From bud formation to flowering: transcriptomic state defines the cherry developmental phases of sweet cherry bud dormancy

Noémie Vimont[1,2,3], Mathieu Fouché[1], José Antonio Campoy[4,5,6], Meixuezi Tong[3], Mustapha Arkoun[2], Jean-Claude Yvin[2], Philip A. Wigge[7], Elisabeth Dirlewanger[1], Sandra Cortijo[3#], Bénédicte Wenden[1#]

[1]UMR 1332 BFP, INRA, Univ. Bordeaux, 33882 Villenave d'Ornon, Cedex France; [2]Agro Innovation International - Centre Mondial d'Innovation - Groupe Roullier, 35400 St Malo, France; [3]The Sainsbury Laboratory, University of Cambridge, Cambridge CB2 1LR, United Kingdom; [4] Universidad Politécnica de Cartagena, Cartagena, Spain; [5] Universidad de Murcia, Murcia, Spain; [6]Current address: Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany; [7]Leibniz-Institute für Gemüse- und Zierpflanzenbau (IGZ), Plant Adaptation, Grossbeeren, Germany

#Corresponding authors: sandra.cortijo@slcu.cam.ac.uk; benedicte.wenden@inra.fr

SUMMARY

- Bud dormancy is a crucial stage in perennial trees and allows survival over winter to ensure optimal flowering and fruit production. Recent work highlighted physiological and molecular events occurring during bud dormancy in trees and we aimed to further explore the global transcriptional changes happening throughout dormancy progression.
- Using next-generation sequencing and modelling, we conducted an in-depth transcriptomic analysis for all stages of flower buds in sweet cherry (*Prunus avium* L.) cultivars displaying contrasted stages of bud dormancy.
- We observed that buds in organogenesis, paradormancy, endodormancy and ecodormancy stages are characterised by specific transcriptional states, associated with different pathways. We further identified that endodormancy can be separated in several phases based on the transcriptomic state. We also found that transcriptional profiles of just seven genes are enough to predict the main cherry tree flower bud dormancy stages.
- Our results indicate that transcriptional changes happening during dormancy are robust and conserved between different sweet cherry cultivars. Our work also sets the stage for the development of a fast and cost effective diagnostic tool to molecularly define the flower bud stages in cherry trees.

KEY WORDS: Transcriptomic, RNA sequencing, time course, *Prunus avium* L., prediction, seasonal timing

37 **INTRODUCTION**

38

39 Temperate trees face a wide range of environmental conditions including highly contrasted
40 seasonal changes. Among the strategies to enhance survival under unfavourable climatic conditions,
41 bud dormancy is crucial for perennial plants since its progression over winter is determinant for
42 optimal growth, flowering and fruit production during the subsequent season. Bud dormancy has long
43 been compared to an unresponsive physiological phase, in which metabolic processes within the buds
44 are halted by cold temperature. However, several studies have shown that bud dormancy progression
45 can be affected in a complex way by temperature and photoperiod (Heide & Prestrud, 2005; Allona *et*
46 *al.*, 2008; Olsen, 2010; Cooke *et al.*, 2012; Maurya *et al.*, 2018). Bud dormancy has traditionally been
47 separated into three main phases: (i) paradormancy, also named "summer dormancy" (Cline &
48 Deppong, 1999); (ii) endodormancy, mostly triggered by internal factors; and (iii) ecodormancy,
49 controlled by external factors (Lang *et al.*, 1987; Considine & Considine, 2016). Progression through
50 endodormancy requires cold accumulation whereas warmer temperatures, i.e. heat accumulation, drive
51 the competence to resume growth over the ecodormancy phase. Dormancy is thus highly dependent
52 on external temperatures, and changes in seasonal timing of bud break and blooming have been
53 reported in relation with global warming. Notably, advances in bud break and blooming dates in spring
54 have been observed in the northern hemisphere, thus increasing the risk of late frost damages (Badeck
55 *et al.*, 2004; Menzel *et al.*, 2006; Vitasse *et al.*, 2014; Fu *et al.*, 2015; Bigler & Bugmann, 2018) while
56 insufficient cold accumulation during winter may lead to incomplete dormancy release associated with
57 bud break delay and low bud break rate (Erez, 2000; Atkinson *et al.*, 2013). These phenological
58 changes directly impact the production of fruit crops, leading to large potential economic losses
59 (Snyder & de Melo-abreu, 2005). Consequently, it becomes urgent to acquire a better understanding
60 of bud responses to temperature stimuli in the context of climate change in order to tackle fruit losses
61 and anticipate future production changes.
62 In the recent years, an increasing number of studies have investigated the physiological and molecular
63 mechanisms of bud dormancy transitions in perennials using RNA sequencing technology, thereby
64 giving a new insight into potential pathways involved in dormancy. The results suggest that the
65 transitions between the three main bud dormancy phases (para-, endo- and eco- dormancy) are
66 mediated by pathways related to phytohormones (Zhong *et al.*, 2013; Chao *et al.*, 2017; Khalil-Ur-
67 Rehman *et al.*, 2017; Zhang *et al.*, 2018), carbohydrates (Min *et al.*, 2017; Zhang *et al.*, 2018),
68 temperature (Ueno *et al.*, 2013; Paul *et al.*, 2014), photoperiod (Lesur *et al.*, 2015), reactive oxygen
69 species (Takemura *et al.*, 2015; Zhu *et al.*, 2015), water deprivation (Lesur *et al.*, 2015), cold
70 acclimation and epigenetic regulation (Kumar *et al.*, 2016). Owing to these studies, a better

71 understanding of bud dormancy has been established in different perennial species (see for example,
72 the recent reviews (Beauvieux *et al.*, 2018; Lloret *et al.*, 2018; Falavigna *et al.*, 2019). However we
73 are still missing a fine-resolution temporal understanding of transcriptomic changes happening over
74 the entire bud development, from bud organogenesis to bud break.
75 Indeed, the small number of sampling dates in existing studies seems to be insufficient to capture all
76 the information about changes occurring throughout the dormancy cycle as it most likely corresponds
77 to a chain of biological events rather than an on/off mechanism. Many unresolved questions remain:
78 What are the fine-resolution dynamics of gene expression related to dormancy? Are specific sets of
79 genes associated with dormancy stages? Since the timing for the response to environmental cues is
80 cultivar-dependant (Campoy *et al.*, 2011; Wenden *et al.*, 2017), are transcriptomic profiles during
81 dormancy different in cultivars with contrasted flowering date?
82 To explore these mechanisms, we conducted a transcriptomic analysis of sweet cherry (*Prunus*
83 *avium* L.) flower buds from bud organogenesis until the end of bud dormancy using next-generation
84 sequencing. Sweet cherry is a perennial species highly sensitive to temperature (Heide, 2008) and we
85 focused on three sweet cherry cultivars displaying contrasted flowering dates and response to
86 environmental conditions. We carried out a fine-resolution time-course spanning the entire bud
87 development, from flower organogenesis in July to spring in the following year when flowering occurs,
88 encompassing para-, enco- and ecodormancy phases. Our results indicate that transcriptional changes
89 happening during dormancy are conserved between different sweet cherry cultivars, opening the way
90 to the identification of key factors involved in the progression through bud dormancy.
91
92 **MATERIAL AND METHODS**
93
94 ***Plant material***
95 Branches and flower buds were collected from four different sweet cherry cultivars with contrasted
96 flowering dates: 'Cristobalina', 'Garnet', 'Regina' and 'Fertard', which display extra-early, early, late
97 and very late flowering dates, respectively. 'Cristobalina', 'Garnet', 'Regina' trees were grown in an
98 orchard located at the Fruit Experimental Unit of INRA in Bourran (South West of France, 44° 19′ 56″
99 N, 0° 24′ 47″ E), under the same agricultural practices. 'Fertard' trees were grown in a nearby orchard
100 at the Fruit Experimental Unit of INRA in Toulenne, near Bordeaux (48° 51′ 46″ N, 2° 17′ 15″ E).
101 During the first sampling season (2015/2016), ten or eleven dates spanning the entire period from
102 flower bud organogenesis (July 2015) to bud break (March 2016) were chosen for RNA sequencing
103 (Table S1; Fig. 1a), while bud tissues from 'Fertard' were sampled in 2015/2016 (12 dates) and
104 2017/2018 (7 dates) for validation by qRT-PCR (Table S1). For each date, flower buds were sampled

105 from different trees, each tree corresponding to a biological replicate. Upon harvesting, buds were
106 flash frozen in liquid nitrogen and stored at -80°C prior to performing RNA-seq.

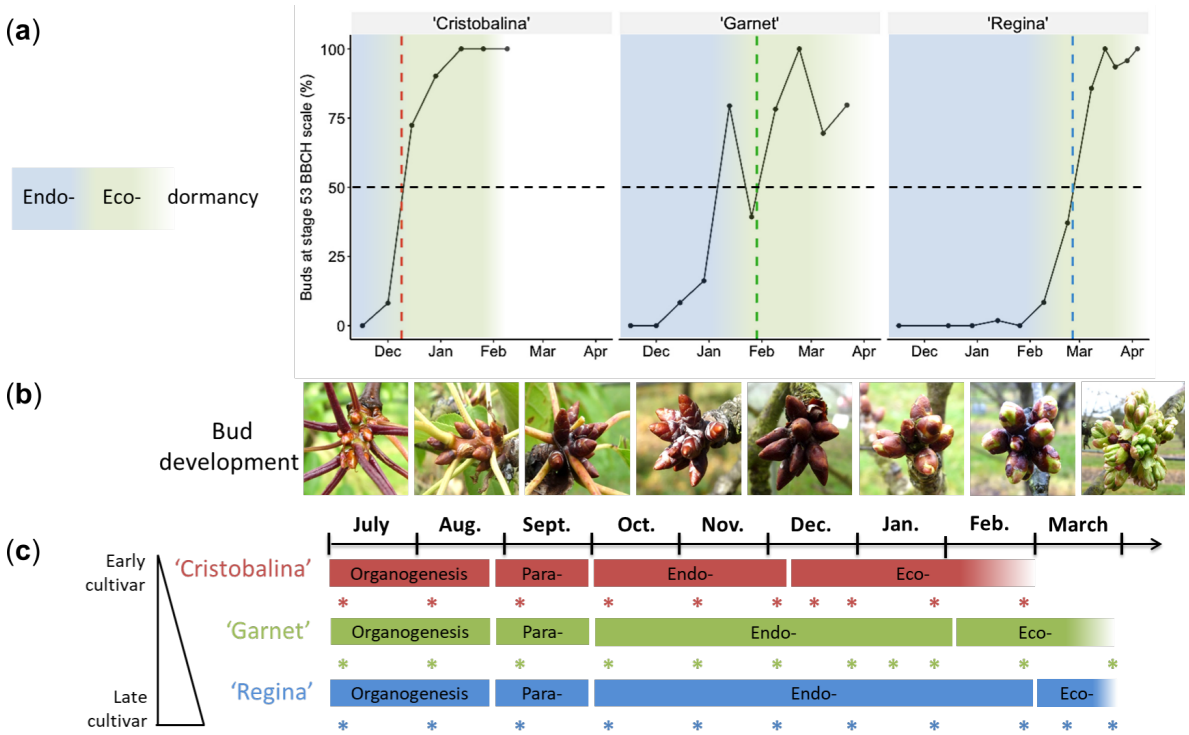

Fig 1 Dormancy status under environmental conditions and RNA-seq sampling dates
(a) Evaluation of bud break percentage under forcing conditions was carried out for three sweet cherry cultivars displaying different flowering dates in 'Cristobalina', 'Garnet' and 'Regina' for the early, medium and late cultivar, respectively. The coloured dotted line corresponds to the dormancy release date, estimated at 50% of buds at BBCH stage 53 (Meier, 2001). (b) Pictures of the sweet cherry buds corresponding to the different sampling dates. (c) Sampling time points for the transcriptomic analysis are represented by coloured stars. Red for 'Cristobalina, green for 'Garnet' and blue for 'Regina'.

107

108 ***Measurements of bud break and estimation of the dormancy release date***

109 For the two sampling seasons, 2015/2016 and 2017/2018, three branches bearing floral buds were
110 randomly chosen fortnightly from 'Cristobalina', 'Garnet', 'Regina' and 'Fertard' trees, between
111 November and flowering time (March-April). Branches were incubated in water pots placed under
112 forcing conditions in a growth chamber (25°C, 16h light/ 8h dark, 60-70% humidity). The water was
113 replaced every 3-4 days. After ten days under forcing conditions, the total number of flower buds that
114 reached the BBCH stage 53 (Meier, 2001; Fadón *et al.*, 2015) was recorded. The date of dormancy
115 release was estimated as the date when the percentage of buds at BBCH stage 53 was above 50% after
116 ten days under forcing conditions (Fig. 1a).

117

118 ***RNA extraction and library preparation***

119 Total RNA was extracted from 50-60 mg of frozen and pulverised flower buds using RNeasy Plant
120 Mini kit (Qiagen) with minor modification: 1.5% PVP-40 was added in the extraction buffer RLT.
121 RNA quality was evaluated using Tapestation 4200 (Agilent Genomics). Library preparation was
122 performed on 1 µg of high quality RNA (RNA integrity number equivalent superior or equivalent to
123 8.5) using the TruSeq Stranded mRNA Library Prep Kit High Throughput (Illumina cat. no. RS-122-
124 2103) for 'Cristobalina', 'Garnet' and 'Regina' cultivars. DNA quality from libraries was evaluated
125 using Tapestation 4200. The libraries were sequenced on a NextSeq500 (Illumina), at the Sainsbury
126 Laboratory Cambridge University (SLCU), using paired-end sequencing of 75 bp in length.
127
128 ***Mapping and differential expression analysis***
129 The raw reads obtained from the sequencing were analysed using several publicly available software
130 and in-house scripts. The quality of reads was assessed using FastQC
131 (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and possible adaptor contaminations and low
132 quality trailing sequences were removed using Trimmomatic (Bolger *et al.*, 2014). Trimmed reads
133 were mapped to the peach (*Prunus persica* (L) Batsch) reference genome v.2 (Verde *et al.*, 2017) using
134 Tophat (Trapnell *et al.*, 2009). Possible optical duplicates were removed using Picard tools
135 (https://github.com/broadinstitute/picard). The total number of mapped reads of each samples are
136 given in Table S2. For each gene, raw read counts and TPM (Transcripts Per Million) numbers were
137 calculated (Wagner, 2003).
138 We performed a differential expression analysis on data obtained from the 'Garnet' samples. First,
139 data were filtered by removing lowly expressed genes (average read count < 3), genes not expressed
140 in most samples (read counts = 0 in more than 75% of the samples) and genes presenting little ratio
141 change (coefficient of variation < 0.3). Then, differentially expressed genes (DEGs) between bud
142 stages (organogenesis – 6 biological replicates, paradormancy – 3 biological replicates, endodormancy
143 – 10 biological replicates, dormancy breaking – 6 biological replicates, ecodormancy – 6 biological
144 replicates, see Table S1) were assessed using DEseq2 R Bioconductor package (Love *et al.*, 2014), in
145 the statistical software R (R Core Team 2018), on filtered data. Genes with an adjusted *p-value* (padj)
146 < 0.05 were assigned as DEGs (Table S3). To enable researchers to access this resource, we have
147 created a graphical web interface to allow easy visualisation of transcriptional profiles throughout
148 flower bud dormancy in the three cultivars for genes of interest (bwenden.shinyapps.io/DorPatterns/).
149
150 ***Principal component analyses and hierarchical clustering***
151 Distances between the DEGs expression patterns over the time course were calculated based on
152 Pearson's correlation on 'Garnet' TPM values. We applied a hierarchical clustering analysis on the

153 distance matrix to define ten clusters (Table S3). For expression patterns representation, we normalized
154 the data using *z-score* for each gene:

155 $$z\ score = \frac{(TPM_{ij} - mean_i)}{Standard\ Deviation}$$

156 where TPM_{ij} is the TPM value of the gene *i* in the sample *j*, $mean_i$ and $standard\ deviation_i$ are the *mean*
157 and *standard deviation* of the TPM values for the gene *i* over all samples.
158 Principal component analyses (PCA) were performed on TPM values from different datasets using the
159 *prcomp* function from R.
160 For each cluster, using data for 'Garnet', 'Regina' and 'Cristobalina', mean expression pattern was
161 calculated as the mean *z-score* value for all genes belonging to the cluster. We then calculated the
162 Pearson's correlation between the *z-score* values for each gene and the mean *z-score* for each cluster.
163 We defined the marker genes as genes with the highest correlation values, i.e. genes that represent the
164 best the average pattern of the clusters. Keeping in mind that the marker genes should be easy to
165 handle, we then selected the optimal marker genes displaying high expression levels while not
166 belonging to extended protein families.
167
168 ***Motif and transcription factor targets enrichment analysis***
169 We performed enrichment analysis on the DEG in the different clusters for transcription factor targets
170 genes and target motifs.
171 Motif discovery on the DEG set was performed using Find Individual Motif occurrences (FIMO)
172 (Grant *et al.*, 2011). Motif list available for peach was obtained from PlantTFDB 4.0 (Jin *et al.*, 2017).
173 To calculate the overrepresentation of motifs, DEGs were grouped by motif (grouping several genes
174 and transcripts in which the motif was found). Overrepresentation of motifs was performed using
175 hypergeometric tests using Hypergeometric {stats} available in R. Comparison was performed for the
176 number of appearances of a motif in one cluster against the number of appearances on the overall set
177 of DEG. As multiple testing implies the increment of false positives, *p-values* obtained were corrected
178 using False Discovery Rate (Benjamini & Hochberg, 1995) correction method using p.adjust{stats}
179 function available in R.
180 A list of predicted regulation between transcription factors and target genes is available for peach in
181 PlantTFDB (Jin *et al.*, 2017). We collected the list and used it to analyse the overrepresentation of
182 genes targeted by TF, using Hypergeometric {stats} available in R, comparing the number of
183 appearances of a gene controlled by one TF in one cluster against the number of appearances on the
184 overall set of DEG. *p-values* obtained were corrected using a false discovery rate as described above.
185 Predicted gene homology to *Arabidopsis thaliana* and functions were retrieved from the data files

186 available for *Prunus persica* (GDR,
187 https://www.rosaceae.org/species/prunus_persica/genome_v2.0.a1).
188
189 ***GO enrichment analysis***
190 The list for the gene ontology (GO) terms was retrieved from the database resource PlantRegMap (Jin
191 *et al.*, 2017). Using the topGO package (Alexa & Rahnenführer, 2018), we performed an enrichment
192 analysis on GO terms for biological processes, cellular components and molecular functions based on
193 a classic Fisher algorithm. Enriched GO terms were filtered with a *p-value* < 0.005 and the ten GO
194 terms with the lowest *p-value* were selected for representation.
195
196 ***Marker genes qRT-PCR analyses***
197 cDNA was synthetised from 1µg of total RNA using the iscript Reverse Transcriptase Kit (Bio-rad
198 Cat no 1708891) in 20 µl of final volume. 2 µL of cDNA diluted to a third was used to perform the
199 qPCR in a 20 µL total reaction volume. qPCRs were performed using a Roche LightCycler 480. Three
200 biological replicates for each sample were performed. Primers used in this study for qPCR are:
201 *PavCSLG3* F:CCAACCAACAAAGTTGACGA, R:CAACTCCCCCAAAAAGATGA; *PavMEE9*:
202 F:CTGCAGCTGAACTGGAACAG, R:ACTCATCCATGGCACTCTCC; *PavSRP*:
203 F:ACAGGATCTGGAAAGCCAAG, R:AGGGTGGCTCTGAAACACAG; *PavTCX2*:
204 F:CTTCCCACAACGCCTTTACG, R:GGCTATGTCTCTCAAACTTGGA; *PavGH127*:
205 F:GCCATTGGTTGTAGGGTTTG, R:ATCCCATTCAGCATTCGTTC; *PavUDP-GALT1*
206 F:CAATGTTGCTGGAAACCTCA, R:GTTATTCCACATCCGACAGC; *PavPP2C*
207 F:CTGTGCCTGAAGTGACACAGA, R:CTGCACTGCTTCTTGATTTG; *PavRPII*
208 F:TGAAGCATACACCTATGATGATGAAG, R:CTTTGACAGCACCAGTAGATTCC; *PavEF1*
209 F:CCCTTCGACTTCCACTTCAG, R:CACAAGCATACCAGGCTTCA. Primers were tested for non-specific
210 products previously by separation on 1.5% agarose gel electrophoresis and by sequencing each
211 amplicon. Realtime data were analyzed using custom R scripts. Expression was estimated for each
212 gene in each sample using a cDNA standard curve. For the visualization of the marker genes' relative
213 expression, we normalized the qRT-PCR results for each marker gene by the average qRT-PCR data
214 for the reference genes *PavRPII* and *PavEF1.*
215
216 ***Bud stage predictive modelling***
217 In order to predict the bud stage based on the marker genes transcriptomic data, we used TPM values
218 for the marker genes to train a multinomial logistic regression. First, all samples were projected into a
219 2-dimensional space using PCA, to transform potentially correlated data to an orthogonal space. The

220 new coordinates were used to train and test the model to predict the five bud stage categories, using
221 the *LogisticRegressionCV* function from the scikit-learn Python package (Pedregosa *et al.,* 2011). The
222 model was 4-fold cross-validated to ensure the robustness of the coefficients and to reduce overfitting.
223 The model accuracy was calculated as the percentage of correct predicted stages in the RNA-seq
224 testing set. In addition, we tested the model on qRT-PCR data for 'Fertard' samples. For the modelling
225 purposes, expression data for the seven marker genes were normalized by the expression
226 corresponding to the October sample. We chose the date of October as the reference because it
227 corresponds to the beginning of dormancy and it was available for all cultivars. For each date, the
228 October-normalized expression values of the seven marker genes were projected in the PCA 2-
229 dimension plan calculated for the RNA-seq data and they were tested against the model trained on
230 'Cristobalina', 'Garnet' and 'Regina' RNA-seq data.
231
232 **RESULTS**
233
234 **Transcriptome accurately captures the dormancy state**
235 In order to define transcriptional changes happening over the sweet cherry flower bud
236 development, we performed a transcriptomic-wide analysis using next-generation sequencing from
237 bud organogenesis to flowering. According to bud break percentage (Fig. 1a), morphological
238 observations (Fig. 1b), average temperatures (Fig. S1) and descriptions from Lang et *al.,* (1987), we
239 assigned five main stages to the early flowering cultivar 'Garnet' flower buds samples (Fig. 1b): i)
240 flower bud organogenesis occurs in July and August, ii) paradormancy corresponds to the period of
241 growth cessation in September, iii) during the endodormancy phase, initiated in October, buds are
242 unresponsive to forcing conditions therefore the increasing bud break percentage under forcing
243 conditions suggests that endodormancy was released on January 29th, 2016, thus corresponding to iv)
244 dormancy breaking, and v) ecodormancy starting from the estimated dormancy release date until
245 flowering.
246 We identified 6,683 genes that are differentially expressed (DEGs) between the defined bud
247 stages for the sweet cherry cultivar 'Garnet' (Table S3). When projected into a two-dimensional space
248 (Principal Component Analysis, PCA), data for these DEGs show that transcriptomes of samples
249 harvested at a given date are projected together (Fig. 2), showing the high quality of the biological
250 replicates and that different trees are in a very similar transcriptional state at the same date. Very
251 interestingly, we also observe that flower bud states are clearly separated on the PCA, with the
252 exception of organogenesis and paradormancy, which are projected together (Fig. 2). The first
253 dimension of the analysis (PC1) explains 41,63% of the variance and clearly represents the strength of

254 bud dormancy where samples on the right of the axis are in endodormancy or dormancy breaking
255 stages. The second dimension of the analysis (PC2) explains 20.24% of the variance and distinguishes
256 two main phases of the bud development: before and after dormancy breaking. We obtain very similar
257 results when performing the PCA on all genes (Fig. S2). These results indicate that the transcriptional
258 state of DEGs accurately captures the dormancy state of flower buds.

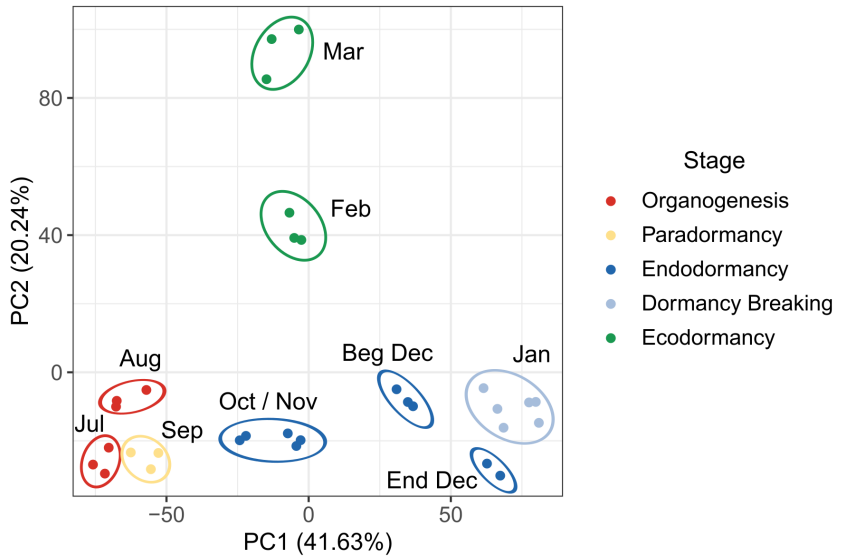

Fig 2 Separation of samples by dormancy stage using differentially expressed genes
The principal component analysis was conducted on the TPM (transcripts per millions reads) values for the differentially expressed genes in the cultivar 'Garnet' flower buds, sampled on three trees between July and March.

259

260 **Bud stage-dependent transcriptional activation and repression are associated with different**
261 **pathways**

262 We further investigated whether specific genes or signalling pathways could be associated with
263 the different flower bud stages. Indeed, the expression of genes grouped in ten clusters clearly shows
264 distinct expression profiles throughout the bud development (Fig. 3). Overall, three main types of
265 clusters can be discriminated: the ones with a maximum expression level during organogenesis and
266 paradormancy (cluster 1: 1,549 genes; cluster 2: 70 genes; cluster 3: 113 genes; cluster 4: 884 genes
267 and cluster 10: 739 genes, Fig. 3), the clusters with a maximum expression level during endodormancy
268 and around the time of dormancy breaking (cluster 5: 156 genes; cluster 6: 989 genes ; cluster 7: 648
269 genes and cluster 8: 612 genes, Fig. 3), and finally the clusters with a maximum expression level during
270 ecodormancy (cluster 9: 924 genes and cluster 10, Fig. 3). This result shows that different groups of
271 genes are associated with these three main flower bud phases. Interestingly, we also observed that,
272 during the endodormancy phase, some genes are expressed in October and November then repressed

273 in December (cluster 4, Fig. 3), whereas another group of genes is expressed in December (clusters 8,
274 5, 6 and 7, Fig. 3) therefore separating endodormancy in two distinct phases.

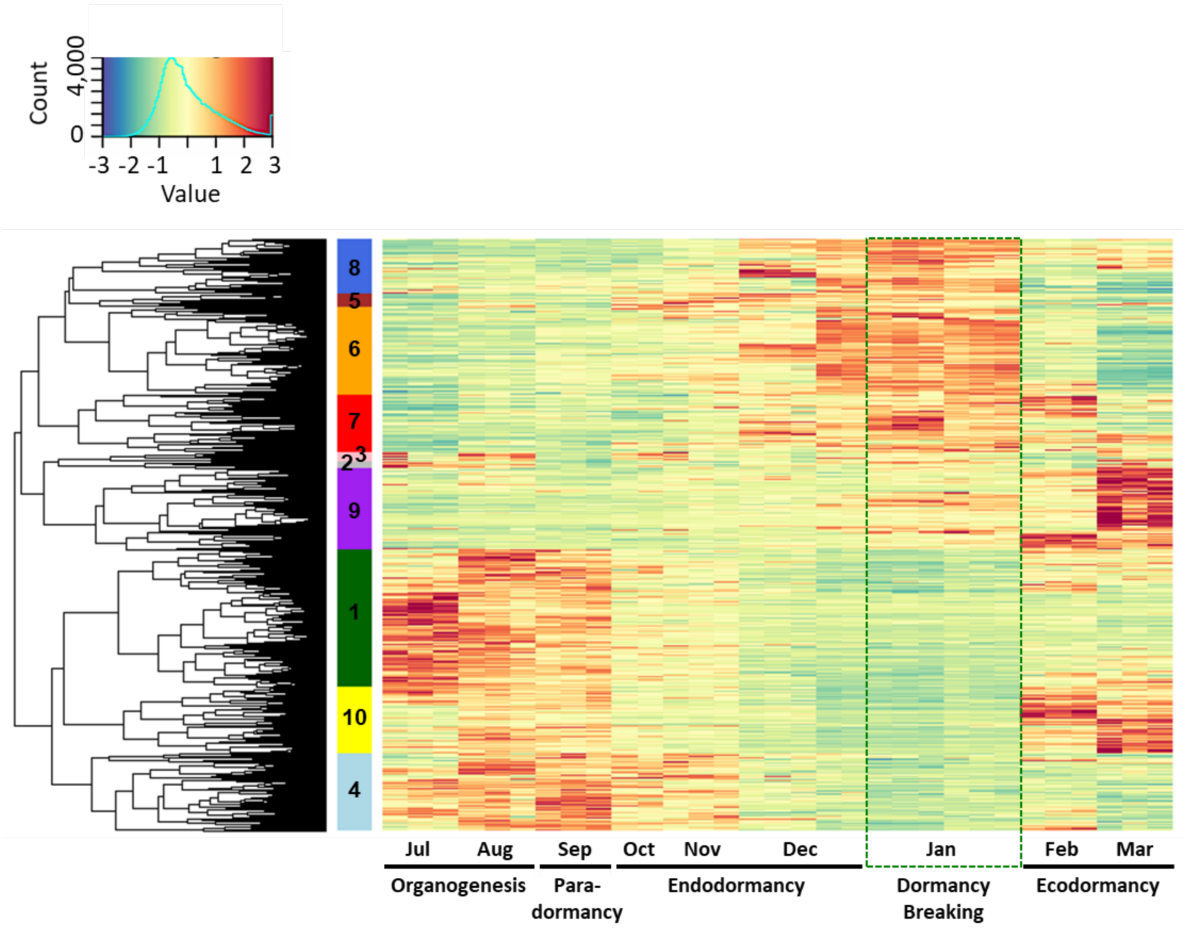

Fig 3 Clusters of expression patterns for differentially expressed genes in the sweet cherry cultivar 'Garnet'
Heatmap for 'Garnet' differentially expressed genes during bud development. Each column corresponds to the gene expression for flower buds from one single tree at a given date. Clusters are ordered based on the chronology of the expression peak (from earliest – July, 1-dark green cluster – to latest – March, 9 and 10). Expression values were normalized and *z-scores* are represented here.

275 In order to explore the functions and pathways associated with the gene clusters, we performed
276 a GO enrichment analysis (Fig. 4, Fig. S3). GO terms associated with the response to stress as well as
277 biotic and abiotic stimuli were enriched in the clusters 2, 3 and 4, with genes mainly expressed during
278 organogenesis and paradormancy. During endodormancy (cluster 5), an enrichment for genes involved
279 in response to nitrate and nitrogen compounds was spotted. On the opposite, at the end of the
280 endodormancy phase (cluster 6, 7 and 8), we highlighted different enrichments in GO terms linked to
281 basic metabolisms such as nucleic acid metabolic processes or DNA replication but also to response
282 to alcohol and abscisic acid. Finally, during ecodormancy, genes in cluster 9 and 10 are enriched in
283 functions associated with transport, cell wall biogenesis as well as oxidation-reduction processes (Fig.

4, Fig. S3). These results show that different functions and pathways are specific to flower bud development stages.

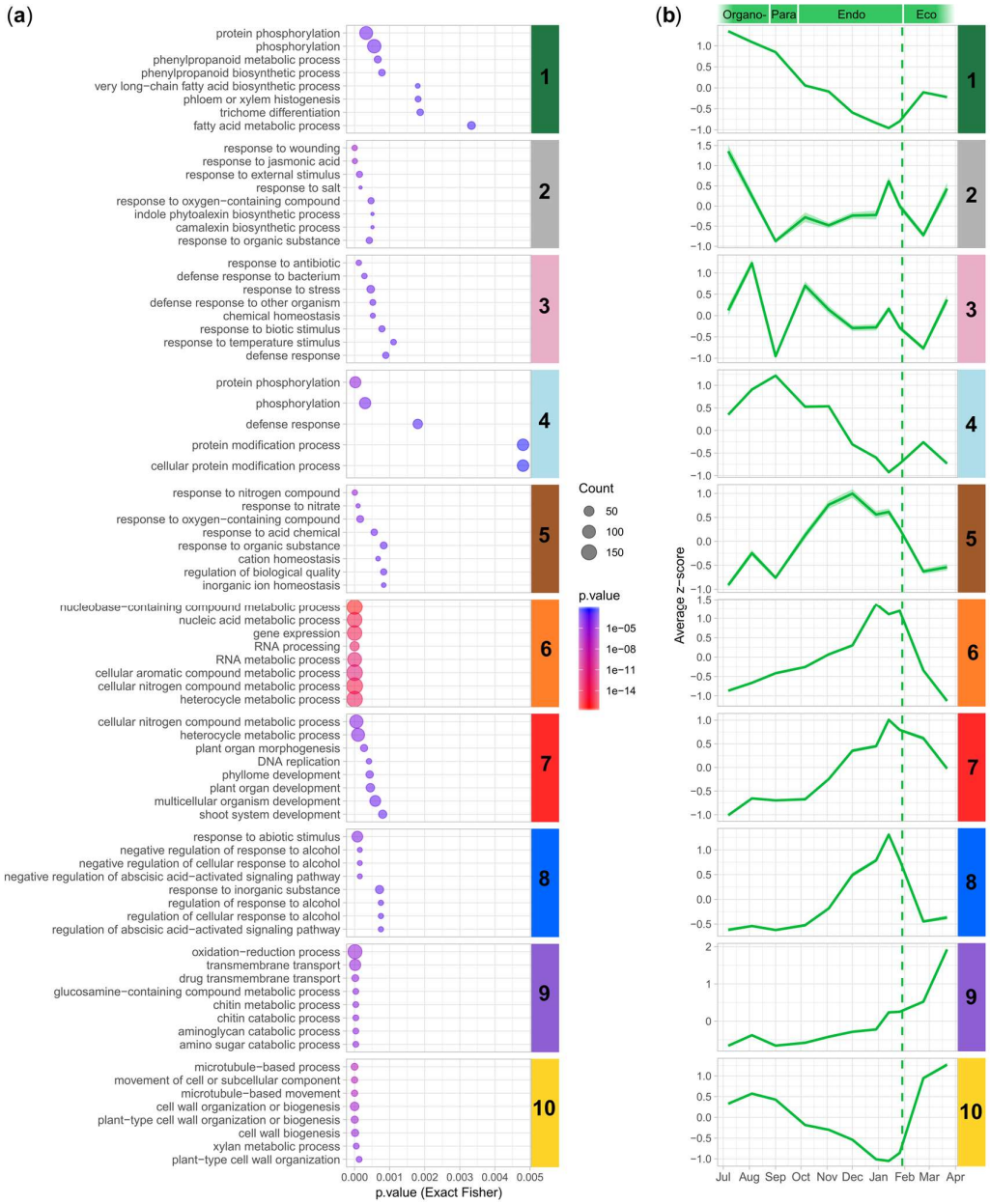

Fig 4 Enrichments in gene ontology terms for biological processes and average expression patterns in the different clusters in the sweet cherry cultivar 'Garnet'

(a) Using the topGO package (Alexa & Rahnenführer, 2018), we performed an enrichment analysis on GO terms for biological processes based on a classic Fisher algorithm. Enriched GO terms with the lowest *p-value* were selected for representation. Dot size represent the number of genes belonging to the clusters associated with the GO term. (b) Average z-score values for each cluster. The coloured dotted line corresponds to the estimated date of dormancy release.

286 **Table 1. Enrichment in transcription factor targets in the different clusters**

287 Based on the gene regulation information available for peach in PlantTFDB (Jin *et al.*, 2017), overrepresentation of genes targeted by transcription factors was
288 performed using hypergeometric tests. *p-values* obtained were corrected using a false discovery rate: (***): adj. p-value < 0.001; (**): adj. p-value < 0.01; (*):
289 adj. p-value < 0.05.

	Gene Name	gene id	Transcription Factor Cluster	Predicted TF family	Arabidopsis homologous	Predicted function	Enrichment p value	Enrichment adjusted p value	
1 - Dark green	PavMYB63	Prupe.4G136300	1 - Dark green	MYB	AT1G79180	Myb-related protein	2,1E-05	6,7E-03	(**)
	PavMYB93	Prupe.6G188300	1 - Dark green	MYB	AT1G34670	Myb-related protein	9,0E-04	3,2E-02	(*)
	PavMYB40	Prupe.3G299000	8 - royal blue	MYB	AT5G14340	Myb-related protein	2,7E-04	1,7E-02	(*)
	PavMYB17	Prupe.2G164300	-	MYB	AT3G61250	Myb-related protein	6,8E-05	7,2E-03	(**)
	PavMYB94	Prupe.5G193200	-	MYB	AT3G47600	Myb-related protein	9,0E-05	7,2E-03	(**)
	PavMYB60	Prupe.7G018400	-	MYB	AT1G08810	Myb-related protein	7,0E-05	7,2E-03	(**)
	PavMYB61	Prupe.6G303300	-	MYB	AT1G09540	Myb-related protein	4,0E-04	2,1E-02	(*)
	PavMYB3	Prupe.1G551400	-	MYB	AT1G22640	Myb-related protein	6,0E-04	2,8E-02	(*)
	PavMYB67	Prupe.4G126900	-	MYB	AT3G12720	Myb-related protein	7,8E-04	3,1E-02	(*)
2 - grey		Prupe.1G122800	-	CAMTA	AT4G16150	Calmodulin-binding transcription activator	3,1E-05	8,0E-03	(**)
3 - pink	PavWRKY40	Prupe.3G098100	3 - pink	WRKY	AT1G80840	WRKY transcription factor	8,4E-05	1,2E-02	(*)
		Prupe.1G122800	-	CAMTA	AT4G16150	Calmodulin-binding transcription activator	4,9E-09	1,4E-06	(***)
	PavWRKY11	Prupe.1G459100	-	WRKY	AT4G31550	WRKY transcription factor	4,7E-04	4,5E-02	(*)
5 - brown	PavCBF4	Prupe.2G289500	-	ERF	AT5G51990	Dehydration-responsive element-binding protein	2,0E-04	5,7E-02	
6 - orange	PavERF110	Prupe.6G165700	8 - royal blue	ERF	AT5G50080	Ethylene-responsive transcription factor	3,1E-04	5,2E-02	
	PavRVE8	Prupe.6G242700	8 - royal blue	MYB	AT3G09600	Homeodomain-like superfamily protein RVE8	4,3E-04	5,2E-02	
	PavRAP2.12	Prupe.3G032300		ERF	AT1G53910	Ethylene-responsive transcription factor	4,9E-04	5,2E-02	
8 - royal blue	PavRVE1	Prupe.3G014900	6 - orange	MYB	AT5G17300	Homeodomain-like superfamily protein RVE1	1,0E-03	3,6E-02	(*)
	PavABI5	Prupe.7G112200	7 - red	bZIP	AT2G36270	ABSCISIC ACID-INSENSITIVE 5	6,6E-05	7,0E-03	(**)
	PavABF2	Prupe.1G434500	8 - royal blue	bZIP	AT1G45249	abscisic acid responsive elements-binding factor	2,4E-06	7,5E-04	(***)
	PavAREB3	Prupe.2G056800	-	bZIP	AT3G56850	ABA-responsive element binding protein	1,4E-05	2,2E-03	(**)
	PavPIL5	Prupe.8G209100	-	bHLH	AT2G20180	phytochrome interacting factor 3-like 5	2,3E-04	1,9E-02	(*)
	PavbZIP16	Prupe.5G027000	-	bZIP	AT2G35530	basic region/leucine zipper transcription factor	4,3E-04	2,7E-02	(*)
	PavSPT	Prupe.7G131400	-	bHLH	AT4G36930	Transcription factor SPATULA	5,6E-04	3,0E-02	(*)
	PavBPE	Prupe.1G263800	-	bHLH	AT1G59640	Transcription factor BPE	1,0E-03	3,6E-02	(*)
	PavPIF4	Prupe.3G179800	-	bHLH	AT2G43010	phytochrome interacting factor 4	9,5E-04	3,6E-02	(*)
	PavGBF3	Prupe.2G182800	-	bZIP	AT2G46270	G-box binding factor 3	1,1E-03	3,6E-02	(*)
9 - purple	PavWRKY50	Prupe.1G407500	-	WRKY	AT5G26170	WRKY transcription factor	1,1E-04	1,8E-02	(*)
	PavWRKY1	Prupe.3G202000	-	WRKY	AT2G04880	WRKY transcription factor	5,8E-05	1,8E-02	(*)
10 - yellow	PavMYB14	Prupe.1G039200	5 - brown	MYB	AT2G31180	Myb-related protein	1,6E-04	3,9E-02	(*)
	PavNAC70	Prupe.8G002500	-	NAC	AT4G10350	NAC domain containing protein	2,4E-04	3,9E-02	(*)

290

291 **Specific transcription factor target genes are expressed during the main flower bud stages**

292 To better understand the regulation of genes that are expressed at different flower bud stages,
293 we investigated the TFs with enriched targets (Table 1) as well as the enriched target promoter motifs
294 (Table S4) in the different gene clusters. Among the genes expressed during the organogenesis and
295 paradormancy phases (clusters 1, 2, 3 and 4), we observed an enrichment for motifs of several MADS-
296 box TFs such as AGAMOUS (AG), APETALA3 (AP3) and SEPALLATA3/AGAMOUS-like 9
297 (SEP3/AGL9) (Table S4), several of them potentially involved in flower organogenesis (Causier *et*
298 *al.*, 2010). On the other hand, for the same clusters, results show an enrichment in MYB-related targets,
299 WRKY and ethylene-responsive element (ERF) binding TFs (Table 1, Table S4). Several members of
300 these TF families have been shown to participate in the response to abiotic factors. Similarly, we found
301 in the cluster 4 target motifs enriched for PavDREB2C (Table S4), potentially involved in the response
302 to cold (Lee *et al.*, 2010). Interestingly, we identified an enrichment in the cluster 5 of targets for
303 CBF4, and of genes with motifs for several ethylene-responsive element binding TFs such as
304 PavDREB2C. We also observed an enrichment in the same cluster for genes with motifs for ABI5
305 (Table S4). All these TFs are involved in the response to cold, in agreement with the fact that genes in
306 the cluster 5 are expressed during endodormancy.

307 Genes belonging to the clusters 6, 7 and 8 are highly expressed during deep dormancy and we
308 found targets and target motifs for many TFs involved in the response to abiotic stresses. For example,
309 we found motifs enriched in the cluster 7 for many TFs of the C2H2 family, which is involved in the
310 response of wide spectrum of stress conditions, such as extreme temperatures, salinity, drought or
311 oxidative stress (Table S4, Kiełbowicz-Matuk, 2012; Liu *et al.*, 2015). Similarly, in the cluster 8, we
312 also identified an enrichment in targets and motifs of many genes involved in the response to abscisic
313 acid (ABA) and to abiotic stimulus, such as *PavABF2*, *PavAREB3*, *PavABI5* and *PavDREB2C*
314 (Koornneef *et al.*, 1998; Lee *et al.*, 2010). We also observe in this same cluster an enrichment for
315 targets of TFs involved in the response to light and temperature, such as *PavPIL5*, *PavSPT*, *PavRVE1*
316 and *PavPIF4* (Table 1, Penfield *et al.*, 2005; Olsen, 2010; Franklin *et al.*, 2011; Doğramacı *et al.*,
317 2014). Interestingly, we found that among the TFs with enriched targets in the clusters, only ten display
318 changes in expression during flower bud development (Table 1, Table S4, Fig. S4), including
319 *PavABF2*, *PavABI5* and *PavRVE1*. Expression profiles for these three genes are very similar, and are
320 also similar to their target genes, with a peak of expression around the estimated dormancy release
321 date, indicating that these TFs are positively regulating their targets (Fig. S4).

322 Finally, genes belonging to the cluster 10 are expressed during ecodormancy and we find an
323 enrichment for targets of PavMYB14 (Table 1). Expression profiles suggest that PavMYB14 represses
324 expression of its target genes during endodormancy (Fig. S4), consistently with the functions of

325 *Arabidopsis thaliana* MYB14 that negatively regulates the response to cold (Chen *et al.*, 2013).
326 Overall, these results show that a small number of TFs specifically regulate target genes during the
327 different flower bud stages.

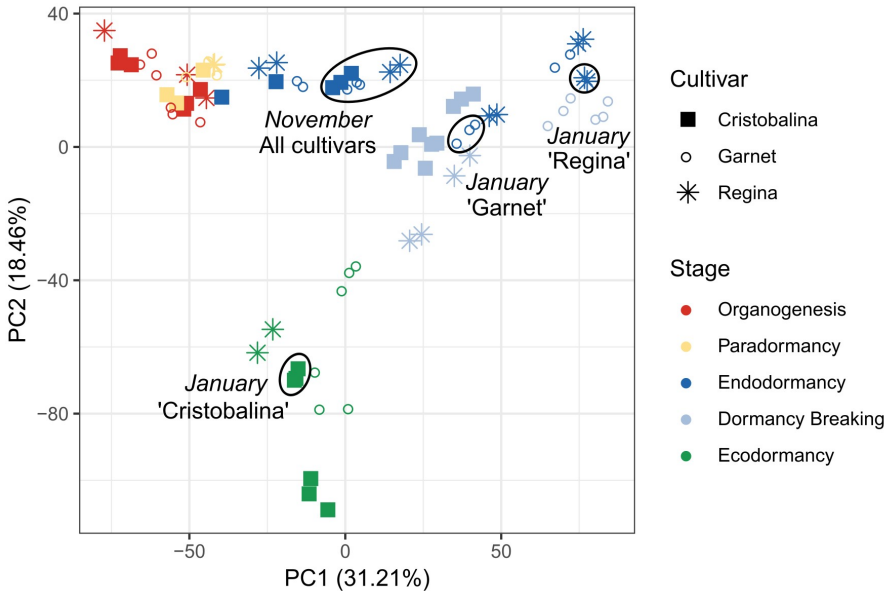

Fig 5 Separation of samples by dormancy stage and cultivar using differentially expressed genes
The principal component analysis was conducted on the TPM (transcripts per millions reads) values for the differentially expressed genes in the flower buds of the cultivars 'Cristobalina' (filled squares), 'Garnet' (empty circles) and 'Regina' (stars). Each point corresponds to one sampling time in a single tree.

328 **Expression patterns highlight bud dormancy similarities and disparities between three cherry**
329 **tree cultivars**

330 Since temperature changes and progression through the flower bud stages are happening
331 synchronously, it is challenging to discriminate transcriptional changes that are mainly associated with
332 one or the other. In this context, we also analysed the transcriptome of two other sweet cherry cultivars:
333 'Cristobalina', characterized by very early flowering dates, and 'Regina', with a late flowering time.
334 The span between flowering periods for the three cultivars is also found in the transition between
335 endodormancy and ecodormancy since ten weeks separated the estimated dates of dormancy release
336 between the cultivars: 9th December 2015 for 'Cristobalina', 29th January 2016 for 'Garnet' and 26th
337 February 2016 for 'Regina' (Fig. 1a). The transition from organogenesis to paradormancy is not well
338 documented and many studies suggest that endodormancy onset is under the strict control of
339 environment. Therefore, we considered that these two transitions occurred at the same time in all three
340 cultivars. However, the two months and half difference in the date of transition from endodormancy
341 to ecodormancy between the cultivars allow us to look for transcriptional changes associated with this
342 transition independently of environmental conditions. To do so, we compared the expression patterns

343 of the previously identified DEGs between the three contrasted cultivars throughout flower bud stages
344 (Fig. 1b). When projected into a PCA 2-components plane, all samples harvested from buds at the
345 same stage cluster together, whatever the cultivar (Fig. 5), suggesting that the stage of the bud has
346 more impact on the transcriptional state than time or external conditions.

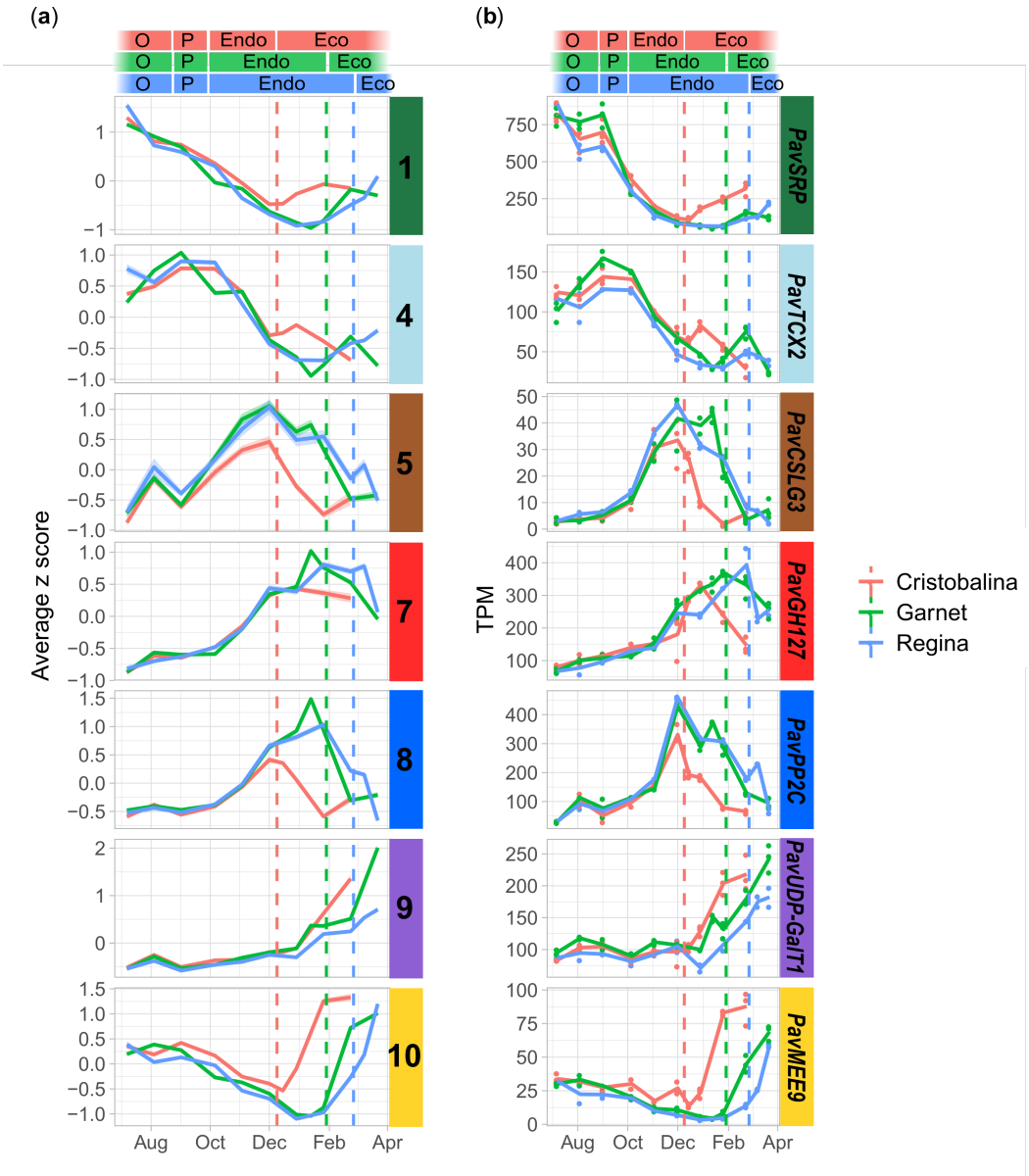

Fig 6 Expression patterns in the selected seven clusters for the three cultivars
Expression patterns were analysed from August to March, covering bud organogenesis (O), paradormancy (P), endodormancy (Endo), and ecodormancy (Eco). Dash lines represent the estimated date of dormancy breaking, in red for 'Cristobalina', green for 'Garnet' and blue for 'Regina'. (a) Average z-score patterns, calculated from the TPM, for the genes belonging to the seven selected clusters and (b) TPM for the seven marker genes from clusters 1, 4, 5, 7, 8, 9 and 10. Lines represent the average TPM, dots are the actual values. *SRP: STRESS RESPONSIVE PROTEIN; TCX2: TESMIN/TSO1-like CXC 2; CSLG3: Cellulose Synthase like G3; GH127: Glycosyl Hydrolase 127; PP2C: Phosphatase 2C; UDP-GalT1: UDP-Galactose transporter 1; MEE9: maternal effect embryo arrest 9.*

347 To go further, we compared transcriptional profiles throughout the time course in all cultivars.
348 For this we analysed the expression profiles in each cultivar for the clusters previously identified for
349 the cultivar 'Garnet' (Fig. 6). Due to the low number of genes, clusters 2, 3 were not further studied in
350 the three cultivars and we considered that the expression patterns for the genes in cluster 6 were
351 redundant with clusters 5 and 7 therefore we simplified the analysis on seven clusters. In general,
352 averaged expression profiles for all clusters are very similar in all three varieties, with the peak of
353 expression happening at a similar period of the year. However, we can distinguish two main phases
354 according to similarities or disparities between cultivars. First, averaged expression profiles are almost
355 similar in all cultivars between July and November. This is especially the case for clusters 1, 4, 7, 8
356 and 9. On the other hand, we can observe a temporal shift in the peak of expression between varieties
357 from December onward for genes in clusters 1, 5, 8 and 10. Indeed, in these clusters, the peak or drop
358 in expression happens earlier in 'Cristobalina', and slightly later in 'Regina' compared to 'Garnet'
359 (Fig. 6), in correlation with their dormancy release dates. These results seem to confirm that the
360 organogenesis and paradormancy phases occur concomitantly in the three cultivars while temporal
361 shifts between cultivars are observed after endodormancy onset. Therefore, similarly to the PCA
362 results (Fig. 5), the expression profile of these genes is more associated with the flower bud stage than
363 with external environmental conditions.
364
365 **Flower bud stage can be predicted using a small set of marker genes**
366 We have shown that flower buds in organogenesis, paradormancy, endodormancy and
367 ecodormancy are characterised by specific transcriptional states. In theory, we could therefore use
368 transcriptional data to infer the flower bud stage. For this, we selected seven marker genes, for clusters
369 1, 4, 5, 7, 8, 9 and 10, that best represent the average expression profiles of their cluster (Fig. 6).
370 Expression for these marker genes not only recapitulates the average profile of the cluster they
371 originate from, but also temporal shifts in the profiles between the three cultivars (Fig. 6b). In order to
372 define if these genes encompass as much information as the full transcriptome, or all DEGs, we
373 performed a PCA of all samples harvested for all three cultivars using expression levels of these seven
374 markers (Fig. S7). The clustering of samples along the two main axes of the PCA using these seven
375 markers is very similar, if not almost identical, to the PCA results obtained using expression for all
376 DEGs (Fig. 5). This indicates that the transcriptomic data can be reduced to only seven genes and still
377 provides accurate information about the flower bud stages.
378 To test if these seven markers can be used to define the flower bud stage, we used a multinomial
379 logistic regression modelling approach to predict the flower bud stage in our dataset based on the
380 expression levels for these seven genes (Fig. 7 and Fig. S8). We obtain a very high model accuracy

381 (100%) when the training and testing sets are randomly picked. These results indicate that the bud
382 stage can be accurately predicted based on expression data by just using seven genes. In order to go
383 further and test the model in an independent experiment, we analysed the expression for the seven
384 marker genes by RT-qPCR on buds sampled from another sweet cherry tree cultivar 'Fertard' for two
385 consecutive years (Fig. 7a). We achieved a high accuracy of 71% for our model when tested on RT-
386 qPCR data to predict the flower bud stage for the 'Fertard' cultivar (Fig. 7c and Fig. S8c). In particular,
387 the chronology of bud stages was very well predicted. This result indicates that these seven genes can
388 be used as a diagnostic tool in order to infer the flower bud stage in sweet cherry trees.

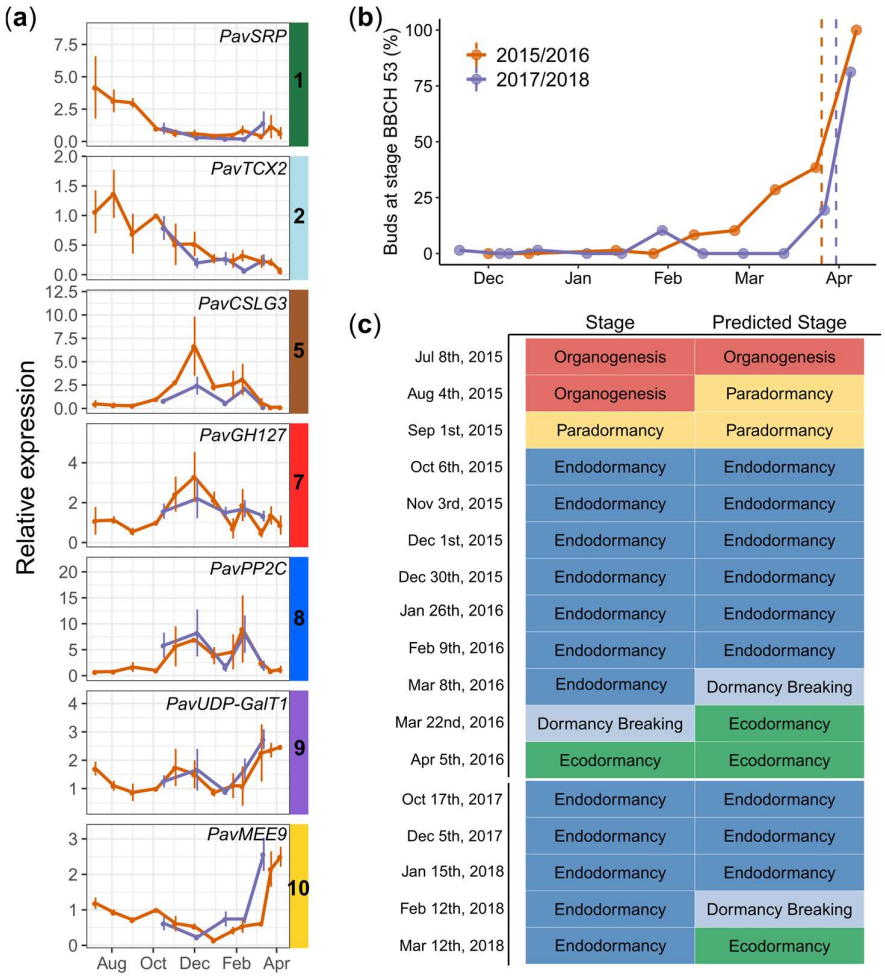

Figure 7. Expression for the seven marker genes allows accurate prediction of the bud dormancy stages in the late flowering cultivar 'Fertard' during two bud dormancy cycles
(a) Relative expressions were obtained by qRT-PCR and normalized by the expression of two reference constitutively expressed genes *PavRPII* and *PavEF1*. (b) Evaluation of the dormancy status in 'Fertard' flower buds during the two seasons using the percentage of open flower buds (BBCH stage 53). (c) Predicted vs experimentally estimated bud stages. *SRP: STRESS RESPONSIVE PROTEIN; TCX2: TESMIN/TSO1-like CXC 2; CSLG3: Cellulose Synthase like G3; GH127: Glycosyl Hydrolase 127; PP2C: Phosphatase 2C; UDP-GalT1: UDP-Galactose transporter 1; MEE9: maternal effect embryo arrest 9.*

Discussion

In this work, we have characterised transcriptional changes at a genome-wide scale happening throughout cherry tree flower bud dormancy, from organogenesis to the end of dormancy. To do this, we have analysed expression in flower buds at 11 dates from July 2015 to March 2016 for three cultivars displaying different dates of dormancy release, generating 82 transcriptomes in total. This resource, with a fine time resolution, reveals key aspects of the regulation of cherry tree flower buds during dormancy (Fig. 8). We have shown that buds in organogenesis, paradormancy, endodormancy and ecodormancy are characterised by distinct transcriptional states (Fig. 2, 3) and we highlighted the different pathways activated during the main cherry tree flower bud dormancy stages (Fig. 4 and Table 1). Finally, we found that just seven genes are enough to accurately predict the main cherry tree flower bud dormancy stages (Fig. 6, 7).

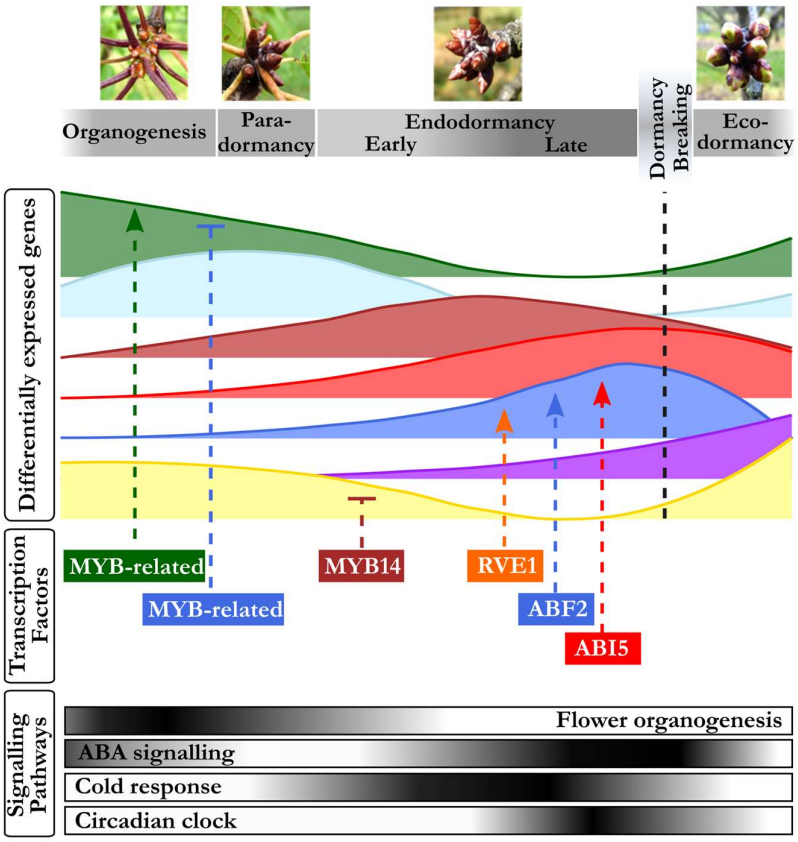

Figure 8. From bud formation to flowering: transcriptomic regulation of flower bud dormancy
Our results highlighted seven main expression patterns corresponding to the main dormancy stages. During organogenesis and paradormancy (July to September), signalling pathways associated with flower organogenesis and ABA signalling are upregulated. Distinct groups of genes are activated during different phases of endodormancy, including targets of transcription factors involved in ABA signalling, cold response and circadian clock. ABA: abscisic acid.

400 **Global lessons from transcriptomic data on the definition of flower bud dormancy stages**

401 Our results show that buds in organogenesis, paradormancy, endodormancy and ecodormancy
402 are characterised by distinct transcriptional states. This result is further supported by the fact that we
403 detected different groups of genes that are specifically expressed at these bud stages (Fig. 3).
404 Specifically, we found that the transcriptional states of flower buds during endodormancy and
405 ecodormancy are very different, indicating that different pathways are involved in these two types of
406 dormancy. This is further supporting previous observations that buds remain in endodormancy and
407 ecodormancy states under the control of different regulation pathways. Indeed, ecodormancy is under
408 the control of external signals and can therefore be reversed by exposure to growth-promotive signals
409 (Lang *et al.*, 1987). On the opposite, endogenous signals control endodormancy onset and maintenance
410 and a complex array of signalling pathways seem to be involved in the response to cold temperatures
411 that subsequently leads to dormancy breaking (see for example Ophir *et al.*, 2009; Horvath, 2009;
412 Considine & Considine, 2016; Singh *et al.*, 2016; Lloret *et al.*, 2018; Falavigna *et al.*, 2019).

413 Another interesting observation is the fact that samples harvested during endodormancy can be
414 separated into two groups based on their transcriptional state: early endodormancy (October and
415 November), and late endodormancy (from December to dormancy breaking). These two groups of
416 samples are forming two distinct clusters in the PCA (Fig. 5), and are associated with different groups
417 of expressed genes. These results indicate that endodormancy could potentially be separated into two
418 periods: early and late endodormancy. However, we have to keep in mind that cold temperatures,
419 below 10°C, only started at the end of November. It is thus difficult to discriminate between
420 transcriptional changes associated with a difference in the bud stage during endodormancy, an effect
421 of the pronounced change in temperatures, or a combination of both. Alternative experiments under
422 controlled environments, similarly to studies conducted on hybrid aspen for example (Ruttink *et al.*,
423 2007), could improve our knowledge on the different levels of endodormancy.

424 We also show that we can accurately predict the different bud stages using expression levels
425 for only seven marker genes (Fig. 7). This suggests that the definition of the different bud stages based
426 on physiological observation is consistent with transcriptomic profiles. However, we could detect
427 substantial discrepancies suggesting that the definition of the bud stages can be improved. Indeed, we
428 observe that samples harvested from buds during phases that we defined as organogenesis and
429 paradormancy cluster together in the PCA, but away from samples harvested during endodormancy.
430 Moreover, most of the genes highly expressed during paradormancy are also highly expressed during
431 organogenesis. This is further supported by the fact that paradormancy is a flower bud stage predicted
432 with less accuracy based on expression level of the seven marker genes. In details, paradormancy is
433 defined as a stage of growth inhibition originating from surrounding organs (Lang *et al.*, 1987)

434 therefore it is strongly dependant on the position of the buds within the tree and the branch. Our results
435 suggest that defining paradormancy for multiple cherry flower buds based on transcriptomic data is
436 difficult and even raise the question of whether paradormancy can be considered as a specific flower
437 bud stage. Alternatively, we propose that the pre-dormancy period should rather be defined as a
438 continuum between organogenesis, growth and/or growth cessation phases. Further physiological
439 observations, including flower primordia developmental context (Fadón *et al.*, 2015), could provide
440 crucial information to precisely link the transcriptomic environment to these bud stages.
441

442 **Highlight on main functions enriched during dormancy: organogenesis, response to cold, to ABA**
443 **and to the circadian clock**

444 We determined different functions and pathways enriched during flower bud organogenesis,
445 paradormancy, endodormancy and ecodormancy. We notably observe an enrichment for GO involved
446 in the response to abiotic and biotic responses, as well as an enrichment for targets of many TFs
447 involved in the response to environmental factors. In particular, our results suggest that *PavMYB14*,
448 which has a peak of expression in November just before the cold period starts, is repressing genes that
449 are expressed during ecodormancy. This is in agreement with the fact that *AtMYB14*, the *PavMYB14*
450 homolog in *Arabidopsis thaliana*, is involved in cold stress response regulation (Chen *et al.*, 2013).
451 Although these results were not confirmed in *Populus* (Howe *et al.*, 2015), two MYB DOMAIN
452 PROTEIN genes (*MYB4* and *MYB14*) were up-regulated during the induction phase of dormancy in
453 grapevine (Fennell *et al.*, 2015). Similarly, we identified an enrichment in target motifs for a
454 transcription factor belonging to the C-REPEAT/DRE BINDING FACTOR 2/DEHYDRATION
455 RESPONSE ELEMENT-BINDING PROTEIN (CBF/DREB) family in genes highly expressed during
456 endodormancy. These TFs have previously been implicated in cold acclimation and endodormancy in
457 several perennial species (Doğramaci *et al.*, 2010; Leida *et al.*, 2012). These results are in agreement
458 with the previous observation showing that genes responding to cold are differentially expressed
459 during dormancy in other tree species (Ueno *et al.*, 2013). Interestingly, we also identified an
460 enrichment in targets for four TFs involved in ABA-dependent signalling. First, *PavWRKY40* is mostly
461 expressed during organogenesis, and its expression profile is very similar to the one of its target genes.
462 Several studies have highlighted a role of *PavWRKY40* homolog in Arabidopsis in ABA signalling, in
463 relation with light transduction (Liu *et al.*, 2013; Geilen & Böhmer, 2015) and biotic stresses (Pandey
464 *et al.*, 2010). On the other hand, *PavABI5* and *PavABF2* are mainly expressed around the time of
465 dormancy release, like their target, and their homologs in Arabidopsis are involved in key ABA
466 processes, especially during seed dormancy (Lopez-Molina *et al.*, 2002). These results are further
467 confirmed by the enrichment of GO terms related to ABA pathway found in the genes highly expressed

468 during endodormancy. Our observations suggest that genes potentially linked to ABA signalling are
469 expressed either during organogenesis or during dormancy release. These results are supported by
470 previous reports where genes involved in ABA signalling are differentially expressed during dormancy
471 in other tree species (Ruttink *et al.*, 2007; Ueno *et al.*, 2013; Zhong *et al.*, 2013; Khalil-Ur-Rehman *et*
472 *al.*, 2017; Zhang *et al.*, 2018). It has also been shown that genes involved in other phytohormones
473 pathways, including auxin, ethylene, gibberellin and jasmonic acid, are differentially expressed
474 between bud stages in other perennial species (Zhong *et al.*, 2013; Khalil-Ur-Rehman *et al.*, 2017).
475 This is in agreement with our observation of an enrichment for GO terms for the response to jasmonic
476 acid, and of targets of TFs involved in the response to ethylene, in genes specifically expressed at
477 different flower bud stages.
478 In addition, we also identified an enrichment of targets for *PavRVE8* and *PavRVE1* among the genes
479 expressed around the time of dormancy release. These TFs are homologs of Arabidopsis MYB
480 transcription factors involved in the circadian clock. In particular, *AtRVE1* seems to integrate several
481 signalling pathways including cold acclimation and auxin (Rawat *et al.*, 2009; Meissner *et al.*, 2013;
482 Jiang *et al.*, 2016) while *AtRVE8* is involved in the regulation of circadian clock by modulating the
483 pattern of H3 acetylation (Farinas & Mas, 2011). Our findings that genes involved in the circadian
484 clock are expressed and potentially regulate genes at the time of dormancy release are in agreement
485 with previous work indicating a role of the circadian clock in dormancy in poplar (Ibáñez *et al.*, 2010).
486 To our knowledge, this is the first report on the transcriptional regulation of early stages of flower bud
487 development. We highlighted the upregulation of several pathways linked to organogenesis during the
488 summer months, including *PavMYB63 and PavMYB93*, expressed during early organogenesis, along
489 their targets, with potential roles in the secondary wall formation (Zhou *et al.*, 2009) and root
490 development (Gibbs *et al.*, 2014).
491
492 **Development of a diagnostic tool to define the flower bud dormancy stage using seven genes**
493 We find that sweet cherry flower bud stage can be accurately predicted with the expression of
494 just seven genes. It indicates that combining expression profiles of just seven genes is enough to
495 recapitulate all transcriptional states in our study. This is in agreement with previous work showing
496 that transcriptomic states can be accurately predicted using a relatively low number of markers (Biswas
497 *et al.*, 2017). Interestingly, when there are discrepancies between the predicted bud stages and the ones
498 defined by physiological observations, the model always predicts that stages happen earlier than the
499 actual observations. For example, the model predicts that dormancy breaking occurs instead of
500 endodormancy, or ecodormancy instead of dormancy breaking. This could suggest that transcriptional
501 changes happen before we can observe physiological changes. This is indeed consistent with the

502 indirect phenotyping method currently used, based on the observation of the response to growth-
503 inducible conditions after ten days. Using these seven genes to predict the flower bud stage would thus
504 potentially allow to identify these important transitions when they actually happen.
505 We also show that the expression level of these seven genes can be used to predict the flower bud stage
506 in other conditions by performing RT-qPCR. This independent experiment has also been done on two
507 consecutive years and shows that RT-qPCR for these seven marker genes as well as two control genes
508 are enough to predict the flower bud stage in cherry trees. It shows that performing a full transcriptomic
509 analysis is not necessary if the only aim is to define the dormancy stage of flower buds. This would
510 offer an alternative approach to methods currently used such as assessing the date of dormancy release
511 by using forcing conditions. In addition, this result sets the stage for the development of a fast and cost
512 effective diagnostic tool to molecularly define the flower bud state in cherry trees. Such diagnostic
513 tool would be very valuable for researchers working on cherry trees as well as for plant growers,
514 notably to define the best time for the application of dormancy breaking agents, whose efficiency
515 highly depends on the state of dormancy progression.
516

517 **Acknowledgments**

518 We thank the Fruit Experimental Unit of INRA (Bordeaux-France) for growing and managing the
519 trees, and Teresa Barreneche, Lydie Fouilhaux, Jacques Joly, Hélène Christman and Rémi Beauvieux
520 for the help during the harvest and for the pictures. Many thanks to Dr Varodom Charoensawan
521 (Mahidol University, Thailand) for providing scripts for mapping and gene expression count
522 extraction. The PhD of Noemie Vimont was supported by a CIFRE grant funded by the Roullier Group
523 (St Malo-France) and ANRT (France).
524

525 **Author contributions**

526 SC, BW, ED and PAW designed the original research. MA and JCY participated to the project design.
527 NV performed the RNA-seq and analysed the RNA-seq with CS and BW. MF performed the RT-
528 qPCR. JAC performed the TF and motifs enrichment analysis. MT developed the model. NV, SC and
529 BW wrote the article with the assistance of all the authors.
530

531 **Data availability**

532 RNA-seq data: Gene Expression Omnibus GSE130426

533 Graphical web interface DorPatterns: bwenden.shinyapps.io\DorPatterns

534

535 **References**

536 **Alexa A, Rahnenführer J**. **2018**. topGO: Enrichment Analysis for Gene Ontology. *R package*
537 *version 2.34.0*.

538 **Allona I, Ramos A, Ibañez C, Contreras A, Casado R, Aragoncillo C**. **2008**. Review. Molecular
539 control of winter dormancy establishment in trees. *Spanish Journal of Agricultural Research* **6**: 201–
540 210.

541 **Atkinson CJ, Brennan RM, Jones HG**. **2013**. Declining chilling and its impact on temperate
542 perennial crops. *Environmental and Experimental Botany* **91**: 48–62.

543 **Badeck FW, Bondeau A, Böttcher K, Doktor D, Lucht W, Schaber JJ, Sitch S, Bottcher K**.
544 **2004**. Responses of spring phenology to climate change. *New Phytologist* **162**: 295–309.

545 **Beauvieux R, Wenden B, Dirlewanger E**. **2018**. Bud Dormancy in Perennial Fruit Tree Species : A
546 Pivotal Role for Oxidative Cues. *Frontiers in Plant Science* **9**: 1–13.

547 **Benjamini Y, Hochberg J**. **1995**. Controlling the false discovery rate: a practical and powerful
548 approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**:
549 289–300.

550 **Bigler C, Bugmann H**. **2018**. Climate-induced shifts in leaf unfolding and frost risk of European
551 trees and shrubs. *Scientific Reports* **8**: 1–10.

552 **Biswas S, Kerner K, Teixeira PJPL, Dangl JL, Jojic V, Wigge PA**. **2017**. Tradict enables accurate
553 prediction of eukaryotic transcriptional states from 100 marker genes. *Nature Communications* **8**.

554 **Bolger AM, Lohse M, Usadel B**. **2014**. Trimmomatic: A flexible trimmer for Illumina sequence
555 data. *Bioinformatics* **30**: 2114–2120.

556 **Campoy J a., Ruiz D, Egea J**. **2011**. Dormancy in temperate fruit trees in a global warming context:
557 A review. *Scientia Horticulturae* **130**: 357–372.

558 **Causier B, Schwarz-Sommer Z, Davies B**. **2010**. Floral organ identity: 20 years of ABCs.
559 *Seminars in Cell and Developmental Biology* **21**: 73–79.

560 **Chao WS, Doğramacı M, Horvath DP, Anderson J V., Foley ME**. **2017**. Comparison of
561 phytohormone levels and transcript profiles during seasonal dormancy transitions in underground
562 adventitious buds of leafy spurge. *Plant Molecular Biology* **94**: 281–302.

563 **Chen Y, Chen Z, Kang J, Kang D, Gu H, Qin G**. **2013**. AtMYB14 Regulates Cold Tolerance in
564 Arabidopsis. *Plant Molecular Biology Reporter* **31**: 87–97.

565 **Cline MG, Deppong DO**. **1999**. The role of apical dominance in paradormancy of temperate woody
566 plants: A reappraisal. *Journal of Plant Physiology* **155**: 350–356.

567 **Considine MJ, Considine JA**. **2016**. On the language and physiology of dormancy and quiescence
568 in plants. *Journal of Experimental Botany* **67**: 3189–3203.

569 **Cooke JE, Eriksson ME, Junttila O**. **2012**. The dynamic nature of bud dormancy in trees:
570 environmental control and molecular mechanisms. *Plant Cell Environ* **35**: 1707–1728.
571 **Doğramaci M, Horvath DP, Chao WS, Foley ME, Christoffers MJ, Anderson J V. 2010**. Low
572 temperatures impact dormancy status, flowering competence, and transcript profiles in crown buds of
573 leafy spurge. *Plant Molecular Biology* **73**: 207–226.
574 **Doğramacı M, Horvath DP, Anderson J V. 2014**. Dehydration-induced endodormancy in crown
575 buds of leafy spurge highlights involvement of MAF3- and RVE1-like homologs, and hormone
576 signaling cross-talk. *Plant Molecular Biology* **86**: 409–424.
577 **Erez A**. **2000**. Bud Dormancy; Phenomenon, Problems and Solutions in the Tropics and Subtropics.
578 In: Temperate Fruit Crops in Warm Climates. 17–48.
579 **Fadón E, Herrero M, Rodrigo J**. **2015**. Flower development in sweet cherry framed in the BBCH
580 scale. *Scientia Horticulturae* **192**: 141–147.
581 **Falavigna V da S, Guitton B, Costes E, Andrés F**. **2019**. I Want to (Bud) Break Free: The
582 Potential Role of DAM and SVP-Like Genes in Regulating Dormancy Cycle in Temperate Fruit
583 Trees. *Frontiers in Plant Science* **9**: 1–17.
584 **Farinas B, Mas P**. **2011**. Histone acetylation and the circadian clock: A role for the MYB
585 transcription factor RVE8/LCL5. *Plant Signaling and Behavior* **6**: 541–543.
586 **Fennell AY, Schlauch KA, Gouthu S, Deluc LG, Khadka V, Sreekantan L, Grimplet J, Cramer**
587 **GR, Mathiason KL**. **2015**. Short day transcriptomic programming during induction of dormancy in
588 grapevine. *Frontiers in Plant Science* **6**: 834.
589 **Franklin KA, Lee SH, Patel D, Kumar SV, Spartz AK, Gu C, Ye S, Yu P, Breen G, Cohen JD,**
590 ***et al.* 2011**. PHYTOCHROME-INTERACTING FACTOR 4 (PIF4) regulates auxin biosynthesis at
591 high temperature. *Proceedings of the National Academy of Sciences* **108**: 20231–20235.
592 **Fu YH, Zhao H, Piao S, Peaucelle M, Peng S, Zhou G, Ciais P, Huang M, Menzel A, Peñuelas**
593 **J, *et al.* 2015**. Declining global warming effects on the phenology of spring leaf unfolding. *Nature*
594 **526**: 104–107.
595 **Geilen K, Böhmer M**. **2015**. Dynamic subnuclear relocalization of WRKY40, a potential new
596 mechanism of ABA-dependent transcription factor regulation. *Plant signaling & behavior* **10**:
597 e1106659.
598 **Gibbs DJ, Voß U, Harding SA, Fannon J, Moody LA, Yamada E, Swarup K, Nibau C, Bassel**
599 **GW, Choudhary A, *et al.* 2014**. AtMYB93 is a novel negative regulator of lateral root development
600 in Arabidopsis. *New Phytologist* **203**: 1194–1207.
601 **Grant CE, Bailey TL, Noble WS**. **2011**. FIMO: Scanning for occurrences of a given motif.
602 *Bioinformatics* **27**: 1017–1018.

603 **Heide OM**. **2008**. Interaction of photoperiod and temperature in the control of growth and dormancy
604 of Prunus species. *Scientia Horticulturae* **115**: 309–314.
605 **Heide OM, Prestrud AK**. **2005**. Low temperature, but not photoperiod, controls growth cessation
606 and dormancy induction and release in apple and pear. *Tree Physiol* **25**: 109–114.
607 **Horvath DP**. **2009**. Common mechanisms regulate flowering and dormancy. *Plant Science* **177**:
608 523–531.
609 **Howe GT, Horvath DP, Dharmawardhana P, Priest HD, Mockler TC, Strauss SH**. **2015**.
610 Extensive Transcriptome Changes During Natural Onset and Release of Vegetative Bud Dormancy
611 in Populus. *Frontiers in plant science* **6**.
612 **Ibáñez C, Kozarewa I, Johansson M, Ogren E, Rohde A, Eriksson ME**. **2010**. Circadian clock
613 components regulate entry and affect exit of seasonal dormancy as well as winter hardiness in
614 Populus trees. *Plant physiology* **153**: 1823–33.
615 **Jiang Z, Xu G, Jing Y, Tang W, Lin R**. **2016**. Phytochrome B and REVEILLE1/2-mediated
616 signalling controls seed dormancy and germination in Arabidopsis. *Nature Communications* **7**: 1–10.
617 **Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G**. **2017**. PlantTFDB 4.0: Toward a
618 central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **45**:
619 D1040–D1045.
620 **Khalil-Ur-Rehman M, Sun L, Li CX, Faheem M, Wang W, Tao JM**. **2017**. Comparative RNA-
621 seq based transcriptomic analysis of bud dormancy in grape. *BMC Plant Biology* **17**: 1–11.
622 **Kiełbowicz-Matuk A**. **2012**. Involvement of plant C2H2-type zinc finger transcription factors in
623 stress responses. *Plant Science* **185**–**186**: 78–85.
624 **Koornneef M, Léon-Kloosterziel KM, Schwartz SH, Zeevaart JAD**. **1998**. The genetic and
625 molecular dissection of abscisic acid biosynthesis and signal transduction in Arabidopsis. *Plant*
626 *Physiology and Biochemistry* **36**: 83–89.
627 **Kumar G, Rattan UK, Singh AK**. **2016**. Chilling-mediated DNA methylation changes during
628 dormancy and its release reveal the importance of epigenetic regulation during winter dormancy in
629 Apple (Malus x domestica Borkh.). *PLoS ONE* **11**: 1–25.
630 **Lang G, Early J, Martin G, Darnell R**. **1987**. Endo-, para-, and ecodormancy: physiological
631 terminology and classification for dormancy research. *Hort Science* **22**: 371–377.
632 **Lee S-J, Kang J-Y, Park H-J, Kim MD, Bae MS, Choi H, Kim SY**. **2010**. DREB2C Interacts with
633 ABF2, a bZIP Protein Regulating Abscisic Acid-Responsive Gene Expression, and Its
634 Overexpression Affects Abscisic Acid Sensitivity. *Plant Physiology* **153**: 716–727.
635 **Leida C, Conesa A, Llácer G, Badenes ML, Ríos G**. **2012**. Histone modifications and expression
636 of DAM6 gene in peach are modulated during bud dormancy release in a cultivar-dependent manner.

637 *New phytologist* **193**: 67–80.
638 **Lesur I, Le Provost G, Bento P, Da Silva C, Leplé JC, Murat F, Ueno S, Bartholomé J, Lalanne**
639 **C, Ehrenmann F, *et al.* 2015**. The oak gene expression atlas: Insights into Fagaceae genome
640 evolution and the discovery of genes regulated during bud dormancy release. *BMC Genomics* **16**:
641 112.
642 **Liu Q, Wang Z, Xu X, Zhang H, Li C**. **2015**. Genome-wide analysis of C2H2 zinc-finger family
643 transcription factors and their responses to abiotic stresses in poplar (Populus trichocarpa). *PLoS*
644 *ONE* **10**: 1–25.
645 **Liu R, Xu Y-H, Jiang S-C, Lu K, Lu Y-F, Feng X-J, Wu Z, Liang S, Yu Y-T, Wang X-F, *et al.***
646 **2013**. Light-harvesting chlorophyll a/b-binding proteins, positively involved in abscisic acid
647 signalling, require a transcription repressor, WRKY40, to balance their function. *Journal of*
648 *Experimental Botany* **64**: 5443–5456.
649 **Lloret A, Badenes ML, Ríos G**. **2018**. Modulation of Dormancy and Growth Responses in
650 Reproductive Buds of Temperate Trees. *Frontiers in Plant Science* **9**: 1–12.
651 **Lopez-Molina L, Mongrand S, McLachlin DT, Chait BT, Chua NH**. **2002**. ABI5 acts
652 downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant Journal*
653 **32**: 317–328.
654 **Love MI, Huber W, Anders S**. **2014**. Moderated estimation of fold change and dispersion for RNA-
655 seq data with DESeq2. *Genome Biology* **15**: 1–21.
656 **Maurya JP, Triozzi PM, Bhalerao RP, Perales M**. **2018**. Environmentally Sensitive Molecular
657 Switches Drive Poplar Phenology. *Frontiers in Plant Science* **9**: 1–8.
658 **Meier U**. **2001**. *Growth stages of mono-and dicotyledonous plants BBCH Monograph*.
659 **Meissner M, Orsini E, Ruschhaupt M, Melchinger AE, Hincha DK, Heyer AG**. **2013**. Mapping
660 quantitative trait loci for freezing tolerance in a recombinant inbred line population of Arabidopsis
661 thaliana accessions Tenela and C24 reveals REVEILLE1 as negative regulator of cold acclimation.
662 *Plant, Cell and Environment* **36**: 1256–1267.
663 **Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kübler K, Bissolli P,**
664 **Braslavská O, Briede A, *et al.* 2006**. European phenological response to climate change matches the
665 warming pattern. *Global Change Biology* **12**: 1969–1976.
666 **Min Z, Zhao X, Li R, Yang B, Liu M, Fang Y**. **2017**. Comparative transcriptome analysis provides
667 insight into differentially expressed genes related to bud dormancy in grapevine (Vitis vinifera).
668 *Scientia Horticulturae* **225**: 213–220.
669 **Olsen JE**. **2010**. Light and temperature sensing and signaling in induction of bud dormancy in
670 woody plants. *Plant Molecular Biology* **73**: 37–47.

671 **Ophir R, Pang X, Halaly T, Venkateswari J, Lavee S, Galbraith D, Or E**. **2009**. Gene-expression
672 profiling of grape bud response to two alternative dormancy-release stimuli expose possible links
673 between impaired mitochondrial activity, hypoxia, ethylene-ABA interplay and cell enlargement.
674 *Plant Molecular Biology* **71**: 403–423.
675 **Pandey SP, Roccaro M, Schön M, Logemann E, Somssich IE**. **2010**. Transcriptional
676 reprogramming regulated by WRKY18 and WRKY40 facilitates powdery mildew infection of
677 Arabidopsis. *Plant Journal* **64**: 912–923.
678 **Paul A, Jha A, Bhardwaj S, Singh S, Shankar R, Kumar S**. **2014**. RNA-seq-mediated
679 transcriptome analysis of actively growing and winter dormant shoots identifies non-deciduous habit
680 of evergreen tree tea during winters. *Scientific Reports* **4**: 1–9.
681 **Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M,**
682 **Prettenhofer P, Weiss R, Dubourg V, *et al.* 2011**. Scikit-learn: Machine Learning in Python Gaël
683 Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX,
684 GRAMFORT ET AL. Matthieu Perrot. *Journal of Machine Learning Research* **12**: 2825–2830.
685 **Penfield S, Josse E-M, Kannangara R, Gilday AD, Halliday KJ, Graham IA**. **2005**. Cold and
686 light control seed germination through the bHLH transcription factor SPATULA. *Current Biology*
687 **15**: 1998–2006.
688 **Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, Andersson CR, Zhao Y, Ljung K,**
689 **Harmer SL**. **2009**. REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and
690 auxin pathways. *Proceedings of the National Academy of Sciences* **106**: 16883–16888.
691 **Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W,**
692 **Rohde A**. **2007**. A molecular timetable for apical bud formation and dormancy induction in poplar.
693 *Plant Cell* **19**: 2370–2390.
694 **Singh RK, Svystun T, AlDahmash B, Jönsson AM, Bhalerao RP**. **2016**. Photoperiod- and
695 temperature-mediated control of growth cessation and dormancy in trees: A molecular perspective.
696 *New Phytologist*.
697 **Snyder RL, de Melo-abreu JP**. **2005**. *Frost Protection : fundamentals , practice and economics*.
698 Rome.
699 **Takemura Y, Kuroki K, Shida Y, Araki S, Takeuchi Y, Tanaka K, Ishige T, Yajima S, Tamura**
700 **F**. **2015**. Comparative transcriptome analysis of the less-dormant taiwanese pear and the dormant
701 Japanese pear during winter season. *PLoS ONE* **10**.
702 **Trapnell C, Pachter L, Salzberg SL**. **2009**. TopHat: Discovering splice junctions with RNA-Seq.
703 *Bioinformatics* **25**: 1105–1111.
704 **Ueno S, Klopp C, Leplé JC, Derory J, Noirot C, Léger V, Prince E, Kremer A, Plomion C, Le**

705 **Provost G**. **2013**. Transcriptional profiling of bud dormancy induction and release in oak by next-
706 generation sequencing. *BMC Genomics* **14**: 236.
707 **Verde I, Jenkins J, Dondini L, Micali S, Pagliarani G, Vendramin E, Paris R, Aramini V, Gaza**
708 **L, Rossini L, *et al.* 2017**. The Peach v2.0 release: high-resolution linkage mapping and deep
709 resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics* **18**: 225.
710 **Vitasse Y, Lenz A, Körner C**. **2014**. The interaction between freezing tolerance and phenology in
711 temperate deciduous trees. *Frontiers in Plant Science* **5**: 541.
712 **Wagner D**. **2003**. Chromatin regulation of plant development. *Current Opinion in Plant Biology* **6**:
713 20–28.
714 **Wenden B, Campoy J, Jensen M, López-Ortega G**. **2017**. Climatic Limiting Factors:
715 Temperature. In: Quero-García J, Iezzoni A, Pulawska J, Lang G, eds. Cherries: Botany, Production
716 and Uses. CABI Publishing, 166–188.
717 **Zhang Z, Zhuo X, Zhao K, Zheng T, Han Y, Yuan C, Zhang Q**. **2018**. Transcriptome Profiles
718 Reveal the Crucial Roles of Hormone and Sugar in the Bud Dormancy of Prunus mume. *Scientific*
719 *Reports* **8**: 1–15.
720 **Zhong W, Gao Z, Zhuang W, Shi T, Zhang Z, Ni Z**. **2013**. Genome-wide expression profiles of
721 seasonal bud dormancy at four critical stages in Japanese apricot. *Plant molecular biology* **83**: 247–
722 64.
723 **Zhou J, Lee C, Zhong R, Ye Z-H**. **2009**. MYB58 and MYB63 Are Transcriptional Activators of the
724 Lignin Biosynthetic Pathway during Secondary Cell Wall Formation in Arabidopsis. *the Plant Cell*
725 *Online* **21**: 248–266.
726 **Zhu Y, Li Y, Xin D, Chen W, Shao X, Wang Y, Guo W**. **2015**. RNA-Seq-based transcriptome
727 analysis of dormant flower buds of Chinese cherry (Prunus pseudocerasus). *Gene* **555**: 362–376.
728
729
730 **Supporting information**
731 **Fig. S1** Field temperature during the sampling season
732 **Fig. S2** Separation of samples by dormancy stage using read counts for all genes
733 **Fig. S3** Enrichments in gene ontology terms in the ten clusters
734 **Fig. S4** Expression patterns for the transcription factors and their targets
735 **Fig. S5** Separation of samples by dormancy stage and cultivar using all genes
736 **Fig. S6** Clusters of expression patterns for differentially expressed genes in the sweet cherry cultivars
737 'Regina', 'Cristobalina' and 'Garnet'
738 **Fig. S7** Separation of samples by dormancy stage and cultivar using the seven marker genes

739 **Fig. S8** Multinomial logistic model details

740

741 **Table S1** Description of the flower bud samples used for RNA-seq and qRT-PCR

742 **Table S2** RNA-seq mapped reads and gene count information

743 **Table S3** 'Garnet' differentially expressed genes and their assigned clusters.

744 **Table S4** Transcription factors with motif enrichment in the clusters.

745