158 phenotype was clearly the main source of variation (PC1, 49 %), while the tissue  
159 specificity factor was also important (PC2, 18%), but significantly lower than the root  
160 color phenotype effect (Fig. 2A).

161 We then assessed the variation in mRNA and ncRNA gene expression between purple  
162 and orange carrot roots in our RNA-seq analysis. A total of 3567 genes were  
163 differentially expressed (DEG) between purple and orange carrots (Bonferroni's  
164 adjusted  $p$ -value < 0.01), divided in 2928 mRNA and 639 lncRNAs (Fig. 2B) and  
165 representing 10% and 15% of the mRNA and lncRNA expressed genes, respectively.  
166 Within the 3567 DEGs, we found 1664 downregulated and 1907 upregulated



167 transcripts. In turn, the downregulated transcripts were distributed into 1343 coding and  
168 319 noncoding transcripts, while the upregulated were divided into 1585 and 320  
169 coding and noncoding transcripts, respectively (Fig. 2B). All information concerning  
170 the differentially expressed analysis and gene annotation is detailed in Supplementary  
171 Table S4.

172 As expected, we identify several differentially expressed genes (DEG) between the two  
173 genotypes known to be involved in carrot root anthocyanin biosynthesis<sup>21,23–26</sup>. Most of  
174 the known genes of the pathway and their main regulators were differentially expressed  
175 between the two genotypes (Supplementary Table S4). Several genes were induced in  
176 purple tissues and they mainly comprised genes representing: i) the early step in the  
177 flavonoid/anthocyanin pathway, like chalcone synthase (*DcCHS1*/DCAR\_030786);  
178 chalcone isomerase (*DcCH11*/DCAR\_027694) and (*DcCH1L*/DCAR\_019805);  
179 flavanone 3-hydroxylase (*DcF3H1*/DCAR\_009483), and flavonoid 3'-hydroxylase  
180 (*DcF3'H1*/DCAR\_014032); ii) cytochrome P450 (CYP450) proteins, putatively related  
181 to the flavonoid and isoflavonoid biosynthesis pathways<sup>23,44</sup>; iii) ATP-binding cassette  
182 (ABC) transporters, potentially related to anthocyanin transport<sup>45,46</sup>; and iv) genes from  
183 the late steps of the pathway, like dihydro-flavonol 4-reductase  
184 (*DcDFR1*/DCAR\_021485), leucoanthocyanidin dioxygenase (*DcLDOX1*/  
185 DCAR\_006772), and UDP-glycosyltransferase (*DcUFGT*/DCAR\_009823) and the  
186 recently described *DcUCGXT1*/DCAR\_021269 and *DcSAT1*/MSTRG.8365, which  
187 were confirmed to be responsible for anthocyanins glycosylation and acylation,  
188 respectively<sup>26,47</sup>. Finally, most significant regulatory genes of the pathway, belonging to  
189 the MYB, bHLH and WD40 TF gene families<sup>21,23–26</sup> were also differentially expressed  
190 between purple and orange genotypes (Supplementary Table S4). We further analyzed  
191 the tissue differential expression distribution of those 26 'MBW' TFs and found that

192 *DcMYB6* and *DcMYB7*, the two most studied TFs associated with anthocyanin  
193 biosynthesis regulation<sup>23–26</sup>, were differentially expressed between purple and orange  
194 carrots, both in phloem and xylem tissues (Supplementary Figure S2). Interestingly,  
195 three genes recently described to be regulated by *DcMYB7*<sup>26</sup> (i.e. *DcbHLH3*,  
196 *DcUCGXT1* and *DcSAT1*) also displayed no tissue specificity. *DcbHLH3* was described  
197 as a co-regulator in anthocyanin biosynthesis, while *DcUCGXT1* and *DcSAT1*  
198 participate in anthocyanin glycosylation and acylation, respectively<sup>26,47</sup>. Additionally,  
199 seven TFs showed xylem preferential expression-specificity, while only one was  
200 preferentially expressed specifically in phloem. Finally, differential expression of 11  
201 TFs was just detected when the 12 libraries were jointly analyzed, presumably by  
202 presenting significant but low expression differences (Supplementary Figure S2).

203

#### 204 **Putative regulation of anthocyanin-related genes by carrot antisense lncRNAs**

205 In order to investigate the putative involvement of carrot lncRNAs in the regulation of  
206 the anthocyanin biosynthesis in different carrot root tissues, we predicted the potential  
207 targets of lncRNAs in *cis*-regulatory relationship, particularly those classified as natural  
208 antisense transcripts (lncNATs). The selection of such lncRNAs was based in three  
209 assumptions: i) both, the lncRNA and the putative target were differentially expressed  
210 between purple and orange tissues (Supplementary Table S4); ii) the lncRNAs were  
211 antisense of the target genes; and iii) the Pearson and Spearman correlation coefficients  
212 between the expression levels of these genes were  $\geq 0.70$  or  $\leq -0.70$ , and  $p < 0.01$ .

213 According to these criteria, we found 19 differentially expressed lncNATs, since the  
214 lncRNAs were located in the antisense orientation (in the opposite strand) to a target  
215 mRNA, being most of them fully overlapping pairs (Supplementary Table S4 and S5).  
216 About 79% of those lncNATs were expressed in concordance with sense strand

217 transcript, while five out of the 19 presented discordant expression (i.e. when the  
218 lncNAT expression increase, the sense strand transcript was repressed) (Supplementary  
219 Table S4 and S5). Interestingly, we detected two lncNATs (MSTRG.27767/asDcMyb6  
220 and MSTRG.9120/asDcMyb7) in antisense relationship to *DcMYB6* and *DcMYB7*,  
221 respectively, with concordant expression correlation (Fig. 3). *DcMYB6* showed a log<sub>2</sub>  
222 fold-change of 7.6 with an adjusted *p*-value of  $4.5 \cdot 10^{-30}$ , while *DcMYB7* presented a  
223 log<sub>2</sub> fold-change of 11.7 with an adjusted *p*-value of  $3.8 \cdot 10^{-37}$ . Accordingly, the two  
224 detected antisense lncRNAs also presented significant differential expression, where  
225 *asDcMYB6* displayed a log<sub>2</sub> fold-change of 6.5 with an adjusted *p*-value of  $2.1 \cdot 10^{-13}$   
226 and *asDcMYB7* presented a log<sub>2</sub> fold-change of 6.1 with an adjusted *p*-value of  $1.3 \cdot 10^{-$   
227 <sup>04</sup> (Supplementary Table S4). Finally, the Pearson and Spearman correlation coefficients  
228 between the expression levels of each sense/antisense pair were  $\geq 0.79$  and *p*-value  $\leq$   
229 0.01 (Supplementary Table S5). On the other hand, as also detailed in Supplementary  
230 Table S4, two out of the four lncNATs showing discordant expression were found in the  
231 antisense relationship with disease resistance-like related genes (a predicted Catalase,  
232 and probable disease resistance protein At5g63020).

233

234

## 235 **DISCUSSION**

236

237 The presence of color in flowers, fruits and other organs and tissues, plays several  
238 biological functions mostly driven by the adaptive behavior of plants in response to the  
239 environment<sup>2,20,48,49</sup>. But in turn, plant organ pigmentation have served as natural  
240 genetic markers since the early works of Mendel<sup>50,51</sup>. Anthocyanins are flavonoid  
241 pigments that accumulate in plant cell vacuoles<sup>52</sup> and are mainly responsible for most

242 tissue and organ coloration<sup>19,20,48</sup>. Genetic analyses using model plant species like  
243 *Arabidopsis*, *petunia* and *maize* allowed the identification of most structural genes in the  
244 anthocyanin biosynthesis pathway as well as the main regulatory genes controlling  
245 pigment synthesis. In carrot, anthocyanin pigmentation is responsible for the purple  
246 phenotype<sup>9,53</sup>. Two main genes, *P<sub>1</sub>* and *P<sub>3</sub>*, have been identified in chromosome 3 and  
247 suggested to be responsible for the two independent mutations underlying the  
248 domestication of purple carrots<sup>17</sup>. Despite several carrot structural genes from the  
249 anthocyanin biosynthesis pathway have shown expression correlation with the purple  
250 phenotype<sup>21,22</sup>, none of them co-localize with *P<sub>1</sub>* and *P<sub>3</sub>*. Similar situation occurred in  
251 other plants like grapevine, where accumulation of anthocyanins correlated with the  
252 expression of several structural genes of the pathway but none of them co-localized with  
253 the ‘color locus’ in chromosome 2<sup>54,55</sup>. Finally, this discrepancy was solved by a study  
254 describing an insertion mutation in the promotor of a R2R3–MYB TF (i.e. *VviMybA1*)<sup>56</sup>  
255 explaining the lack of color of white grapevine cultivars. In the same direction, several  
256 recent works<sup>16,23–25,47</sup> focused on the role of carrot TFs putatively involved in the  
257 regulation of anthocyanin biosynthesis in purple genotypes, particularly those belonging  
258 to the ‘MBW’ complex (i.e., R2R3–MYB, basic helix-loop-helix -bHLH- and WD-  
259 repeat TFs). Two recent reports showed that three R2R3–MYB TFs are involved in the  
260 *P<sub>1</sub>* and *P<sub>3</sub>* loci: *DcMYB113* has been suggested to correspond to *P<sub>1</sub>*<sup>47</sup>, while *DcMYB6*  
261 and *DcMYB7* were proposed as the two main candidate TFs underlying the carrot root  
262 anthocyanin pigmentation in the *P<sub>3</sub>* locus<sup>25</sup>. However, knockdown and overexpression  
263 functional analyses demonstrated that *DcMYB7* (but not *DcMYB6*) is the *P<sub>3</sub>* gene  
264 controlling purple pigmentation in carrot roots<sup>26</sup>. Likewise described for the grapevine  
265 *VviMybA1* gene<sup>56</sup>, non-purple carrot genotypes seems to arise by an insertion mutation  
266 in the promoter region of *DcMYB7*<sup>26</sup>, yet the authors imply the existence of an

267 additional genetic factor suppressing the expression of *DcMYB7* in non-purple  
268 pigmented peridermal carrot root tissues.

269 In this work, we performed a thorough transcriptomic analysis by comparing two carrot  
270 hybrids with contrasted anthocyanin pigmentation phenotypes (i.e. purple vs. orange),  
271 both in phloem and xylem tissues. The study corroborates the involvement of the  
272 principal reported structural genes of the anthocyanin biosynthesis pathway<sup>21,22</sup>, but  
273 mostly, the key TF genes reported as the main regulators explain the carrot purple  
274 phenotype (i.e. *DcMYB6* and *DcMYB7*)<sup>16,25,26</sup>. Interestingly, the performed dissection  
275 between phloem and xylem purple samples, allowed us to show that there is no tissue-  
276 specific expression of such key genes, contrary to previously suggested for *DcMYB6*  
277 and *DcMYB7*<sup>16,23,25</sup>.

278 We showed here the first whole genome identification and annotation of lncRNAs in  
279 carrot by combining a high throughput stranded RNA-Seq based approach with a  
280 focused bioinformatic pipeline. Through this process, we identified 6373 novel  
281 lncRNAs, as compared to the 915 sequences annotated in the original carrot genome  
282 assembly<sup>42</sup>. Moreover, 10% of them (641 genes) can be defined as anthocyanin  
283 biosynthesis-related lncRNAs since we found them differentially expressed between  
284 purple and orange carrots. For the functional annotation of such lncRNAs, we focused  
285 on those showing an antisense relationship to differentially expressed protein coding  
286 genes, known (or putatively) involved in carrot anthocyanin biosynthesis and depicted  
287 in the precedent paragraph. Additionally, the selected lncRNAs had to present a  
288 statistically significant Pearson and Spearman correlation with their putative targets.  
289 This led us to identify 19 differentially expressed lncRNAs between purple and orange  
290 carrots. Interestingly, we found two of these lncRNAs (*asDcMYB6* and *asDcMYB7*)  
291 transcribed in opposite direction to *DcMYB6* and *DcMYB7*, respectively. Moreover,

292 *asDcMYB6* and *asDcMYB7* exhibited concordant expression patterns with their  
293 corresponding sense transcripts opening the possibility that non-coding RNA antisense  
294 transcription is a new player in the regulation of carrot anthocyanin biosynthesis,  
295 putatively affecting the expression of *DcMYB7* (and/or *DcMYB6*). This regulation  
296 maybe linked to the previously described genetic factors<sup>26</sup>.

297 Antisense transcripts, particularly lncNATs, present in many genomes of diverse  
298 kingdoms, showed either positively or negatively correlated expression with their  
299 corresponding sense transcripts. In many cases, lncNATs regulate the expression of  
300 their sense transcripts in a negative or positive way, by means of different  
301 transcriptional or post-transcriptional mechanisms. In particular cases, upregulation of  
302 sense gene expression may be explained by the participation of a lncNAT in the  
303 inhibition of other factors at translational level, such as efficient translation initiation or  
304 elongation<sup>57-59</sup>.

305 In plants, both repression and activation roles have been assigned to some lncNATs in  
306 response to environmental conditions. While *COOLAIR* and *COLD AIR* negatively  
307 regulates *FLC* in vernalization responses<sup>39,60</sup>, and *SVALK A* controls *CBF1* expression to  
308 consequently regulate freezing tolerance<sup>38</sup>, the expression of another member of the  
309 *FLC* family (*MAF4*) is activated by the lncNAT MAS to fine-tune flowering time<sup>37</sup>.

310 Anthocyanins are known to participate in abiotic stress responses and adaptation to  
311 environmental variations<sup>3,4,61</sup>, so the evolutionary role of the newly identified antisense  
312 transcripts *asDcMYB7* and *asDcMYB6* may be linked to the activation of anthocyanin  
313 biosynthesis through *DcMYB7* and *DcMYB6*. Hence, our work hints to new antisense  
314 regulations potentially involved in the variable expression of anthocyanin genes among  
315 carrot ecotypes.

316

317

## 318 **METHODS**

319

### 320 **Sample preparation and plant material**

321 Total RNA was obtained independently from three biological replicates of phloem and  
322 xylem root samples of two *Daucus carota* L genotypes: ‘Nightbird’, a purple root  
323 hybrid (purple phloem and xylem) and ‘Musica’, a non-anthocyanin pigmentated root  
324 hybrid. Plants were germinated from seeds and roots were collected after 12 weeks.  
325 Frozen samples were grinded using liquid nitrogen and RNA was extracted using TRI  
326 Reagent® (Sigma-Aldrich) and purified using SV Total RNA Isolation System  
327 (Promega). RNA samples were quantified, and purity measured using a  
328 spectrophotometer (AmpliQuant AQ-07). RNA integrity and potential genomic DNA  
329 contaminations were checked through agarose gel electrophoresis.

330

### 331 **Libraries construction and RNA sequencing**

332 Twelve samples (two genotypes x two tissues x three biological replicates) were sent to  
333 the Macrogen sequencing service (Seoul, Korea). Once in destiny they were checked for  
334 total RNA integrity using a Bioanalyzer RNA Nano 6000 chip. All the samples  
335 qualified to proceed with the library construction having an RNA Integrity Number  
336 (RIN)  $\geq 7$ . NGS transcriptomic libraries were constructed using a TruSeq Stranded  
337 mRNA LT Sample Prep Kit (Illumina). To verify the size of PCR enriched fragments,  
338 the template size distribution was checked on an Agilent Technologies 2100  
339 Bioanalyzer using a DNA 1000 chip. The sequencing of libraries was performed as  
340 paired-end 101 bp reads on an Illumina HiSeq 2500 platform. The quality of the raw  
341 reads in the FastQ files was checked through FastQC<sup>62</sup> and were then trimmed for

342 sequencing adaptor and low quality sequences using Trimmomatic<sup>63</sup> using  
343 ‘ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:21 TRAILING:21 MINLEN:30’  
344 as parameters. For removing reads corresponding to remaining ribosomal RNA,  
345 trimmed reads were mapped to the rRNA reference using SortMeRNA<sup>64</sup> using ‘--ref  
346 silva-bac-16s-id90.fasta --ref silva-bac-23s-id98.fasta --ref silva-euk-18s-id95.fasta --ref  
347 silva-euk-28s-id98.fasta --paired\_in --fastx --log -e 1e-07 -a 4 -v’ as parameters.

348

### 349 **New transcripts assembling and lncRNA identification**

350 Clean filtered reads were aligned on the *Daucus carota* genome<sup>42</sup> using the STAR  
351 aligner<sup>65</sup> using ‘--alignIntronMin 20 --alignIntronMax 20000 --outSAMtype BAM  
352 SortedByCoordinate --outReadsUnmapped Fastx’ as parameters. Subsequently, the  
353 aligned reads were assembled by means of StringTie<sup>66</sup> and new transcripts were  
354 extracted and annotated using the GffCompare<sup>67</sup> program (GffCompare classes ‘u’, ‘x’,  
355 to adjust). Only new transcripts whose length was greater than 200nt were kept. The  
356 classification of the newly predicted transcript was performed according the following  
357 steps: i) the transcripts were classified as coding if their predicted open reading frame  
358 (ORF) was greater than 120 aa or if they were predicted as coding by CPC2<sup>68</sup> calculator  
359 and classify as structural RNA in case of homology with structural RNA (tRNA, rRNA,  
360 snRNA or snoRNA) or as non-coding in case of homology with known structured non-  
361 coding RNA (miRNA precursors, lncRNA). The transcript without any homology in  
362 Rfam and classified as non-coding par CPC2 were classified as non-coding.

363

### 364 **Differential expression analysis**

365 We performed a strand-specific read counting of coding and non-coding gene using on  
366 the carrot official annotation and the newly predicted genes of this study for each of the



367 12 aligned BAM files by means of the featureCount<sup>70</sup> software included in the Rsubread  
368 package<sup>71</sup>. The resulted count data was used for differential expression analysis with  
369 DEseq2<sup>72</sup>. Differentially expressed genes were declared as having a Bonferroni's  
370 adjusted  $p$ -value  $< 0.01$ . Reads corresponding to the strand specific expression of  
371 mRNAs and their lncRNAs were visualized with the Integrative Genomics Viewer  
372 (IGV) software<sup>73</sup>. Additional Venn diagrams were performed with Venny v2.1<sup>74</sup>.

373

374

## 375 REFERENCES

376

- 377 1. Herrmann, K. M. & Weaver, L. M. The Shikimate Pathway. *Annu. Rev. Plant*  
378 *Physiol. Plant Mol. Biol.* **50**, 473–503 (1999).
- 379 2. Harborne, J. B. & Williams, C. A. Advances in flavonoid research since 1992.  
380 *Phytochemistry* **55**, 481–504 (2000).
- 381 3. Koes, R. E., Quattrocchio, F. & Mol, J. N. M. M. The flavonoid biosynthetic  
382 pathway in plants: Function and evolution. *BioEssays* **16**, 123–132 (1994).
- 383 4. Shirley, B. W. Flavonoid biosynthesis: 'new' functions for an 'old' pathway.  
384 *Trends Plant Sci.* **1**, 377–382 (1996).
- 385 5. Hatier, J.-H. B. & Gould, K. S. Anthocyanin Function in Vegetative Organs. in  
386 *Anthocyanins* 1–19 (Springer New York, 2008). doi:10.1007/978-0-387-77335-  
387 3\_1
- 388 6. Gould, K. S. & Lister, C. Flavonoid function in plants. *Flavonoids Chem.*  
389 *Biochem. Appl.* 397–441 (2006).
- 390 7. Lila, M. A. Anthocyanins and human health: An in vitro investigative approach.  
391 *Journal of Biomedicine and Biotechnology* **2004**, 306–313 (2004).

- 392 8. Iorizzo, M. *et al.* Genetic structure and domestication of carrot (*Daucus carota*  
393 subsp. *sativus*) (Apiaceae). *Am. J. Bot.* **100**, 930–938 (2013).
- 394 9. Simon, P. W. Domestication, Historical Development, and Modern Breeding of  
395 Carrot. in *Plant Breeding Reviews* 157–190 (John Wiley & Sons, Inc., 2010).  
396 doi:10.1002/9780470650172.ch5
- 397 10. Arscott, S. A. & Tanumihardjo, S. A. Carrots of Many Colors Provide Basic  
398 Nutrition and Bioavailable Phytochemicals Acting as a Functional Food. *Compr.*  
399 *Rev. Food Sci. Food Saf.* **9**, 223–239 (2010).
- 400 11. Simon, P. W., Pollak, L. M., Clevidence, B. A., Holden, J. M. & Haytowitz, D.  
401 B. Plant Breeding for Human Nutritional Quality. in *Plant Breeding Reviews* **31**,  
402 325–392 (John Wiley & Sons, Inc., 2009).
- 403 12. Montilla, E. C., Arzaba, M. R., Hillebrand, S. & Winterhalter, P. Anthocyanin  
404 composition of black carrot (*Daucus carota* ssp. *sativus* var. *atrorubens* Alef.)  
405 Cultivars *antonina*, *beta* sweet, *deep purple*, and *purple haze*. *J. Agric. Food*  
406 *Chem.* **59**, 3385–3390 (2011).
- 407 13. Kammerer, D., Carle, R. & Schieber, A. Quantification of anthocyanins in black  
408 carrot extracts (*Daucus carota* ssp. *sativus* var. *atrorubens* Alef.) and evaluation  
409 of their color properties. *Eur. Food Res. Technol.* **219**, 479–486 (2004).
- 410 14. Kammerer, D., Carle, R. & Schieber, A. Detection of peonidin and pelargonidin  
411 glycosides in black carrots (*Daucus carota* ssp. *sativus* var. *atrorubens* Alef.) by  
412 high-performance liquid chromatography/electrospray ionization mass  
413 spectrometry. *Rapid Commun. Mass Spectrom.* **17**, 2407–2412 (2003).
- 414 15. Mazza, G. & Miniati, E. (Enrico). *Anthocyanins in fruits, vegetables, and grains*.  
415 (CRC Press, 1993).
- 416 16. Xu, Z. S., Feng, K., Que, F., Wang, F. & Xiong, A. S. A MYB transcription

- 417 factor, DcMYB6, is involved in regulating anthocyanin biosynthesis in purple  
418 carrot taproots. *Sci. Rep.* **7**, 1–9 (2017).
- 419 17. Cavagnaro, P. F. *et al.* A gene-derived SNP-based high resolution linkage map of  
420 carrot including the location of QTL conditioning root and leaf anthocyanin  
421 pigmentation. *BMC Genomics* **15**, 1118 (2014).
- 422 18. Jaakola, L. *et al.* Expression of genes involved in anthocyanin biosynthesis in  
423 relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry  
424 fruit development. *Plant Physiol* **130**, 729–39 (2002).
- 425 19. Koes, R., Verweij, W. & Quattrocchio, F. Flavonoids: a colorful model for the  
426 regulation and evolution of biochemical pathways. *Trends Plant Sci* **10**, 236–242  
427 (2005).
- 428 20. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics,  
429 biochemistry, cell biology, and biotechnology. *Plant Physiol* **126**, 485–493  
430 (2001).
- 431 21. Yildiz, M. *et al.* Expression and mapping of anthocyanin biosynthesis genes in  
432 carrot. *Theor. Appl. Genet.* **126**, 1–14 (2013).
- 433 22. Xu, Z. S. *et al.* Transcript profiling of structural genes involved in cyanidin-based  
434 anthocyanin biosynthesis between purple and non-purple carrot (*Daucus carota*  
435 L.) cultivars reveals distinct patterns. *BMC Plant Biol.* **14**, (2014).
- 436 23. Bannoud, F. *et al.* Dissecting the genetic control of root and leaf tissue-specific  
437 anthocyanin pigmentation in carrot (*Daucus carota* L.). *Theor. Appl. Genet.* **132**,  
438 2485–2507 (2019).
- 439 24. Kodama, M. *et al.* Identification of transcription factor genes involved in  
440 anthocyanin biosynthesis in carrot (*Daucus carota* L.) using RNA-Seq. *BMC*  
441 *Genomics* **19**, 811 (2018).

- 442 25. Iorizzo, M. *et al.* A Cluster of MYB Transcription Factors Regulates  
443 Anthocyanin Biosynthesis in Carrot (*Daucus carota* L.) Root and Petiole. *Front.*  
444 *Plant Sci.* **9**, 1927 (2018).
- 445 26. Xu, Z.-S., Yang, Q.-Q., Feng, K. & Xiong, A.-S. Changing Carrot Color:  
446 Insertions in DcMYB7 Alter the Regulation of Anthocyanin Biosynthesis and  
447 Modification. *Plant Physiol.* pp.00523.2019 (2019). doi:10.1104/pp.19.00523
- 448 27. Ariel, F., Romero-Barrios, N., Jégu, T., Benhamed, M. & Crespi, M. Battles and  
449 hijacks: noncoding transcription in plants. *Trends Plant Sci.* **20**, 362–371 (2015).
- 450 28. Forestan, C. *et al.* Stress-induced and epigenetic-mediated maize transcriptome  
451 regulation study by means of transcriptome reannotation and differential  
452 expression analysis. *Sci. Rep.* **6**, 1–20 (2016).
- 453 29. Forestan, C. *et al.* Epigenetic signatures of stress adaptation and flowering  
454 regulation in response to extended drought and recovery in *Zea mays*. *Plant Cell*  
455 *Environ.* **43**, 55–75 (2020).
- 456 30. Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in  
457 epigenetic regulation. *Nature Structural and Molecular Biology* **20**, 300–307  
458 (2013).
- 459 31. Wu, H. J., Wang, Z. M., Wang, M. & Wang, X. J. Widespread long noncoding  
460 RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol.* **161**,  
461 1875–1884 (2013).
- 462 32. Zhang, G. *et al.* Transcriptomic and functional analyses unveil the role of long  
463 non-coding RNAs in anthocyanin biosynthesis during sea buckthorn fruit  
464 ripening. *DNA Res.* **25**, 465–476 (2018).
- 465 33. Yang, T. *et al.* Systematic identification of long noncoding <sc>RNA</sc> s  
466 expressed during light-induced anthocyanin accumulation in apple fruit. *Plant J.*

- 467           **100**, 572–590 (2019).
- 468   34.   Li, Z. & Rana, T. M. Molecular mechanisms of RNA-triggered gene silencing  
469           machineries. *Acc. Chem. Res.* **45**, 1122–1131 (2012).
- 470   35.   Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs.  
471           *Annu. Rev. Biochem.* **81**, 145–166 (2012).
- 472   36.   Bonasio, R. & Shiekhattar, R. Regulation of Transcription by Long Noncoding  
473           RNAs. *Annu. Rev. Genet.* **48**, 433–455 (2014).
- 474   37.   Zhao, X. *et al.* Global identification of Arabidopsis lncRNAs reveals the  
475           regulation of MAF4 by a natural antisense RNA. *Nat. Commun.* **9**, 1–12 (2018).
- 476   38.   Kindgren, P., Ard, R., Ivanov, M. & Marquardt, S. Transcriptional read-through  
477           of the long non-coding RNA SVALKKA governs plant cold acclimation. *Nat.*  
478           *Commun.* **9**, (2018).
- 479   39.   Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long  
480           intronic noncoding RNA. *Science (80-. )*. **331**, 76–79 (2011).
- 481   40.   Thieffry, A. *et al.* Characterization of Arabidopsis thaliana promoter  
482           Bidirectionality and Antisense RNAs by Depletion of Nuclear RNA Decay  
483           Pathways. *Plant Cell* tpc.00815.2019 (2020). doi:10.1105/tpc.19.00815
- 484   41.   Ewing, B., Hillier, L. D., Wendl, M. C. & Green, P. Base-calling of automated  
485           sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185  
486           (1998).
- 487   42.   Iorizzo, M. *et al.* A high-quality carrot genome assembly provides new insights  
488           into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **48**, 657–  
489           666 (2016).
- 490   43.   Ramírez-Sánchez, O., Pérez-Rodríguez, P., Delaye, L. & Tiessen, A. Plant  
491           Proteins Are Smaller Because They Are Encoded by Fewer Exons than Animal

- 492 Proteins. *Genomics, Proteomics Bioinforma.* **14**, 357–370 (2016).
- 493 44. Ayabe, S. I. & Akashi, T. Cytochrome P450s in flavonoid metabolism.  
494 *Phytochemistry Reviews* **5**, 271–282 (2006).
- 495 45. Kovinich, N. *et al.* Arabidopsis MATE45 antagonizes local abscisic acid  
496 signaling to mediate development and abiotic stress responses. *Plant Direct* **2**,  
497 e00087 (2018).
- 498 46. Francisco, R. M. *et al.* ABCC1, an ATP Binding Cassette Protein from Grape  
499 Berry, Transports Anthocyanidin 3-O-Glucosides. *Plant Cell Online* (2013).  
500 doi:10.1105/tpc.112.102152
- 501 47. Xu, Z., Yang, Q., Feng, K., Yu, X. & Xiong, A. DcMYB113, a root-specific  
502 R2R3-MYB, conditions anthocyanin biosynthesis and modification in carrot.  
503 *Plant Biotechnol. J.* pbi.13325 (2020). doi:10.1111/pbi.13325
- 504 48. Mol, J., Grotewold, E. & Koes, R. How genes paint flowers and seeds. *Trends*  
505 *Plant Sci* **3**, 212–217 (1998).
- 506 49. Winkel-Shirley, B. Biosynthesis of flavonoids and effects of stress. *Curr Opin*  
507 *Plant Biol* **5**, 218–223 (2002).
- 508 50. Henig, R. M. *The monk in the garden: the lost and found genius of Gregor*  
509 *Mendel, the father of genetics.* (Houghton Mifflin Harcourt, 2017).
- 510 51. Hartl, D. L. & Orel, V. What did Gregor Mendel think he discovered? *Genetics*  
511 **131**, 245 (1992).
- 512 52. Petrucci, E. *et al.* Plant Flavonoids—Biosynthesis, Transport and Involvement in  
513 Stress Responses. *Int. J. Mol. Sci.* **14**, 14950–14973 (2013).
- 514 53. Simon, P. W. Inheritance and expression of purple and yellow storage root color  
515 in carrot. *J. Hered.* **87**, 63–66 (1996).
- 516 54. Kobayashi, S., Ishimaru, M., Ding, C. K., Yakushiji, H. & Goto, N. Comparison

- 517 of UDP-glucose:flavonoid 3-O-glucosyltransferase (UFGT) gene sequences  
518 between white grapes (*Vitis vinifera*) and their sports with red skin. *Plant Sci*  
519 **160**, 543–550 (2001).
- 520 55. Kobayashi, S., Ishimaru, M., Hiraoka, K. & Honda, C. Myb-related genes of the  
521 Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta* **215**,  
522 924–933 (2002).
- 523 56. Kobayashi, S., Goto-Yamamoto, N. & Hirochika, H. Retrotransposon-induced  
524 mutations in grape skin color. *Science (80-. )*. **304**, 982 (2004).
- 525 57. Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription.  
526 *Nature Reviews Genetics* **14**, 880–893 (2013).
- 527 58. Wight, M. & Werner, A. The functions of natural antisense transcripts. *Essays*  
528 *Biochem.* **54**, 91–101 (2013).
- 529 59. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts.  
530 *Nature Reviews Molecular Cell Biology* **10**, 637–643 (2009).
- 531 60. Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long  
532 antisense transcripts of an Arabidopsis Polycomb target. *Nature* **462**, 799–802  
533 (2009).
- 534 61. Rienth, M. *et al.* Day and night heat stress trigger different transcriptomic  
535 responses in green and ripening grapevine (*vitis vinifera*) fruit. *BMC Plant Biol.*  
536 **14**, 108 (2014).
- 537 62. Andrews, S. FastQC: a quality control tool for high throughput sequence data.  
538 (2010). Available at:  
539 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 540 63. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for  
541 Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

- 542 64. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of  
543 ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217  
544 (2012).
- 545 65. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,  
546 15–21 (2013).
- 547 66. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome  
548 from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
- 549 67. GffCompare. A Program for processing GTF/GFF files. Available at:  
550 <https://ccb.jhu.edu/software/stringtie/gffcompare.shtml>.
- 551 68. Kang, Y.-J. *et al.* CPC2: a fast and accurate coding potential calculator based on  
552 sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
- 553 69. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology  
554 searches. *Bioinformatics* **29**, 2933–2935 (2013).
- 555 70. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose  
556 program for assigning sequence reads to genomic features. *Bioinformatics* **30**,  
557 923–930 (2014).
- 558 71. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster,  
559 cheaper and better for alignment and quantification of RNA sequencing reads.  
560 *Nucleic Acids Res.* **47**, e47–e47 (2019).
- 561 72. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and  
562 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- 563 73. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics  
564 Viewer (IGV): high-performance genomics data visualization and exploration.  
565 *Brief. Bioinform.* **14**, 178–192 (2012).
- 566 74. Oliveros, J. C. Venny. *Venny. An interactive tool for comparing lists with Venn's*



567 *diagrams.* (2015).

568

569

## 570 **ACKNOWLEDGEMENTS**

571 This work was supported by Agencia Nacional de Promoción Científica y Tecnológica

572 (ANPCyT): PICT2015-0822 & PICT-2016-3134; LabEx Sciences des Plantes de Saclay

573 (SPS). Travel support to Diego Lijavetzky was funded by Programa de Movilidad para

574 Docentes de la UNCuyo, INRA and Universite Paris Sud. We thank to Dr. Pablo

575 Cavagnaro for providing and growing the plant materials.

576

## 577 **AUTHOR CONTRIBUTIONS**

578 C.C. performed the samples collection, laboratory work for library preparation,

579 participated in the draft of the manuscript and figures preparation; T.B. designed the

580 bioinformatic pipeline related to raw data processing, new transcripts identification and

581 differential expression analysis. M.C provided the main insight on the design of the

582 transcriptomic experiment. D.L. wrote the final manuscript, designed and coordinated

583 the experiments. All authors carefully read and helped to improve the final content of

584 the manuscript.

585

586

## 587 **ADDITIONAL INFORMATION**

588 **Data access:** Sequence files generated during this study have been deposited into the

589 NCBI BioProject database accession PRJNA668894.

590 **Competing Interests:** The authors declare no competing interests.

591

## 592 **FIGURE LEGENDS**

593 **Figure 1.** Characteristics of carrot transcripts. (A) Distribution of coding, noncoding  
594 and structural sequences between the known and newly annotated transcripts. (B)  
595 Transcript length distributions for the total coding, noncoding and structural RNAs. (C)  
596 Number of exons per transcript for the total coding and noncoding RNAs. (D)  
597 Proportional distribution of the total coding, noncoding and structural RNAs along each  
598 chromosome. (E) Violin plot of the expression levels of carrot total coding and  
599 noncoding RNAs. The y-axis represents the average  $\log_2(\text{TPM})$  values. T-test  $p$ -value <  
600 0.01 is considered to be significantly different.

601

602 **Figure 2.** Expression of carrot coding and noncoding RNAs. (A) PCA analysis of the  
603 global gene expression of the 12 evaluated libraries (three replicates for each color-  
604 phenotype and tissue type combination). (B) Differentially expressed genes (up- and  
605 down-regulated) between purple and orange carrots (Bonferroni's adjusted  $p$ -value <  
606 0.01) distributed by coding and noncoding transcripts.

607

608 **Figure 3.** Strand specific expression of R2R3-MYB TFs and their lncNATs. Coverage  
609 data for the sense (green) and antisense (red) strands corresponding to  
610 *DcMYB7/asDcMYB7* (A) and *DcMYB6/as DcMYB6* (B), respectively. Tracks  
611 correspond to four carrot libraries: two phloem samples Purple\_F1 and Orange F1; and  
612 two xylem samples Purple\_X1 and Orange\_X1. Data Range of each track was set to  
613 allow an even visualization of the mRNA and lncRNA transcripts by enlarging the last  
614 ones (20x).

615

616 **Supplementary Figure S1.** Picture of the purple and orange hybrids used for the RNA-  
617 seq experiment and scheme of the performed dissection between xylem and phloem  
618 tissues.

619 **Supplementary Figure S2.** Tissue specific differential expression of the 26 ‘MBW’  
620 TFs identified in the experiment. a) Genes differentially expressed between purple and  
621 orange carrots both in xylem and phloem tissues; b) genes differentially expressed  
622 between purple and orange carrots just in xylem; c) gene differentially expressed  
623 between purple and orange carrots just in phloem; d) genes differentially expressed  
624 between purple and orange carrots detected after the join analysis of phloem and xylem  
625 samples.

626

627 **Supplementary Table S1.** Summary of NGS and quality control data regarding the 12  
628 sequenced libraries.

629 **Supplementary Table S2.** Genome annotation of the newly identified transcripts.

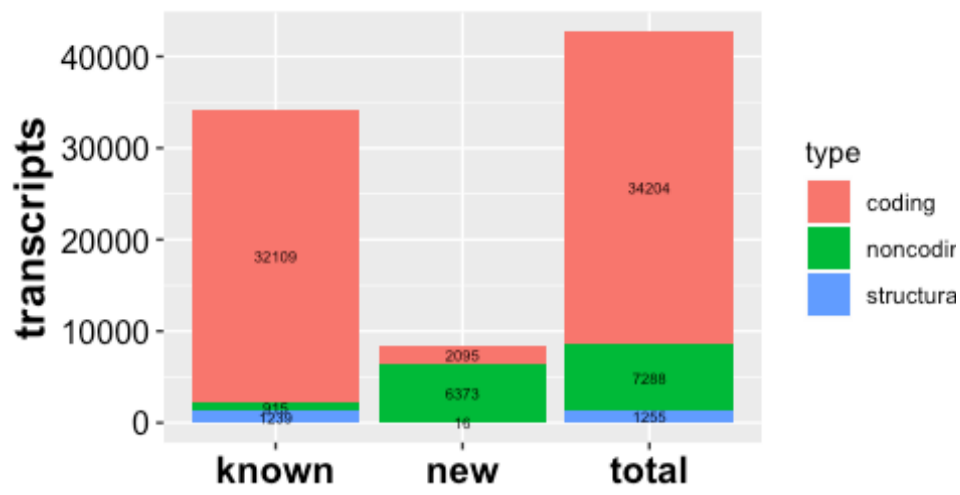
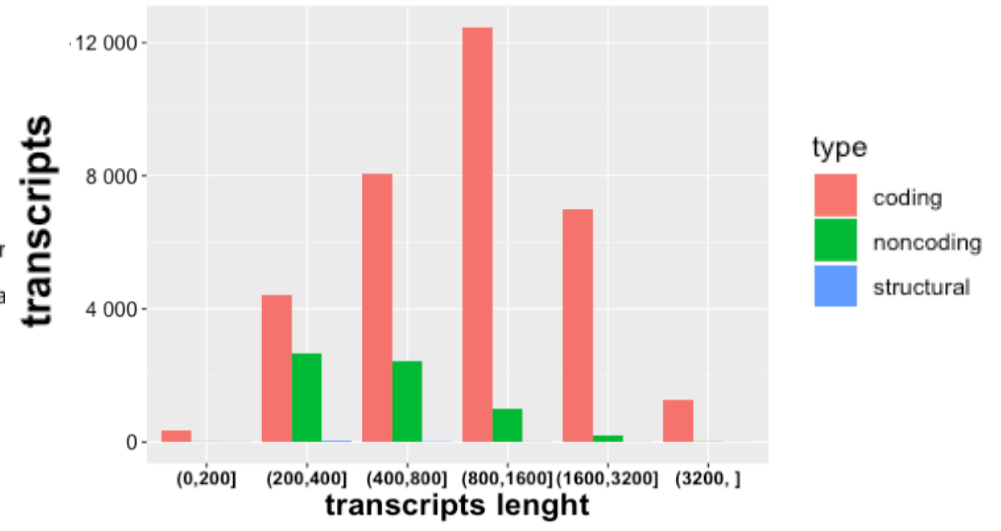
630 **Supplementary Table S3.** Known and newly annotated carrot genes classified as  
631 coding, noncoding and structural transcripts.

632 **Supplementary Table S4.** Overall differentially expressed genes (DEGs) list, including  
633 statistical tests, *cis*-located sequences, gene lengths and gene products. The 21 identified  
634 lncNAT/coding transcript pairs are sorted on top of the list.

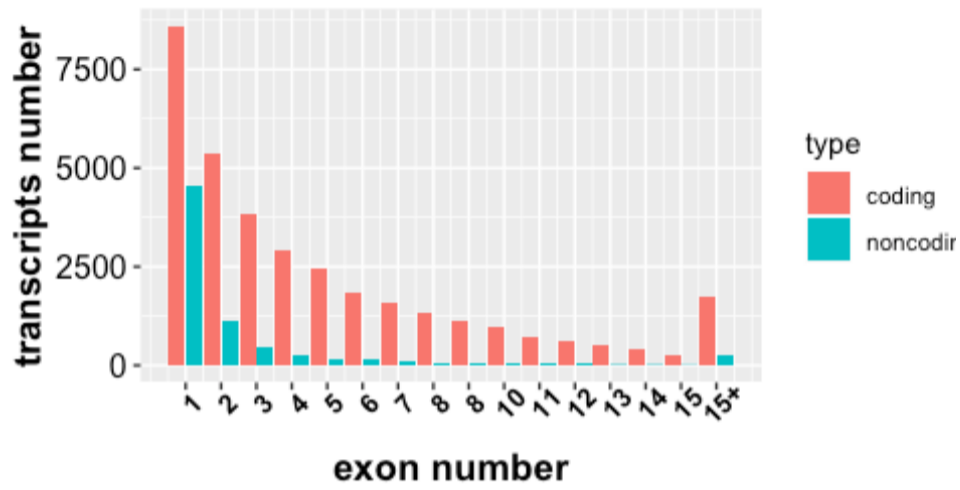
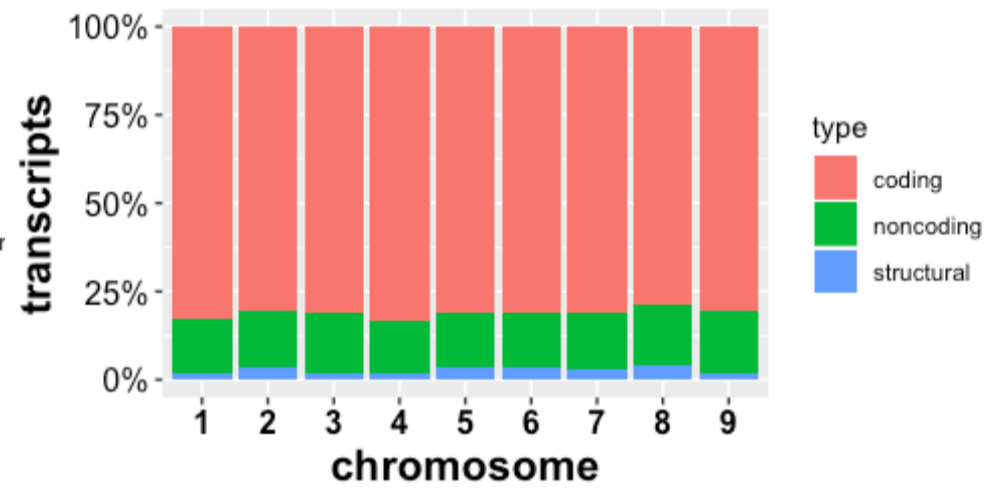
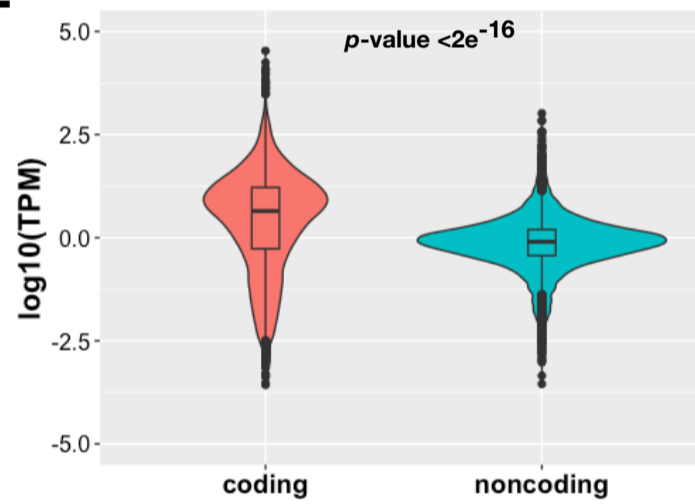
635 **Supplementary Table S5.** Pearson and Spearman correlation coefficients between the  
636 expression levels of the 19 identified lncNAT/coding transcript pairs across the 12  
637 analyzed libraries.

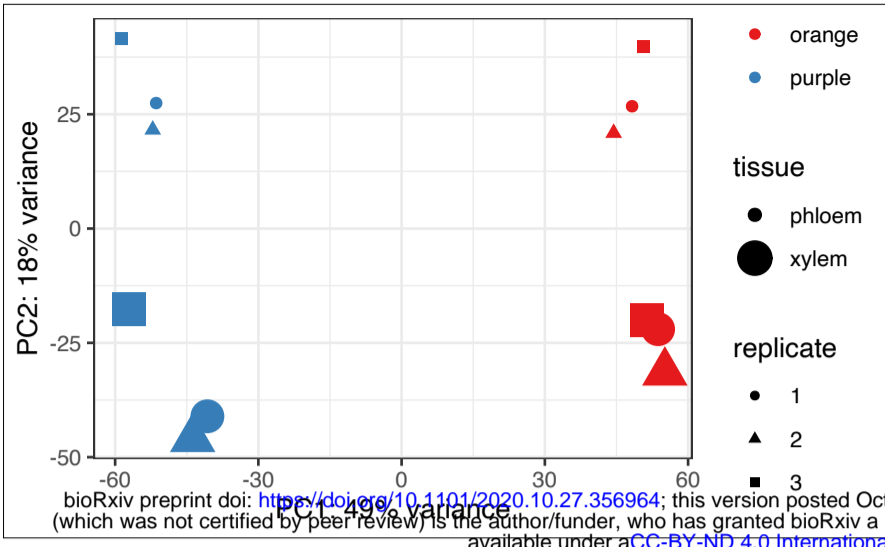
638 **Supplementary File S1.** FASTA sequences of the newly annotated transcripts.



**A****B**

bioRxiv preprint doi: <https://doi.org/10.1101/2020.10.27.356964>; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

**C****D****E**

**A****B**