bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

1 # Characterization of genes and alleles involved in the control of flowering
2 # time in grapevine

3

4 Nadia Kamal[1,#a], Iris Ochßner[2], Anna Schwandner[2], Prisca Viehöver[1], Ludger Hausmann[2], Reinhard
5 Töpfer[2], Bernd Weisshaar[1*] and Daniela Holtgräwe[1]

6

7 [1] Bielefeld University, Faculty of Biology & Center for Biotechnology, Bielefeld, Germany
8 [2] Julius Kühn-Institute (JKI), Institute for Grapevine Breeding Geilweilerhof, Siebeldingen, Germany
9 [#a] Helmholtz Center Munich, Plant Genome and Systems Biology, Munich, Germany

10

11 Email addresses:
12 Nadia Kamal (NK); nadia.kamal@helmholtz-muenchen.de
13 Iris Ochßner (IO); i.ochssner@gmx.de
14 Anna Schwandner (AS); Anna.schwandner@julius-kuehn.de
15 Prisca Viehöver (PV); viehoeve@cebitec.uni-bielefeld.de
16 Ludger Hausmann (LH); ludger.hausmann@julius-kuehn.de
17 Reinhard Töpfer (RT); reinhard.toepfer@julius-kuehn.de
18 Bernd Weisshaar (BW); bernd.weisshaar@uni-bielefeld.de
19 Daniela Holtgräwe (DH); dholtgra@cebitec.uni-bielefeld.de
20 * Corresponding Author
21 E-mail: dholtgra@cebitec.uni-bielefeld.de (DH)

22

23 ORCIDS
24 Ludger Hausmann; ORCID: 0000-0002-4046-9626
25 Reinhard Töpfer; ORCID: 0000-0003-1569-2495
26 Bernd Weisshaar; ORCID: 0000-0002-7635-3473
27 Daniela Holtgräwe; ORCID: 0000-0002-1062-4576
28 Nadia Kamal; ORCID: 0000-0002-3258-4130

29

30

31

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

32

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

33 # Abstract

34 Grapevine (*Vitis vinifera*) is one of the most important perennial crop plants in worldwide.
35 Understanding of developmental processes like flowering, which impact quality and quantity of yield
36 in this species is therefore of high interest. This gets even more important when considering some of
37 the expected consequences of climate change. Earlier bud burst and flowering, for example, may
38 result in yield loss due to spring frost. Berry ripening under higher temperatures will impact wine
39 quality. Knowledge of interactions between a genotype or allele combination and the environment
40 can be used for the breeding of genotypes that are better adapted to new climatic conditions. To this
41 end, we have generated a list of more than 500 candidate genes that may play a role in the timing of
42 flowering. The grapevine genome was exploited for flowering time control gene homologs on the
43 basis of functional data from model organisms like *A. thaliana.* In a previous study, a mapping
44 population derived from early flowering GF.GA-47-42 and late flowering 'Villard Blanc' was
45 analyzed for flowering time QTLs. In a second step we have now established a workflow combining
46 amplicon sequencing and bioinformatics to follow alleles of selected candidate genes in the F_1
47 individuals and the parental genotypes. Allele combinations of these genes in individuals of the
48 mapping population were correlated with early or late flowering phenotypes. Specific allele
49 combinations of flowering time candidate genes within and outside of the QTL regions for flowering
50 time on chromosome 1, 4, 14, 17, and 18 were found to be associated with an early flowering
51 phenotype. In addition, expression of many of the flowering candidate genes was analyzed over
52 consecutive stages of bud and inflorescence development indicating functional roles of these genes in
53 the flowering control network.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

54 # Introduction

55 The reproductive developmental cycle of grapevine spans two years (S1 Figure). Grapevine plants
56 need intense light and high temperatures to initiate inflorescences during spring, which develop and
57 flower during the subsequent summer [1]. The ongoing tendency to higher temperatures in spring
58 due to global warming causes earlier bud burst and flowering [2]. As a consequence, late spring
59 frost is an increasing risk to viticulture, which may cause significant crop loss [3]. Together with
60 flowering the onset of ripening is shifted towards earlier dates [4,5] and the ripening process occurs
61 under warmer conditions. This influences berry composition [6], affects wine quality and promotes
62 e.g. fungi infection. Grapevine breeding programs aim to keep the production of high quality grapes
63 in a changing environment consistent. Making use of late flowering genotypes may be one approach
64 to compensate for earlier ripening. Understanding the flowering process in grapevine and
65 determining factors that lead to early or late flowering may help to control variation in berry
66 production [7].

67 Detailed knowledge of pathways controlling flowering is available in crop species and the woody
68 plant poplar, but especially the model species *A. thaliana* and rice [8,9]. With the availability of a
69 *Vitis* reference genome sequence [10-14], gene homologs to *A. thaliana* floral development pathway
70 genes or genes involved in photoperiod or vernalization responses could be identified in the
71 grapevine genome. Most of these are flowering signal integrators, floral meristem identity genes,
72 and flower organ identity genes, such as MADS box genes, like *VvMADS8* that promotes early
73 flowering and the *VvFT/TFL1* gene family [15-17]. The expression of *VvFT* - the ortholog of the *A.*
74 *thaliana FLOWERING LOCUS T* - is associated with seasonal flowering induction in latent buds
75 and the development of inflorescences, flowers, and fruits [18]. The expression of the *LEAFY*
76 ortholog *VvFL* is correlated with inflorescence and flower development [15]. *VvFUL-L* and *VvAP1* -
77 homologs of the *A. thaliana* genes *FUL* and *AP1* - are suggested to act on the specification of
78 flower organ identity as their expression appears in early developmental stages of lateral meristems
79 and is maintained in both inflorescence and tendril primordia [16,19].

80 Due to the high heterozygosity and severe inbreeding depression, the first filial generation (F_1) is
81 used for QTL (quantitative trait loci) mapping in *V. vinifera.* This is different to other crop or model
82 species (and is called a double pseudo test cross approach; [20,21]). Several QTL for the timing of
83 developmental stages such as flowering time have been identified [2,22,23]. One locus contributing
84 to flowering time control (FTC) was reported in 2006 [24]. Six QTL on different chromosomes
85 (chr) in the mapping population GF.GA-47-42 x ‘Villard Blanc’ were described in [23]. The

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

86 detected QTL are localized on chr 1, 4, 8, 14, 17, 18 and 19. Three of them (chr 1, 14 and 17) were
87 also found in another mapping population derived from the genotypes V3125 and 'Börner' [23].
88 MADS-box genes with a proposed impact on flowering time such as *VvFL, VvFUL-L* and *VvAP1*
89 were annotated within FTC QTL regions in *Vitis*. Further, examples of flowering time gene
90 homologues in such QTL regions include *CONSTANS-like* genes on chr 1, 4 and 14 and the MADS-
91 box genes, *VvFLC1* und *VvFLC2 (Vitis vinifera FLOWERING LOCUS C 1 & 2),* which are highly
92 expressed in buds [25].

93 The observation that either very early or very late flowering seems to be inherited by specific
94 combinations of alleles at several loci, while all mixed combinations lead to an intermediate
95 flowering type indicates an additive effect. The data further suggest a dominant effect for early
96 flowering, with the responsible alleles being inherited from either 'Bacchus' or 'Seyval', the parents
97 of the breeding line GF.GA-47-42 [23]. In order to link certain alleles of the sequenced genes to the
98 flowering time phenotype, the two allele sequences of a given gene in a heterozygous diploid plant
99 have to be determined (allele phasing).

100 Short read sequencing technologies still suffer from producing ambiguous haplotype phase
101 sequences. Determining the haplotype phase of an individual is computationally challenging and
102 experimentally expensive; but haplotype phase information is crucial in various analyses, such as
103 genetic association studies, the reconstruction of phylogenies and pedigrees, genomic imputation,
104 linkage disequilibrium, and SNP tagging [26,27][28,29]. In diploid organisms like grapevine,
105 generally both alleles of a given gene are expressed. Different alleles can show different expression
106 patterns, which can consequently result in varying manifestations of traits. The determination of
107 these alleles is an important step in the dissection of corresponding traits. Among other approaches,
108 haplotypic information can be obtained from DNA sequence fragments to reconstruct the two
109 haplotypes of a diploid individual. A sequence fragment that covers at least two variant sites in a
110 genome can link those variants together and thus phase them. When fragments are long enough to
111 encompass multiple variant sites and the sequencing coverage is sufficiently high to provide
112 overlaps between fragments, fragments can be assembled to reconstruct longer haplotypes [30].

113 For haplotype or allele phasing a variant discovery process is necessary beforehand. The two
114 mainly used methods are based on Shotgun Genome Assembly (SGA) or on amplicon sequencing.
115 SGA generates phasing information without knowledge of the surrounding sequence, the library
116 coverage needs to be high and it is computationally very challenging to distinguish paralogous
117 repeats from polymorphism but it does not require sequence information for the loci. Amplicon
118 sequencing, which includes the amplification of a genomic region by PCR, requires sequence

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

119 information of the target locus for primer design and can be done very effectively. However, it is
120 not practical for large-scale projects [31].

121 In this work, we used a F_1 population of *V. vinifera*, with the aim to associate allele sequences of
122 several FTC candidate genes with the phenotype of flowering time in order to identify alleles
123 influencing and controlling this trait using amplicon sequencing. Gene expression was analyzed in
124 different time courses of bud and flower development in order to further investigate and confirm the
125 role of FTC candidate genes.

126 Materials and Methods

127 Plant material

128 The mapping population GF.GA-47-42 x 'Villard Blanc' was crossed in 1989 using the breeding
129 line GF.GA-47-42 ('Calardis Musque'; 'Bacchus Weiss' x 'Seyval') and the cultivar 'Villard
130 Blanc' (Seibel 6468 x 'Subereux'). The 151 F_1 individuals were planted in the vineyards at the
131 Institute for Grapevine Breeding Geilweilerhof in Siebeldingen (49°13'05.0"N 8°02'45.0"E) in
132 Southwestern Germany (www.julius-kuehn.de/en/grapevine-breeding) in 1996. The offspring
133 shows notable segregation for the trait "flowering time" as the maternal breeding line GF.GA-47-42
134 and its parents are early flowering while the paternal line 'Villard Blanc' as well as its parents
135 flower rather late. QTL analysis for flowering time was carried out using a SSR marker-based
136 genetic map of the biparental population [32].

137 Phenotyping of the mapping population GF.GA-47-42 x 'Villard Blanc' was performed for
138 flowering time (full bloom) in nine years (1999, 2009 - 2016) as described in [23] (Table 1, S1
139 Table). For determination of the median of flowering time for each individual, the days of the
140 flowering period of each year were numbered whereas the first day of the flowering period was
141 numbered with one, the second day with two, etc. These numbers were then divided by the length of
142 the flowering period. The resulting values were used to calculate the median. Values for global
143 radiation and accumulated temperature from November 1st of the previous year until the day of full
144 bloom were obtained from the DLR (www.wetter.rlp.de) and refer to the location of the vineyard at
145 Siebeldingen, Germany. For gene expression analysis of FTC target genes, leaves, buds, and
146 inflorescences from early flowering GF.GA-47-42 were collected at several consecutive time points
147 starting from latent winter buds until inflorescences shortly before full bloom within the
148 developmental cycle that was completed over the two consecutive years 2012 and 2013. Moreover,
149 in 2013, sampling of buds on consecutive time points before dormancy in winter was continued.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

150 The development of the plants was described using BBCH codes [33]. Plant tissue from four
151 different GF.GA-47-42 plants was harvested into liquid nitrogen. We decided in favor of single
152 samples but many time points to detect trends in expression levels. Table 2 shows an overview of
153 the collected samples.

154 Table 1: Dates of flowering periods of the mapping population GF.GA-47-42 x 'Villard Blanc' and
155 the amount of global radiation at the location of the vineyards (Geilweilerhof) if available.

Year	Start of flowering period (days after January 1st)	End of flowering period (days after January 1st)	Length of flowering period (days)	Global radiation at beginning of flowering period (KWh/ m²)	Global radiation at end of flowering period (KWh/ m²)
1999	165	178	14	/	/
2009	156	170	15	/	/
2010	151	180	19	/	/
2011	147	157	11	531	579
2012	153	169	17	511	596
2013	168	183	16	516	597
2014	150	161	12	518	567
2015	156	167	12	536	595
2016	168	177	10	502	548

156

157 Table 2: Samples collected from grapevine genotype GF.GA-47-42 for the analysis of trends in
158 gene expression levels. Listed are developmental stage and the corresponding BBCH code.

Date of sample collection	**Developmental stage**	**BBCH code**
Developmental cycle 2012/2013:		
December 20th, 2012	dormant bud	BBCH 0
March 8th, 2013	dormant bud	BBCH 0

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

March 22nd, 2013	swelling bud	BBCH 0-5
April 12th, 2013	swelling bud	BBCH 5-9
April 26th, 2013	swelling bud/ first leaves	BBCH 11
May 3rd, 2013	buds/ first leaves	BBCH 11-13
June 7th, 2013	inflorescences & leaves	BBCH 53
June 14th, 2013	inflorescences & leaves	BBCH 55
June 17th, 2013	inflorescences	BBCH 57
Developmental cycle 2013/2014:		
July 22nd, 2013	buds & leaves	/
August 2nd, 2013	buds	/
August 8th, 2013	buds & leaves	/
August 16th, 2013	buds	/
August 22nd, 2013	buds & leaves	/
September 5th, 2013	buds	/
September 19th, 2013	leaves	/

159

160 ## FTC candidate gene prediction

161 For the identification and characterization of putative flowering time control (FTC) genes,
162 functional data from well studied model species was used to exploit the grapevine genome for
163 homologous genes. Using BLAST (e-value cut off below 1e-25) [34] protein sequences of
164 candidate genes from *A. thaliana* and other model species were compared against the *Vitis* protein
165 sequences (PN40024-12xv0, Genoscope gene prediction 12X.v0
166 (www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/) and the CRIBI gene prediction 12X.v2
167 [12]). Results were manually checked for additional evidence from the literature.

168 For functional annotation of FTC candidate genes, the method of reciprocal best hits (RBH) [35]
169 was applied. A RBH pair consists of two sequences from different sets of sequences, each
170 displaying the highest genome wide score in the other data set. Genomic sequences of FTC genes
171 were compared against protein sequences of *V. vinifera* and *A. thaliana* with blastx. If a gene

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

displayed several transcripts, the longest sequence was used. Using tBlastn the hit showing the highest score was compared back against *V. vinifera* coding genes. When the original query was found to have the highest score, the resulting RBH pair was considered.

To establish unique genes, we used the *Vv* (*Vitis vinifera*) prefix followed, for almost all genes, by the gene name deduced from the *Arabidopsis* annotation. In many cases the *Vitis* genome holds several putative homologs for known FTC genes from model crops, leading to low number of RBHs between *Vitis* and *Arabidopsis* genes. In order to distinguish these *Vitis* genes, the one with the highest BLAST score to the query gene got the name extension “a”, the second best the “b”.

Amplimer design

Genes for targeted allele phasing (target genes) through amplicon sequencing were selected out of the identified FTC candidate genes. The cDNA sequences of target genes were used as query in a BLAST against the grapevine reference sequence PN40024-12xv0. Genomic DNA sequences were extracted in addition to 1,000 bp from the 5'- and 3'-UTR regions. Primers were designed for overlapping amplimers of up to 8 kb using the tool Primer3 [36].

DNA isolation and amplicon generation

Extraction of genomic DNA was performed from young leaf tissue. The leaf material was grounded under liquid nitrogen and subsequently used for DNA isolation with the DNeasy® Plant Maxi Kit (Qiagen, Hilden, Germany) according to manufacturer’s protocols. The purified DNA was quality checked via gel electrophoresis and quantified using a NanoDrop spectrophotometer (Peqlab, Erlangen, Germany). Amplicons were amplified by long range PCR (98 °C 30 sec, 15 cycles of 10 sec 98 °C, 30 sec 72 °C – 57 °C, 5 min 72 °C, 25 cycles 10 sec 98 °C, 30 sec 58 °C, 5 min 72 °C and finally 2 min 72 °C).

Target gene sequences were amplified from 37 individuals of the mapping population GF.GA-47-42 x ‘Villard Blanc’ including the parental lines and 35 F_1 individuals with early, intermediate and, late flowering time phenotypes (S2 Table).

Library preparation and amplicon sequencing

Amplicon sequencing was carried out on a MiSeq (Illumina, San Diego, USA) in seven runs. All amplicons belonging to a respective individual were pooled in equimolar amounts, fragmented by sonification using a Bioruptor (Diagenode, Denville, USA) and subsequently used for library preparation. The libraries were prepared as recommended by Illumina (TruSeq DNA Sample

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

Preparation v2 Guide). Adaptor-ligated fragments were size selected on a two percent low melt agarose gel to an average insert size of 500 bp. Fragments that carry adaptors on both ends were enriched by PCR. Final libraries were quantified using PicoGreen (Quant-iT, Fisher Scientific, Schwerte, Germany) on a Fluostar platereader (BMG labtech, Ortenberg, Germany) and quality checked by HS-Chips on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Up to 20 libraries were pooled and sequenced on an Illumina MiSeq platform with 2 x 250 bp read length using the Illumina MiSeq v2 reagents. After sequencing, basecalling and demultiplexing and FASTQ file generation was performed using a casava-based in house script.

Read processing and mapping

Adapter trimming of raw reads and quality filtering of reads with a window of four consecutive bases that exhibited a quality value below 30 was performed using Trimmomatic [37]. Bases at the heads and tails of the reads with quality values below 30 were cropped using Trimmomatic. Before and after trimming the tool FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) was used to check the quality of the reads. Between 11.5 and 35.6% (20.2% on average [standard deviation (SD): 5.5%]) of reads were dropped through trimming. Trimmed reads were mapped to the grapevine reference sequence PN40024 12x.v2 [14] using the BWA-MEM algorithm which is suitable for long reads with default parameters [38]. Mapping was performed for each individual separately. Instead of the entire reference sequence the target gene sequences only were chosen for mapping in order to prevent false positive mapping results. The SAM format files were converted to BAM format files and sorted using SAMtools [39]. Readgroups were added and duplicated reads removed using Picard Tools (https://broadinstitute.github.io/picard/). Besides PCR duplicates unpaired reads were removed from the mapping files. About 15% of amplicons failed to be amplified or sequencing depth was below 20.

Allele phasing of target genes

In order to separate the two alleles of the sequenced target genes (phasing), a workflow using the Genome Analysis Toolkit (GATK) [40] was established (Fig. 1). After read alignment, the quality of the alignments was improved in two ways. Firstly, local realignments around InDels were performed using InDelRealigner of GATK [40] to reduce the number of misalignments. Occasionally, the presence of insertions or deletions in individuals with respect to the reference genome sequence leads to misalignments of reads to the reference, especially when InDels are covered at the start or end of a read. Such misalignments lead to many false positive SNPs. Secondly, base quality scores of reads in the aligned mapping files were recalibrated using

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

234 BaseRecalibrator of GATK in order to correct for variation in quality with machine cycle and
235 sequence context. Thus, more accurate and more widely dispersed quality scores are provided.

236 Fig. 1: Workflow using the Genome Analysis Toolkit (GATK), which uses the high-coverage
237 genotype sequence variation information and the family relationship for phasing.

238 Using the HaplotypeCaller of GATK variants were called for each individual separately. The ploidy
239 parameter was set to 12 for variant calling. It was performed in gVCF mode for F_1 individuals and
240 the parental lines of the population GF.GA-47-42 x ‘Villard Blanc’. Cases of allele dropout were
241 identified, in which the missing allele leads to genotyping errors. Since we were working with an F_1
242 population and by applying Mendelian constraints it was possible to determine which allele was
243 missing within the population GF.GA-47-42 x ‘Villard Blanc’, but its sequence remained unknown.
244 After variant calling, resulting variant files from individuals of the population were merged using
245 GATK's GenotypeGVCFs in order to apply further downstream steps on all samples together. At
246 each position of the input gVCFs, this tool combines all spanning records and outputs them to a
247 new variant file. Raw variants were hard-filtered according to GATK's "Best Practices"
248 recommendations [41,42]. In addition, variants with read coverage depth and genotype quality
249 below 20 were filtered out. For the determination of allele-specific sequences initially physical
250 phasing was performed using HapCUT [30]. Fragments were defined from the sequenced reads.
251 Haplotype-informative reads that cover at least two heterozygous variants were extracted from the
252 aligned file using the tool extractHairs from HapCut and used for the assembly of haplotypes. The
253 information of polymorphic sites was passed to HapCUT through a variant file. A maximum
254 number of 600 iterations were used to run HapCut and the reference sequence was provided in order
255 to extract reads covering both SNPs and InDels. Using various python scripts, intervals in which
256 phasing could be performed in individuals of the population GF.GA-47-42 x ‘Villard Blanc’
257 including the parents and F_1 individuals were determined and homozygous alternative variants were
258 added to the variant files. Using GATKs FastaAlternateReferenceMaker FASTA-format files with
259 alternate sequences were created for each individual within the regions in which allele phasing
260 could be performed.

261 A nomenclature system was created for the alleles of genes within the population GF.GA-47-42 x
262 ‘Villard Blanc’ (S3 Table). The system distinguishes between fourteen different cases, where four,
263 three, or two different allele sequences can be present at a locus or all sequences can be identical.
264 Moreover, it distinguishes between various combinations of two or three different sequences. E, as
265 in E1, E2 and E0, refers to “early” and originates from early flowering GF.GA-47-42, while L, as in
266 L1, L2 and L0 refers to “late” and originates from late flowering ‘Villard Blanc’. N means that both

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

267 GF.GA-47-42 and 'Villard Blanc' share one or more alleles. N1 means that E1 and L1 are alike,
268 while N2 means that E2 and L2 are alike. N means that either L2 and E1 or E2 and L1 are alike. Na
269 means that E1, E2, and L1 are alike. Nb means that E1, E2, and L2 are alike. Nc means that E1, L1,
270 and L2 are alike. Nd means that E2, L1, and L2 are alike. Descriptions for allele combinations that
271 distinguish between which of the two alleles of one parental line is alike the two alleles of the other
272 line (as in NaNa x NaL2) was implemented in order to be able to track patterns of allele
273 combinations throughout QTL regions and closely neighboring genes.

274 Correlation analysis

275 To test for the correlation of an allele and the flowering time phenotype, a Wilcoxon Rank-Sum test
276 was carried out between a dichotomous variable (the presence or absence of an allele) and a
277 continuous variable (flowering time). The null hypothesis assumed that the median of flowering
278 time between groups of individuals carrying a certain allele or not is equal. When p-values below
279 5% were found, the null hypothesis was rejected and an association between an allele and the
280 flowering time phenotype was found to exist.

281 Marker development and testing of the whole mapping population

282 After creating haplotype specific allele sequences through amplicon sequencing and the subsequent
283 bioinformatic pipeline, markers were designed for haplotype specific PCRs. Obtained allele
284 sequences of target genes were scanned for InDel structures differing between the parental alleles.
285 Variants with low coverage or low quality were filtered out. In the case that InDels were filtered
286 out, the actual allele sequence can be greater than the calculated one. The sequence information was
287 used for subsequent STS (Sequence-Tagged Sites) marker design with the Primer3 tool [36].
288 Primers had an optimum Tm of 58 - 60 °C, with PCR products differing in size between 100 – 400
289 bp for multiplexing purposes (S7 Table). Forward primers were labeled at the 5'end with one of the
290 fluorescent dyes 6-FAM (blue), HEX (green), TAMRA (yellow) or ROX (red). Allele distributions
291 were analyzed over all 151 F_1 individuals of the mapping population GF.GA-47-42 x 'Villard
292 Blanc'. PCRs were carried out with the QIAGEN multiplex PCR kit (Qiagen GmbH, Hilden,
293 Germany) following the instructions of the manufacturer in three multiplexes combining different
294 product sizes and fluorescent dyes. Resulting PCR products were analyzed on an ABI 3110xl
295 Genetic Analyzer (Applied Biosystems, Foster City, USA) and the results compared with the
296 respective phenotype of the tested individual (i.e. early, intermediate or late flowering).

297 RNA extraction and sequencing

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

Total RNA was extracted from up to 100 mg of liquid nitrogen ground tissue using the Spectrum™ Plant Total RNA kit (Sigma-Aldrich, Taufkirchen, Germany) according to the manufacturer's instructions for protocol B. After on-column DNase treatment with the DNase I Digest Set (Sigma-Aldrich, Taufkirchen, Germany) the RNA was quantified. RNA-libraries for each time point were prepared according to the Illumina TruSeq RNA Sample Preparation v2 Kit using an input of 1 µg of total RNA. RNA-Seq (1x 135 bp) was performed on an Illumina Rapid HiSeq-1500 Run. One barcoded library was created for each of the time points.

RNA-Seq read processing for analysis of gene expression kinetics

Read trimming and quality control was performed as described above in "Read processing and mapping". Sequence read data are available from SRA accession SRP153932. The reads were mapped to the grapevine reference sequence PN40024 12x.v2 [14] using tophat2 [43] which is capable of performing split read mapping. The maximal intron size was set to 3000, otherwise default parameters were used. Resulting BAM-format files were sorted and indexed using SAMtools [39]. With HTSeq [44] mapped reads were counted for each gene. Differential gene expression was analyzed using the R-package DESeq2 [45]. In order to perform an analysis of expression without replicates, the counts were modeled as a smooth function of time, and an interaction term of the condition with the smooth function was included. Likelihood ratio test of DESeq2s with a reduced design, which does not include the interaction term, was then applied. Genes with small p-values from this test are those showing a time-specific effect.

Results

Phenotypic evaluation of the mapping population

The 151 F_1 individuals of the segregating population and their parental lines were phenotyped for time of full bloom as indicated in S2 Figure showing the timing of flowering in days after January 1st. The length of the flowering period varied considerably between 10 days (2016) and 17 days (2012) [23]. In the year 2010 the flowering period was heavily extended compared to the other years. The greatest portion of individuals within the population reached full bloom in approximately the first third of the flowering period. Within the mapping population, early flowering is inherited from the maternal genotype GF.GA-47-42.

Identification of FTC candidate genes

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

Functional data from *A. thaliana* and other model organisms was systematically exploited to identify FTC candidate genes in the *Vitis* reference genome sequence. More than 500 homologous genes were identified which are distributed over all chromosomes including the unanchored, random part of the sequence (S4 Table). Some of the genes are absent from the CRIBI annotations, but were included in the previous annotations, provided by Genoscope. To our knowledge the majority of the identified FTC candidate genes was not analyzed or even mentioned in a previous publication. As expected, an enrichment of the FTC candidate genes (75) annotated within the FTC QTL regions was found. In several cases we identified more than one homologous sequence in the grapevine genome with a single copy *Arabidopsis* query. In these cases not necessarily the gene with the highest sequence similarity is the one in the FTC QTL region, nor the one with the highest expression in flowering related tissues. For instance the *RAV* genes *VvRAV1b* and *VvRAV1c* are located within the QTL regions on chr 1 and chr 14, respectively, whereas the *RAV1a* is located on chr 11 outside of any FTC QTL.

Many of the FTC candidate genes are transcription factors involved in flower development and morphogenesis such as members of the AP2/EREBP family [46] and homeodomain proteins [47]. About eight MYB-transcription factors that participate in cell cycle control in many living taxa [48] were among the identified FTC candidate genes in *Vitis*. Several other protein families were among the FTC candidate genes, such as a dozen GRAS and FRIGIDA proteins that are involved in flowering time and plant development. FRIGIDA proteins are required for the regulation of flowering time by upregulating *FLC* expression. Allelic variation at the FRIGIDA locus is an important determinant of natural variation in the timing of flowering [49]. The GRAS (GAI, RGA, SCR) family is a very important family of proteins involved in flowering in grapevine. GRAS proteins participate in GA signaling, which influences numerous aspects of plant growth and development [50]. Remarkably sixteen SQUAMOSA PROMOTER BINDING PROTEIN (SBP)-domain proteins, that are known from other plants as transcriptional activators involved in a variety of processes such as flower and fruit development, plant architecture, GA signaling, and the control of early flower development [51] are candidates.

Allele phasing

From our comprehensive list of *V. vinifera* FTC candidates the 72 most promising genes were chosen as targets for amplicon sequencing (S5 Table), many of which are located in flowering related QTL regions on chr 1, 14, and 17 [23]. The average read depth of coverage was 286 (SD: 276) and for most samples sequencing depth was between 100 and 300. Variants in the analyzed

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

359 lines were detected with a density between 1.02 and 1.63 variants per 100 bp most of which were
360 SNPs.

361 In order to link certain alleles of the sequenced candidate genes to the flowering time phenotype,
362 the two alleles of genes had to be reconstructed from the mix of sequenced fragments of the two
363 alleles. The phasing of alleles was performed on the basis of sites polymorphic between the two
364 alleles of a gene.

365 Aside from recombination events, a parent-offspring pair must share one haplotype for each
366 chromosome and thus one identical-by-descent allele for every gene. Hence, Mendelian constraints
367 could be applied to validate the obtained allele-specific sequence. Alleles of the chosen 72 target
368 genes studied could be identified in 46 cases (S5 Table).

369 In 23 cases four different allele sequences could be found, three allele sequences in 18 cases, two in
370 four cases and in one case (VIT_217s0000g00150; *VvFL*) only one allele sequence, meaning that all
371 individuals of the population were homozygous for the respective locus. This fits the expectation
372 since grapevine is highly heterozygous. The number of allele sequences has been deduced from
373 regions of the genes in which phasing was performed. The lengths of the phased intervals were
374 between 204 and 8,285 bp (S5 Table).

375 Correlation analysis

376 Allele sequences of the progeny of the mapping population GF.GA-47-42 x 'Villard Blanc' were
377 compared against the allele sequences of the parental lines to determine the inheritance pattern
378 within the population for each gene. In order to find alleles correlating with the phenotype of
379 flowering time, a correlation analysis between the phased alleles of FTC target genes and flowering
380 time phenotypes was performed. Several sets of phenotypic data were used. For the years 1999,
381 2009 - 2016 a correlation analysis was performed using days after January 1st of the respective year.
382 Additionally for the years 2011 - 2016 values of accumulated temperature above 3°C from
383 November 1st of the previous year and global radiation in KWh/m² from January 1st were
384 considered.

385 After the reconstruction of inheritance patterns within the parental lines and the 35 analyzed F_1
386 individuals of the mapping population GF.GA-47-42 x 'Villard Blanc' through the amplicon
387 sequencing approach and subsequent bioinformatic analysis, the numbers of individuals harboring
388 each of the alleles was determined and a correlation analysis between alleles of FTC target genes
389 and the flowering time phenotype was performed for 43 genes. A correlation between alleles and

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

flowering time could be observed for several genes on chr 1, 4, 14, 17, 18, and within unassigned contigs. Correlation values differed depending on whether days, accumulated temperature or global radiation was used as phenotypic data. As an example Fig. 2 shows allele combinations in the parental lines of the population GF.GA-47-42 x 'Villard Blanc' and the p-values of the correlation of alleles unique to one of the lines. Values equal and below 0.05 were considered to be significant and the lower the p-value the higher is the correlation. In total for 16 FTC target gene alleles a significant correlation with either an early or late flowering phenotype could be found.

Fig 2: Correlation between alleles of FTC target genes and flowering time phenotype. Given are the allele constitutions of the parental lines for each gene and the allele counts of the amplicon sequenced F_1 individuals. The median of flowering time (calculated from days after January 1st of the years 1999 and 2010-2016) of individuals carrying the counted is given in brackets. The higher the value of the median, the later the flowering phenotype of the F_1 individuals. Color coded are the p-values for the E alleles and L alleles in the up to 35 F_1 individuals. Significant correlation values are in bold and italic. Genes located in QTL regions are marked in grey. Differences in allele counts between the years are due to missing data points. "E" alleles are inherited from GF.GA-47-42, while "L" alleles originate from 'Villard Blanc'. "N" means that both GF.GA-47-42 and 'Villard Blanc' share one or more alleles. "E0": E1=E2, "L0": L1=L2, "N1": E1=L1, "N2": E2=L2. "N": L2=E1 or E2=L1, "Na": E1=E2=L1, "Nb": E1=E2=L2. „n.d." : not determined. Further explanations are given in S3 Table.

The L2 alleles, inherited from the paternal line 'Villard Blanc', of *VvSEP4 (SEPALLATA 4), VvBS2, VvHUA2a, VvRAV1b*, and *VvGAI1* (chr 1) correlate with late flowering, strengthen the importance of the FTC QTL on chr1. The E1 alleles of the two genes *VvWNK6 (V. vinifera WITH NO LYSIN KINASE 6)* and *VvTM6* (*V. vinifera TOMATO MADS-BOX 6*), both located on chr 4 and inherited from the early flowering maternal line, were found to strongly correlate with early flowering. The p-values calculated from the median (Fig 2 is p = 0.007 and values down to p = 0.003 were observed for single years. Table 3 shows the p-values of correlation for different sets of phenotypic data related to *VvWNK6* and *VvTM6*. Most of the significant correlations are obvious regardless the year or scale of phenotyping (days after January 1st, accumulated temperature or global radiation). The differences in correlation among years are due to the seasonal weather conditions of the respective year, which influence both the flowering time and the length of the flowering period. A significant correlation between the E1 allele of Vv*WNK6* and the flowering time phenotype could not be observed in 2016 for neither days after January 1st, accumulated temperature or global radiation. In 2015, the correlation was not significant for days after January 1st but, albeit only

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

423 slightly, for the other two sets of phenotypic data. Other genes, such as *VvMFT* (*V. vinifera*
424 *MOTHER of FT and TFL1*) showed significant correlation in 2016 but not in 2013.

425 Table 3: P-values of the correlation between the E1 allele distribution of *VvWNK6* and *VvTM6* in
426 relation to different sets of phenotypic data using 35 amplicon sequenced F_1 individuals.

427

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

	VvWNK6 (E1)	*VvTM6* (E1)
Days after January 1st/ 1999	0.032	0.023
Days after January 1st/ 2009	0.012	0.009
Days after January 1st/ 2010	0.44	0.37
Days after January 1st/ 2011	0.033	0.063
Days after January 1st/ 2012	0.047	0.041
Days after January 1st/ 2013	0.008	0.012
Days after January 1st/ 2014	0.015	0.029
Days after January 1st/ 2015	0.067	0.063
Days after January 1st/ 2016	0.177	0.098
Median for days after January 1st/ 1999-2016	0.012	0.009
Accumulated Temp. above 3°C/ 2011	0.027	0.109
Accumulated Temp. above 3°C/ 2012	0.03	0.091
Accumulated Temp. above 3°C/ 2013	0.004	0.058
Accumulated Temp. above 3°C/ 2014	0.003	0.016
Accumulated Temp. above 3°C/ 2015	0.046	0.186
Accumulated Temp. above 3°C/ 2016	0.177	0.098
Global radiation (KWh/m²)/ 2011	0.027	0.109
Global radiation (KWh/m²)/ 2012	0.03	0.091
Global radiation (KWh/m²)/ 2013	0.004	0.058
Global radiation (KWh/m²)/ 2014	0.003	0.016
Global radiation (KWh/m²)/ 2015	0.046	0.186
Global radiation (KWh/m²)/ 2016	0.177	0.098

428

429 Compared to the reference sequence, the E1 allele of *VvWNK6* (chr 4) was found to harbor a
430 variation in the terminal exon (SNP at chr4:21997435/ C → T) leading to an amino acid exchange
431 from threonine to methionine. Fig. 3 shows the distribution of allele combinations for *VvWNK6*

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

among individuals of the mapping population. Early flowering is associated with the E1 allele inherited from the maternal 'Bacchus' allele of GF.GA-47-42.

Fig 3: Distribution of allele combinations for *VvWNK6* (chr 4) among 35 selected individuals of the mapping population GF.GA-47-42 x 'Villard Blanc'. The date of flowering was counted in days from the 1st of January and the data was subsequently classified according to six stages for flowering time following (1 = very early flowering; 2 = early flowering; 3 = medium early flowering; 4 = medium late flowering; 5 = late flowering; 6 = very late flowering). For visualization flowering classes 1 and 2, 3 and 4, and 5 and 6 were merged.

Application of the pipeline for amplicon sequencing in a heterozygous plant for subsequent marker design

Amplicon sequencing was performed in 35 F_1 individuals and the parents of the mapping population. In order to investigate the resulting allele distributions over all 151 F_1 individuals of the mapping population GF.GA-47-42 x 'Villard Blanc', STS markers were designed from the allele sequences that enabled an easy allele-specific genotyping. The information obtained from amplicon sequencing of the FTC target genes proved usable for both deduction of segregation patterns and marker design for investigating allele distribution over the whole mapping population. Table 4 gives an overview of the segregation patterns as analyzed for all 151 F_1 individuals. From 15 markers 12 showed a segregation pattern matching the segregation pattern that was obtained through allele phasing. The markers GAVBInd_019 and GAVBInd_020 were not designed using the obtained allele sequences of GF.GA-47-42 and 'Villard Blanc', since suitable InDels were not available. Therefore, these markers were designed based on InDels upstream of the phased regions. Observed product sizes can deviate from the expected ones by 1 – 2 bp due to the limited accuracy of the used fragment analyzing method. Markers GAVBInd_004, GAVBInd_014, and GAVBInd_019 showed two different segregation patterns since the measuring method cannot reliably resolve differences of 1 – 2 bp. See S6 Table for further details.

Table 4: Comparison of the expected and observed allele sizes (bp) and segregation patterns of several FTC target genes. Expected data were obtained through amplicon sequencing; observed data were gained by analyzing 151 F_1 individuals of the mapping population GF.GA-47-42 x 'Villard Blanc' with STS markers located within the FTC target genes. ab x cd: four alleles/both parents heterozygous, hk x hk: 2 alleles/both parents heterozygous, ef x eg: 3 alleles/both parents heterozygous, lm x ll: 2 alleles/ mother heterozygous, nn x np: 2 alleles, father heterozygous. x: amplification failed. See Table S6 for further information.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

		Expected product sizes		**Expected segregation**	**Observed product sizes**		**Observed segregation**
Gene	**Marker name**	**GF.GA-47-42**	**'Villard Blanc'**		**GF.GA-47-42**	**'Villard Blanc'**	
VvHLH74	GAVBInd_009	147/155	155/155	lmxll	137/146	146/146	lmxll
VvBHLH74	GAVBInd_010	230/233	234/230	hkxhk	231/237	231/237	hkxhk
VvHUA2	GAVBInd_001	452/455	437/456	bbxcd	454/454	441/454	nnxnp
VvCOL10	GAVBInd_004	197/195	197/196	hkxhk	195/197	195/197	hkxhk
VvCOL10	GAVBInd_004	197/195	197/196	efxeg	194/196	195/196	efxeg
VvCOL10	GAVBInd_005	146/115	155/155	lmxll	110/142	110/110	lmxll
VvWNK6	GAVBInd_019	N/A	N/A	N/A	210/x	217/218	abxcd
VvWNK6	GAVBInd_019	N/A	N/A	N/A	210/x	X	lmxll
VvFPA	GAVBInd_007	362/353	363/365	efxeg	363/x	X	lmxll
VvGAMYBc	GAVBInd_014	401/422	435/426	abxcd	407/428	432/437	abxcd
VvGAMYBc	GAVBInd_014	401/422	435/426	abxcd	457/x	X	lmxll
VvCOL5	GAVBInd_015	309/312	312/312	lmxll	195/198	198/198	lmxll
VvTOE3	GAVBInd_016	276/275	268/275	efxeg	279/282	272/279	efxeg
VvTOE3	GAVBInd_017	139/144	144/144	lmxll	136/140	140/1740	lmxll
VvPRR37b	GAVBInd_018	281/286	286/286	lmxll	281/286	286/286	lmxll
VvGAIb	GAVBInd_006	231/241	231/236	efxeg	231/245	231/237	efxeg
VvFLKa	GAVBInd_012	133/182	182/182	lmxll	128/178	178/178	lmxll
VvFLKa	GAVBInd_013	213/217	213/213	abxcd	211/215	211/215	hkxhk
VvFUL2	GAVBInd_020	N/A	N/A	N/A	443/433	414/444	abxcd
VvSVP2	GAVBInd_008	236/247	248/248	lmxll	238/244	X	lmxll

464

465 Using the results of marker segregation across the 151 F_1 individuals, a correlation analysis between
466 alleles and flowering time phenotypes was performed. The correlation results of marker analysis
467 support those of allele phasing (Table 5). See S7 Table for further details.

468 Table 5: P-values of correlation between alleles and the phenotype of flowering time from both the
469 allele phasing workflow (first row) and marker analysis (second row) based on days after January
470 1st on the median of the years 1999 and 2009-2016. Marker analysis was performed in 151 F_1
471 individuals of the population GF.GA-47-42 x 'Villard Blanc', while allele phasing was performed
472 in 35 F_1 individuals. Number of alleles over the analyzed F_1 individuals and the median of each, are
473 given in the same order as in column 3. ab x cd: four alleles/both parents heterozygous, ef x eg: 3
474 alleles/both parents heterozygous, lm x ll: 2 alleles/ mother heterozygous, nn x np: 2 alleles, father
475 heterozygous.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

Gene	Marker name	Segregation GF.GA-47-42 x 'Villard Blanc'	p-values of correlation between alleles and phenotypes		Allele numbers				Median			
			GF.GA-47-42	'Villard Blanc'	GF.GA-47-42		'Villard Blanc'		GF.GA-47-42		'Villard Blanc'	
VvbHLH49		E1E2 x L1L2	0.438	0.77	13	14	12	15	0.48	0.62	0.56	0.48
	GAVBInd_009	lm x ll (m = E1)	0.32	/	64	67	/	/	0.48	0.5	/	/
VvHUA2		E1E2 x L1L2	0.076	0.04	14	12	11	15	0.41	0.665	0.41	0.66
	GAVBInd_001	nn x np (p = L1)	/	0.02	/	/	67	65	/	/	0.47	0.51
VvCOL10		E1E2 x L1L2	0.018	0.097	12	15	12	15	0.41	0.67	0.48	0.66
	GAVBInd_005	lm x ll (m = E2)	0.295		66	68	/	/	0.46	0.505	/	/
	GAVBInd_004	ef x eg (f=E2, g=L2)	0.4	0.39	54	75	68	61	0.48	0.5	0.5	0.5
VvWNK6		E1N2 x L1N2	0.007	0.703	25	10	19	16	0.55	0.725	0.61	0.575
	GAVBInd_019	ab x cd	0	0.573	81	67	66	82	0.45	0.54	0.475	0.5
	GAVBInd_019	lm x ll	0	/	81	67	/	/	0.45	0.54	/	/
VvFPA		E1E2 x L1L2	0.536	0.257	7	28	20	15	0.55	0.6	0.6	0.55
	GAVBInd_007	lm x ll (m = E2)	0.7	/	53	79	/	/	0.48	0.5	/	/
VvGAMYBc		E1E2 x L1L2	0.691	0.68	16	19	27	8	0.57	0.61	0.6	0.625
	GAVBInd_014	ab x cd (a = E1, b=E2, c= L1, d L2)	0.45	0.79	70	59	96	33	0.5	0.5	0.5	0.46
VvCOL5		E1E2 x L1L2	0.445	0.567	15	19	19	15	0.66	0.6	0.59	0.6
	GAVBInd_015	lm x ll (m = E2)	0.81	/	67	64	/	/	0.48	0.5	/	/
VvTOE3		E1E2 x L1L2	0.231	0.943	19	15	12	22	0.56	0.61	0.605	0.57

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

	GAVBInd_016	ef x eg (f=E2, g=L2)	0.23	0.85	69	66	68	67	0.5	0.49	0.5	0.5
	GAVBInd_017	lm x ll (m = E2)	0.13	/	71	65	/	/	0.5	0.48	/	/
VvPRR37b		E1E2 x L1L2	0.73	0.431	14	14	9	19	0.595	0.61	0.6	0.59
	GAVBInd_018	lm x ll (m = E1)	0.88	/	67	66	/	/	0.48	0.5	/	/
VvGAIb		E1E2 x L1L2	0.027	0.059	12	16	13	15	0.665	0.41	0.69	0.46
	GAVBInd_006	ef x eg (f=E1, g=L2)	0	0.25	80	55	70	65	0.45	0.58	0.48	0.51
VvFLKa		E1E2 x L1L2	0.069	0.392	7	19	15	11	0.67	0.49	0.6	0.55
	GAVBInd_012	lm x ll (m = E2)	0.01	/	79	59	/	/	0.46	0.56	/	/
VvFUL2		E1E2 x L0L0	0.003	/	12	13	/	/	0.725	0.41	/	/
	GAVBInd_020	Abxcd	0	0.93	58	89	67	80	0.575	0.44	0.5	0.49
VvSVP2		E1E2 x L1L2	0.05	0.064	17	17	15	19	0.49	0.6	0.41	0.61
	GAVBInd_008	lm x ll (m = E2)	0.86	/	62	66	/	/	0.48	0.5	/	/

476

477 # Analysis of gene expression kinetics

478 Variation in expression could be detected in both time courses 2012/2013 and 2013/2014 for
479 various FTC candidate and target genes when testing for time-specific effects. Between consecutive
480 developmental stages of bud differentiation before dormancy (August 2nd to September 5th, 2013
481 time series 1, Table 2) differences in expression could be detected for the MADS transcription
482 factor *VvTM8* as well as the protein kinase encoding gene Vv*WNK5*. *VvTM8* encodes a MIKC
483 transcription factor whose *A. thaliana* homologue *AtTM8* has been shown to be involved in the
484 specification of flower organ identity [25].

485 In a time course of dormant buds (BBCH 0) until after bud burst when leaf formation had already
486 begun (BBCH 11-13), 58 of the FTC candidate genes were found to show a BBCH or
487 developmental stage-dependent expression. Several of these genes are squamosa binding proteins,

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

MADS- and MYC-transcription factors that are known to influence floral development. Most of these genes show a variation in gene expression due to an up or down regulation towards developmental stages during inflorescence maturation. In order to test for expression variation between consecutive developmental stages of bud development before inflorescence structures become externally visible, inflorescences collected after bud break were excluded from the analysis. Genes with different expression kinetics when the time course was extended to include visible inflorescences, are those showing a clear variation in gene expression between buds and inflorescence. In total 67 of such "inflorescence-specific genes" were identified (S8 Table).

After excluding inflorescences, several genes were found showing an obvious time-dependent expression. They cluster into two groups: genes upregulated in winter during bud dormancy (Fig 4, upper part) and genes upregulated towards inflorescence development (Fig 4, lower part). Most of these genes encode BZIP-, MADS- or MYC-transcription factors, which regulate other flowering related genes. Downregulation towards bud burst and inflorescence maturation was found for transcription factor genes involved in circadian rhythm such as *VvGRP2A (Glycine Rich Protein 2A), VvRVE1 (REVEILLE), VvTICb* (*TIME FOR COFFEE*) and *VvELF3* (*EARLY FLOWERING3*). Moreover, genes coding for transcription factors involved in gibberellic acid (GA) biosynthesis were found to be upregulated during bud dormancy. Numerous other genes like *VvHUA2b (ENHANCER OF AGAMOUS)*, which is involved in the repression of floral transition and flower development, were found to be upregulated during bud dormancy.

Fig 4: Heatmap of FTC candidate genes showing variations in their expression over consecutive time points of bud development from dormancy until appearance of inflorescence in grapevine variety GF.GA-47-42. Time series from December 20th, 2012 to May 3rd, 2013. LFC-threshold: 2 = expression fourfolded, -2 = expression quartered. Shown are rlog transformed counts.

For most of the genes (Fig 4) an up- or downregulation in expression is observed between the first and the second time point during bud dormancy. Many genes also show an up- or downregulation in expression between the third and the fourth time point when swelling buds are developing.

The gene expression for the amplicon sequenced target genes in buds and inflorescences is shown in Fig 5. Some genes are not expressed at all, while some are only expressed before dormancy or in inflorescence tissue. However, up- or downregulation in gene expression mainly occurs when swelling buds develop. Genes involved in floral development, such as *VvSEP3* and *4*, *VvAP1*, and *VvTM6* show an increased expression in developing inflorescences. *VvTM6* is a MADS-box B-class floral identity gene influencing the development of petals and stamen [52,53]. In *Vitis* all three B-

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

520 class floral homeotic genes (*VvPI*, *VvAP3* and *VvTM6*) are highly expressed in inflorescences (S3
521 Figure).

522 Fig 5: Heatmap of gene expression of amplicon sequenced FTC candidate genes in GF.GA-47-42 at
523 different developmental stages of buds and inflorescences. LFC-threshold: 1 = expression doubled,
524 -1 = expression halved. rlog transformed counts are shown.

525 For three selected time points, bud/inflorescence samples and the corresponding leaf from the same
526 node were collected and differential gene expression was analyzed between leaves and the
527 associated bud/inflorescence. Fig 6 shows a heatmap of the FTC candidate genes with expression
528 differences between leaves and buds/inflorescences. With few exceptions, all genes with expression
529 differences between leaves and buds or inflorescences are downregulated or not expressed in leaves.

530 Fig 6: Heatmap of FTC candidate genes showing expression variations between leaves and their
531 prompt buds/ inflorescences. LFC-threshold: 2 = expression fourfolded, -2 = expression quartered.
532 Shown are rlog transformed counts.

533 Discussion

534 FTC candidate genes

535 A large number of FTC candidate genes inside and outside of known flowering QTLs in grapevine
536 were identified. Although the identification relies mostly on sequence homology to previously
537 known genes from other plants, the putative functional connection via e.g. Pfam, literature search or
538 the performed RNA-Seq experiments substantiate the reliability of the prediction. This
539 comprehensive gene list opens the door for investigations on e.g. flowering time networks in the
540 future. One the one hand, compared to *Arabidopsis thaliana* there is probably an overestimation of
541 FTC candidate genes in *Vitis*. On the other hand the high complexity and long duration of bud
542 initiation and flower development may require a large number of genes.

543 Allele phasing of target genes

544 A workflow for the phasing of amplicon sequenced genes using Illumina short-read sequencing of
545 a diploid organism was established and successfully applied to separate alleles in regions with a
546 length of up to 8.3 kb. By analyzing inheritance patterns within a family of parents and F_1
547 individuals, we could show that the inheritance of alleles of neighboring genes within a QTL
548 remains largely constant throughout the QTL. Since grapevine has a highly heterozygous genome

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

549 and suffers from inbreeding depression, we used a F_1 mapping population and followed a double
550 pseudo-testcross strategy [54]. Therefore, a lower recombination frequency was expected compared
551 to typical F_2 mapping populations in other plant species. The constancy of the inheritance pattern of
552 alleles of closely neighboring genes indicates the functionality and applicability of the established
553 allele phasing method.

554 For the phasing of alleles, a mapped read or read pair needs to encompass two or more
555 heterozygous sequence positions. The phase of the heterozygous sequence positions can be
556 determined since each read or pair of reads is obtained from a single haplotype. Read lengths after
557 trimming was distributed between 80 and 300 bp with an average insert size of ~500 bp. When
558 variants were located farther apart than the maximum length that could be spanned by a read pair,
559 alleles could not be phased despite the presence of variants. Moreover, the allele frequency,
560 calculated from the read coverage of variants can vary despite being amplified from the same allele.
561 The amount of reads covering a variant can differ from one variant to the next. When dealing with
562 extremely biased allele frequencies, this can lead to some variants being detected while others
563 remain undetected. In such cases allele phasing was unsuccessful. Some amplicons could hardly be
564 amplified at all. This is likely due to a high diversity at the primer binding sites between the
565 reference sequence and the plant lines analyzed in this work.

566 The use of paired-end sequencing is highly advantageous in haplotype phasing as it covers variants
567 that are spaced at distances longer than the technology's read length limit. Read length in high-
568 throughput sequencing is constantly increasing and technologies are evolving rapidly. With the rise
569 of third generation technologies, capable of producing even longer reads, many of the difficulties
570 associated with haplotype phasing might soon be alleviated as such data may permit direct phasing
571 from sequence reads [26].

572 Correlation analysis

573 We were able to detect a correlation between alleles of FTC target genes and flowering time for
574 several QTL regions, which supports the role of these regions in the timing of flowering. Flowering
575 time is highly dependent on the weather conditions of the respective and previous year. Therefore,
576 correlation values vary between the years, as observed e.g., for Vv*WNK6* in 2016 (Table 3).

577 Alleles of FTC target genes within a QTL region on chr 1 were found to rather correlate with late,
578 while QTL regions on chr 4 and 14 were found to correlate with early flowering. With one
579 exception, all analyzed F_1 individuals carrying alleles correlating with flowering time from two of

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

580 the QTL regions on chr 1, 4, and 14 or all three of them were either intermediate-early, early, or
581 very early flowering. The correlation for the QTL regions on chr 4 and 14 was more stable than for
582 chr 1 indicating a stronger affect of these QTLs in the timing of flowering. The investigation of
583 epistatic effects between these QTL regions could contribute to the clarification of the genetic
584 factors that influence and control flowering time in grapevine.

585 Correlation values between alleles of FTC target genes and flowering time phenotypes could be
586 largely supported by genetic marker analysis. Deviations can be due to the measuring method that
587 can occasionally lead to deviations of up to two bp in product size. In order to distinguish the
588 maximum putative number of alleles at a single locus within a bi-parental F_1 population of a diploid
589 organism, the marker needs to be capable of distinguishing between four different alleles.

590 Classic high informative marker analysis requires InDels / SSRs that distinguish between the
591 maximum number of different alleles with polymorphic differences of at least two bp in size at a
592 specific locus. The usage of blocks of tightly linked polymorphisms and treating each haplotype of
593 these blocks as a separate allele can produce highly polymorphic markers. In addition, it also uses
594 SNPs and InDels shorter than two bp to distinguish between the alleles. This leads to a higher
595 resolution compared to classic marker analysis and the detection of a higher number of different
596 alleles.

597 The correlation of alleles of FTC genes with flowering time phentoypes is based on the genotypic
598 data on one hand, which is obtained through the allele phasing workflow, from amplicon
599 sequencing, mapping and variant calling to the final establishment of allele sequences. On the other
600 hand, the correlation analysis is based on the phenotypic data, which is also prone to errors.
601 Phenotyping of flowering time was performed on a daily basis throughout the flowering phase.
602 Differences in the timing of flowering shorter than one day are therefore not recorded. Moreover,
603 phenotyping is a subjective process when different people work on the recording of phenotypic data
604 and hence a possible error source.

605 As already mentioned, the timing of flowering depends clearly on environmental parameters,
606 especially weather and climatic conditions. These are most probably non-genetic factors causing the
607 differences in the flowering periods between the respective years. In 2016, for example, flowering
608 in the population GF.GA-47-42 x 'Villard Blanc' started on June 17th being very late compared to
609 other years (Table 2). However, the flowering period was very short, ending after only 10 days on
610 June 26th. Global radiation is distributed between ~502 and ~536 KWh/m² at the beginning of
611 flowering in the analyzed population and between ~548 and ~597 KWh/m² at the end of it. While

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

612 flowering occurred very late in 2016 compared to other years, the amount of global radiation until
613 the first day of the flowering period was less than in the other years. This shows that the amount of
614 solar radiation before flowering initiation was small which might have had an impact on the timing
615 of flowering.

616 In some cases the p-value of correlation is significant although the medians are nearly equal or
617 equal. This is because the Wilcoxon Rank-Sum test is a rank sum tests and not a median test. It
618 ranks all of the observations from both groups and then sums the ranks from one of the groups and
619 compares it with the expected rank sum. Therefore, it is in rare cases for groups possible to have
620 different rank sums and yet have equal or nearly equal medians.

621 *VvHUA2a* of which an amplicon sequenced allele from 'Villard Blanc' was found to correlate with
622 late flowering is a floral homeotic gene. It's homologue in *A. thaliana, HUA2*, regulates the
623 expression of the floral homeotic class-C gene *AGAMOUS (AG)* and *FLC [55]*. This suggests a role
624 of *VvHUA2* in the delay of flowering.

625 An allele of *VvGAI1* from late flowering 'Villard Blanc' was found to correlate with late flowering.
626 Mutants of *VvGAI1* are insensitive to gibberellic acid and form inflorescences instead of tendrils.
627 These mutants show a correlation between inflorescence development and increased *VvFL*
628 expression, a floral developmental gene [56]. In *A. thaliana, GAI* acts as a repressor of *LFY* and
629 *SOC1* and thus represses flowering.

630

631 From the amplicon sequenced and early flowering individuals (median data) of the population
632 GF.GA-47-42 x 'Villard Blanc', 90% were found to carry the *VvTM6* E1 allele inherited from
633 GF.GA-47-42. Only 10% of plants that carry the other maternal allele are early flowering. *VvTM6* is
634 a MADS-box B-class floral identity gene and influences the development of petals and stamen. In
635 *A. thaliana*, mutants exhibit a transformation of petals to sepals and stamen to carpels. B-class floral
636 homeotic genes either belong to the paleoAPETALA3 or to the PISTILLATA (PI) gene lineage,
637 which are paralogous and resulted from a duplication event before the emergence of angiosperms.
638 The paleoAP3 lineage underwent a further duplication event at the base of the core eudicots
639 resulting in the two sublineages *euAP3* and *TM6* (named after the Tomato MADS-box gene 6) [57].
640 A *TM6* homologue is absent in *A. thaliana* [52,57,58]. In grapevine, all three B-class floral
641 homeotic genes were found to be highly expressed in inflorescences (Fig 5) but not in leaves (Fig
642 6). [25] showed that *VvTM6 (VvAP3.2)* is expressed in fruits, while the expression of *VvAP3*

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

643 (*VvAP3.1*) and *VvPI* is more restricted to flowers. Also, [52] showed that the expression of *VvTM6*
644 is higher in carpels, fruits, and seeds than in petals. Due to the expression of *VvTM6* in carpels and
645 during berry development and ripening, it was suggested to play an important role in grapevine fruit
646 development [25]. The expression of *VvTM6* increases towards inflorescence maturation, which is
647 followed by berry formation and ripening. This is consistent with its role during berry development
648 and ripening.

649 All early flowering amplicon sequenced individuals of the population GF.GA-47-42 x 'Villard
650 Blanc' were observed to carry the E1 allele of *VvWNK6* (Fig 3). In *A. thaliana WNK6* has been
651 shown to be involved in circadian rhythm [59]. WNKs are a subfamily of serine/threonine protein
652 kinases with a lysine residue essential for ATP-binding, which is located in kinase subdomain I
653 instead of subdomain II as common among all other kinases [60]. It has been suggested that *WNK*
654 gene family members regulate flowering time in *A. thaliana* by modulating the photoperiod
655 pathway. For instance, APRR3, a component of the clock-associated APRR1/TOC1 quintet is a
656 substrate of WNK1 in *A. thaliana*. T-DNA knockout mutants of *AtWNK1* are delayed in flowering
657 time while T-DNA knockout mutants of *AtWNK2, 5,* and *8* flower early [61]. *WNK6* transcription is
658 downregulated in *AtABI4* mutants, which show an early flowering phenotype [62]. In *A. thaliana,*
659 *ABI4* negatively regulates flowering through directly promoting *FLC* transcription, a negative
660 regulator of flowering [63]. This might indicate that *VvWNK6* is involved in the delay of flowering.
661 *VvWNK6* expression was detected in leaves, buds, and inflorescences of the early flowering
662 GF.GA-47-42. Both alleles E1 and E2 are expressed at a similar level. However, all individuals of
663 the mapping population carrying the E1 allele of *VvWNK6* flower early. This suggests that either the
664 E1 allele of *VvWNK6* itself might contribute to early flowering or alleles of other nearby-genes
665 inherited together with E1 of *VvWNK6.* Further analysis should include the investigation of
666 sequence variations leading to an alteration of the amino acid sequence and the functionality of the
667 protein.

668 Gene expression kinetics

669 Many of the analyzed FTC candidate genes show variations in expression pattern in the course of
670 the developmental cycle, supporting their role in flowering time control. Genes coding for
671 transcription factors and other proteins involved in inflorescence architecture, floral transition and
672 flower development are usually upregulated after bud burst, while genes coding for proteins that
673 repress flowering in diverse manners typically show an upregulation during bud dormancy (Fig 5).
674 Among the genes showing downregulation towards bud burst and inflorescence maturation are
675 transcription factors involved in circadian rhythm such as *VvGRP2A (Glycine Rich Protein 2A)*,

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

676 *VvRVE1 (REVEILLE1)*, *VvTICb (TIME FOR COFFEE)* and *VvELF3 (EARLY FLOWERING3)*. It is
677 not unexpected to detect different gene expression kinetics for genes involved in circadian rhythm
678 since sampling was performed at the same time of the day over the entire time course. However, the
679 period from daybreak until the time of sampling varies throughout the year and the different
680 seasons.

681 *AtGRP7,* the homologue of *VvGRP2A* in *A. thaliana,* undergoes circadian oscillations with peak
682 levels in the evening [64]. *RVE* is a MYB-like transcription factor that controls auxin levels,
683 promotes free auxin and hence plant growth during the day [65]. *TIC* and *ELF3* are components of
684 the circadian clock in *A. thaliana*. *ELF3* is a circadian clock gene that contributes to photoperiod-
685 dependent flowering in plants [66-68]. Our findings thus suggest a similar impact of these genes in
686 grapevine.

687 Moreover, genes coding for transcription factors involved in GA biosynthesis were found to be
688 upregulated during bud dormancy. GAs are inhibitors of flowering in many fruit species but their
689 role in grapevine varies with the stage of bud development. The initiation and development of
690 lateral meristems is promoted by GAs as well as their development into tendrils, while
691 inflorescence development is suppressed by GAs. Thus GA is a promoter of flowering at an early
692 stage but acts as an inhibitor of flowering later on and promotes vegetative growth [19]. *SPY*
693 *(SPINDLY),* whose *Vitis* homologue *VvSPY* was found to be upregulated during bud dormancy, is a
694 negative regulator of GA response in *A. thaliana* and functions with *GI (GIGANTEA)* in pathways
695 controlling flowering [69]. In *Vitis* the role of *SPY* in GA signaling is still unclear. It could be
696 shown that treatment of grapevine plants at pre-bloom stage with GA led to rachis elongation and a
697 downregulation of *VvSPY* in the rachis [70]. In *A. thaliana* GA signaling is initiated through its
698 binding to the GA INSENSITIVE DWARF1 (GID1) receptors. This allows subsequent interaction
699 between GID1 and DELLA proteins (GA INSENSITIVE [GAI], REPRESSOR OF GAI-3 [RGA],
700 RGA-LIKE1 [RGL1], RGL2, and RGL3). DELLA proteins are transcriptional repressors and
701 downregulate GA response genes. In the presence of gibberellin, the stable GID1-GA-DELLA
702 complex is recognized by the SCF^{SLY1} complex which ubiquintylates the DELLA proteins and
703 causes their degradation by the 26S proteasome [71,72]. It has been reported previously that GID1-
704 transcripts are upregulated during bud dormancy in grapevine while transcripts of DELLA are
705 downregulated [73]. Similarly, we found that the GID1B receptor transcript is upregulated during
706 bud dormancy while the DELLA-protein SLR1-like (SLENDER RICE 1 LIKE) are downregulated.
707 This confirms the promoting role of GID1B in plant growth, and the development of lateral

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

708 meristems in dormant buds and indicates that SLR1-like is responsive for the mediation of the
709 suppression of inflorescence development through GA.

710 In our analyses, numerous other genes involved in the repression of floral transition and flower
711 development were found to be upregulated during bud dormancy. *HUA2*-like genes, which play a
712 role in the repression of floral transition [74], are upregulated during bud dormancy in *Vitis*. The
713 *KNOTTED1*-like homeobox gene *BP (BREVIPEDICELLUS)* was found to be upregulated towards
714 grapevine bud burst and inflorescence maturation. In *A. thaliana BP* controls distal pedicel growth
715 and thus inflorescence architecture [75,76]. *ER (ERECTA)* and other *KNAT (KNOTTED-LIKE)*
716 genes, are involved in inflorescence architecture in *A. thaliana* [77,78], were also found to be
717 upregulated towards bud burst, which indicates their function in inflorescence development. Genes
718 for SQUAMOSA promoter-binding proteins, known to be involved in flower development [79],
719 were downregulated during bud dormancy while upregulated during flower formation in grapevine.
720 The *BEL*-like gene (*VvBELa* and *b*) and the *Vitis STM* orthologue *VvSBH1* were also found to be
721 upregulated during bud dormancy. STM and the *A. thaliana* homeobox-gene *BEL1* build a
722 complex, which maintains the indeterminacy of the inflorescence meristem [80].

723 MYC transcription factors *VvbHLH74* and *VvbHLH63* show large variations in gene expression
724 over time with a peak in expression around March when buds are swelling. *CIB1 (cryptochrome-*
725 *interacting basic-helix-loop-helix),* the *A. thaliana* homologue of *VvbHLH63,* plays a role in CRY2
726 (cryptochrome 2)-dependent regulation of flowering time. Cryptochromes (CRY) are blue-light
727 receptors that mediate light response. In yeast and *A. thaliana,* CIB1 interacts with CRY2 when
728 blue light is available. It promotes CRY2-dependent floral initiation together with additional CIB1-
729 related proteins and stimulates *FT* transcription [81]**.** Hence, *VvbHLH74* and *VvbHLH63* might be
730 involved in light dependent floral initiation.

731 *ELF*-like genes as well as a *CONSTANS*-like gene (*VvCOL16*) and *CDF* genes (*CYCLING DOF*
732 *FACTORS*) were upregulated during bud dormancy. DOF proteins delay flowering by repressing
733 *CO* transcription [82]. *ELF3*, *ELF4*, and *TOC1* function in the primary, phytochrome-mediated
734 light-input pathway to the circadian oscillator in *A. thaliana. TOC1* is necessary for light-induced
735 *CCA1* (CIRCADIAN CLOCK ASSOCIATED 1)/ LHY (*LATE ELONGATED HYPOCOTYL*)
736 expression [83]. Mutants of *elf4* show attenuated expression of *CCA1* and early flowering in non-
737 inductive photoperiods, which is probably caused by elevated amounts of *CONSTANS (CO)*, a gene
738 that promotes floral induction [84]. *ELF4* is a flowering pathway gene that may play a key role in
739 signaling processes regulating dormancy induction in grapevine [85].

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

740 *MBD9*, whose *A. thaliana* homologue - *AtMBD9* is related to the inhibition of flowering [86] and
741 suggested to have a role in bud development through an interaction with FLC in *A. thaliana* [85], is
742 upregulated during bud dormancy in grapevine. *VvSPAR2* (*SUPPRESSOR OF PHYA RELATED2*)
743 is upregulated during bud dormancy and downregulated towards inflorescence development. Its
744 homologue in *A. thaliana* represses photomorphogenesis by negatively regulating the transcription
745 factor *HY5 (ELONGATED HYPOCOTYL 5)*, which promotes photomorphogenesis [87,88].

746 # Conclusion

747 Here, we have reported a new workflow for amplicon sequencing including allele phasing in the
748 highly heterozygous species grapevine. Our genetic association study revealed a significant
749 correlation between alleles of selected FTC target genes and flowering time phenotypes within and
750 outside of previously mapped QTL regions for flowering time on chr 1, 4, 14, 17, and 18. The
751 discovery of a correlation between alleles of FTC target genes and the timing of flowering for genes
752 within previously defined QTL regions supports the role of these QTLs in the timing of flowering.
753 The analysis of gene expression kinetics revealed strong changes in expression pattern for many
754 FTC candidate genes over the consecutive developmental stages. A shift between an up- or
755 downregulation in expression mostly occurred between dormant and swelling buds, or toward
756 inflorescence maturation when the young inflorescence structures at the shoots grow out of the buds
757 and become externally visible. These time-dependent expression profiles underline the role of many
758 FTC candidate genes in the control of flowering time. Moreover, many FTC candidate genes were
759 found to be expressed in buds and inflorescences but not in leaves. This tissue specificity further
760 confirms their role in flowering time and floral development.
761 The knowledge of genes and loci that influence flowering time and play a role in early flowering
762 may allow the selection of genotypes not carrying these alleles through grapevine breeding
763 programs. To meet the expected change of climate conditions late flowering cultivars might be
764 better adapted, especially in the present cool climate areas.
765 For future research, grapevine cultivars are to be analysed for alleles of flowering time control
766 genes correlating with early or late flowering in order to further investigate the role of these alleles
767 in the timing of flowering and study epistatic and additive effects between QTL regions influencing
768 the timing of flowering.

769 # Author contributions

770 N.K., L.H. and D.H. concieved and planned the experiments. N.K., I.O., A.S., P.V. designed and
771 performed the experiments. I.O., L.H. and A.S. performed the phenotyping. N.K., I.O. and P.V.
772 carried out the sample preparation. N.K., I.O. and L.H. calculated the data. R.T., B.W., D.H.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

conceived the original idea and supervised the project. All authors interpreted and discussed results. N.K. and D.H. wrote the manuscript with input from all authors.

Acknowledgements

We thank Willy Keller, Johanna Müschner, Thomas Rosleff Sörensen, and Andreas Preiss for the excellent technical support, and Martin Sagasser for editing the manuscript. In addition, the authors wish to thank the members of the SPP1530 team for their support. The research was funded by the German research foundation (DFG) - Project number 172418080/15 - SPP1530.

References

1. Carmona MJ, Chaïb J, Martínez-Zapater JM, Thomas MR (2008) A molecular genetic perspective of reproductive development in grapevine. Journal of Experimental Botany 59: 2579-2596.

2. Duchêne E, Butterlin G, Dumas V, Merdinoglu D (2012) Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages. Theoretical and Applied Genetics 124: 623-635.

3. Mullins MG, Bouquet A, Williams LE (1992) Biology of the Grapevine. Cambridge: Cambridge University Press.

4. Tomasi D, Jones GV, Giust M, Lovat L, Gaiotti F (2011) Grapevine Phenology and Climate Change: Relationships and Trends in the Veneto Region of Italy for 1964-2009. American Journal of Enology and Viticulture 62: 329-339.

5. Zyprian A, Eibach R, Trapp O, Schwander F, Töpfer R (2018) Grapevine breeding under climate change: Applicability of a molecular marker linked to veraison. Vitis 57: 119-123.

6. Rienth M, Torregrosa L, Sarah G, Ardisson M, Brillouet JM, et al. (2016) Temperature desynchronizes sugar and organic acid metabolism in ripening grapevine fruits and remodels their transcriptome. BMC Plant Biology 16: 164.

7. Boss PK, Buckeridge EJ, Poole A, Thomas MR (2003) New insights into grapevine flowering. Functional Plant Biology 30: 593-606.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

800 8. Andrés F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. Nature
801 Reviews Genetics 13: 627-639.

802 9. Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, et al. (2009) Candidate gene association
803 mapping of Arabidopsis flowering time. Genetics 183: 325-335.

804 10. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, et al. (2007) The grapevine genome sequence
805 suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463-467.

806 11. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, et al. (2007) A high quality
807 draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2.

808 12. Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, et al. (2014) A deep survey of
809 alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress
810 condition and genotype. BMC Plant Biology 14: 99.

811 13. Adam-Blondon A-F, Jaillon O, Vezzulli S, Zharkikh A, Troggio M, et al. (2011) Genome
812 Sequence Initiatives. In: Adam-Blondon A-F, Martinez-Zapater JM, Kole C, editors. Genetics,
813 Genomics and Breeding of Grapes.

814 14. Canaguier A, Grimplet J, Di Gaspero G, Scalabrin S, Duchêne E, et al. (2017) A new version of
815 the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). Genomics
816 Data 14: 56-62.

817 15. Carmo Vasconcelos M, Greven M, Winefield CS, Trought MCT, Raw V (2009) The Flowering
818 Process of Vitis vinifera: A Review. American Journal of Enology and Viticulture 60: 411-434.

819 16. Calonje M, Cubas P, Martínez-Zapater JM, Carmona MJ (2004) Floral meristem identity genes
820 are expressed during tendril development in grapevine. Plant Physiology 135: 1491-1501.

821 17. Joly D, Perrin M, Gertz C, Kronenberger J, Demangeat G, et al. (2004) Expression analysis of
822 flowering genes from seedling-stage to vineyard life of grapevine cv. Riesling. Plant Science 166:
823 1427-1436.

824 18. Sreekantan L, Thomas MR (2006) VvFT and VvMADS8, the grapevine homologues of the
825 floral integrators FT and SOC1, have unique expression patterns in grapevine and hasten flowering
826 in Arabidopsis. Functional Plant Biology 33: 1129-1139.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

19. Carmona MJ, Cubas P, Calonje M, Martinez-Zapater JM (2007) Flowering transition in grapevine (Vitis vinifera L.). Canadian Journal of Botany 85: 701-711.

20. Grattapaglia D, Sederoff R (1994) Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121-1137.

21. Cipriani G, Di Gaspero G, Canaguier A, Jusseaume J, Tassin J, et al. (2011) Molecular Linkage Maps: Strategies, Resources and Achievements. In: Adam-Blondon A-F, Martinez-Zapater JM, Kole C, editors. Genetics, Genomics and Breeding of Grapes.

22. Costantini L, Battilana J, Lamaj F, Fanizza G, Grando MS (2008) Berry and phenology-related traits in grapevine (Vitis vinifera L.): from quantitative trait loci to underlying genes. BMC Plant Biology 8.

23. Fechter I, Hausmann L, Zyprian E, Daum M, Holtgräwe D, et al. (2014) QTL analysis of flowering time and ripening traits suggests an impact of a genomic region on linkage group 1 in Vitis. Theoretical and Applied Genetics 127: 1857-1872.

24. Gökbayrak Z, Özer C, Söylemezoglu G (2006) Preliminary Results on Genome Mapping of an Italia x Mercan Grapevine Population. Turk J Agric For 30: 273-280.

25. Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ (2009) Genome-wide analysis of MIKCC-type MADS box genes in grapevine. Plant Physiology 149: 354-369.

26. Aguiar D, Istrail S (2012) HapCompass: a fast cycle basis algorithm for accurate haplotype assembly of sequence data. Journal of Computational Biology 19: 577-590.

27. Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. Nature Reviews Genetics 12: 703-714.

28. Keurentjes JJ, Fu J, De Vos RCH, Lommen A, Hall RD, et al. (2006) The genetics of plant metabolism. Nature Genetics 38: 842-849.

29. Martin M, Patterson M, Garg S, Fischer SO, Pisanti N, et al. (2016) WhatsHap: fast and accurate read-based phasing. Software.

30. Bansal V, Bafna V (2008) HapCUT: an efficient and accurate algorithm for the haplotype assembly problem. Bioinformatics 24: 153-159.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

855 31. Lancia G, Bafna V, Istrail S, Lippert R, Schwartz R (2001) SNPs Problems, Complexity, and
856 Algorithms. ESA: 182-193.

857 32. Zyprian E, Ochßner I, Schwander F, Šimon S, Hausmann L, et al. (2016) Quantitative trait loci
858 affecting pathogen resistance and ripening of grapevines. Molecular Genetics and Genomics 291:
859 1573-1594.

860 33. Lorenz DH, Eichhorn KW, Bleiholder H, Klose R, Meier U, et al. (1995) Phenological growth
861 stages of the grapevine (Vitis vinifera L. ssp. vinifera)-Codes and descrptions according to the
862 extended BBCH scale. Australian Journal of Grape and Wine Research 1: 100-110.

863 34. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. (2009) BLAST+:
864 architecture and applications. BMC Bioinformatics 10: 421.

865 35. Ward N, Moreno-Hagelsieb G (2014) Quickly finding orthologs as reciprocal best hits with
866 BLAT, LAST, and UBLAST: how much do we miss? PLoS ONE 9: e101850.

867 36. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, et al. (2012) Primer3--new
868 capabilities and interfaces. Nucleic Acids Research 40.

869 37. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence
870 data. Bioinformatics 30: 2114-2120.

871 38. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform.
872 Bioinformatics 25: 1754-1760.

873 39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map
874 format and SAMtools. Bioinformatics 25: 2078-2079.

875 40. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome
876 Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.
877 Genome Research 20: 1297-1303.

878 41. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, et al. (2013) From FastQ
879 data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current
880 Protocols in Bioinformatics 11: 1110.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

42. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43: 491-498.

43. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14: R36.

44. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166-169.

45. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550.

46. Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biological Chemistry 379: 633-646.

47. Gehring WJ (1992) The homeobox in perspective. Trends in Biochemical Sciences 17: 277-280.

48. Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in Arabidopsis thaliana. Current Opinion in Plant Biology 4: 447-456.

49. Johanson U, West J, Lister C, Michaels S, Amasino R, et al. (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science 290: 344-347.

50. Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, Benfey PN (1999) The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. The Plant Journal 18: 111-119.

51. Chen X, Zhang Z, Liu D, Zhang K, Li A, et al. (2010) SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development. Journal of Integrative Plant Biology 52: 946-951.

52. Poupin MJ, Federici F, Medina C, Matus JT, Timmermann T, et al. (2007) Isolation of the three grape sub-lineages of B-class MADS-box TM6, PISTILLATA and APETALA3 genes which are differentially expressed during flower and fruit development. Gene 404: 10-24.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

908 53. Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, et al. (2000) A short history of MADS-
909 box genes in plants. Plant Molecular Biology 42: 115-149.

910 54. Grattapaglia D, Bertolucci FL, Sederoff RR (1995) Genetic mapping of QTLs controlling
911 vegetative propagation in Eucalyptus grandis and E. urophylla using a pseudo-testcross strategy and
912 RAPD markers. Theoretical and Applied Genetics 90: 933-947.

913 55. Chen X, Meyerowitz, E M (1999) HUA1 and HUA2 are two members of the floral homeotic
914 AGAMOUS pathway. . Molecular Cell 3: 349-360.

915 56. Boss PK, Thomas MR (2002) Association of dwarfism and floral induction with a grape 'green
916 revolution' mutation. Nature 416: 847-850.

917 57. Causier B, Castillo R, Xue Y, Schwarz-Sommer Z, Davies B (2010) Tracing the evolution of
918 the floral homeotic B- and C-function genes through genome synteny. Molecular Biology and
919 Evolution 27: 2651-2664.

920 58. Kramer EM, Dorit RL, Irish VF (1998) Molecular evolution of genes controlling petal and
921 stamen development: duplication and divergence within the APETALA3 and PISTILLATA
922 MADS-box gene lineages. Genetics 149: 765-783.

923 59. Nakamichi N, Murakami-Kojima M, Sato E, Kishi Y, Yamashino T, et al. (2002) Compilation
924 and characterization of a novel WNK family of protein kinases in Arabiodpsis thaliana with
925 reference to circadian rhythms. Bioscience, Biotechnology, and Biochemistry 66: 2429-2436.

926 60. Hong-Hermesdorf A, Brüx A, Grüber A, Grüber G, Schumacher K (2006) A WNK kinase binds
927 and phosphorylates V-ATPase subunit C. FEBS Letters 580: 932-939.

928 61. Wang Y, Liu K, Liao H, Zhuang C, Ma H, et al. (2008) The plant WNK gene family and
929 regulation of flowering time in Arabidopsis. Plant Biology 10: 548-562.

930 62. Foyer CH, Kerchev PI, Hancock RD (2012) The ABA-INSENSITIVE-4 (ABI4) transcription
931 factor links redox, hormone and sugar signaling pathways. Plant Signaling & Behavior 7: 276-281.

932 63. Shu K, Chen Q, Wu Y, Liu R, Zhang H, et al. (2016) ABSCISIC ACID-INSENSITIVE 4
933 negatively regulates flowering through directly promoting Arabidopsis FLOWERING LOCUS C
934 transcription. Journal of Experimental Botany 67: 195-205.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

64. Heintzen C, Nater M, Apel K, Staiger D (1997) AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 94: 8515-8520.

65. Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, et al. (2009) REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. Proceedings of the National Academy of Sciences of the United States of America 106: 16883-16888.

66. Boden SA, Weiss D, Ross JJ, Davies NW, Trevaskis B, et al. (2014) EARLY FLOWERING3 Regulates Flowering in Spring Barley by Mediating Gibberellin Production and FLOWERING LOCUS T Expression. The Plant Cell 26: 1557-1569.

67. Hall A, Bastow RM, Davis SJ, Hanano S, McWatters HG, et al. (2003) The TIME FOR COFFEE gene maintains the amplitude and timing of Arabidopsis circadian clocks. The Plant Cell 15: 2719-2729.

68. Ding Z, Millar AJ, Davis AM, Davis SJ (2007) TIME FOR COFFEE encodes a nuclear regulator in the Arabidopsis thaliana circadian clock. The Plant Cell 19: 1522-1536.

69. Tseng TS, Salomé PA, McClung CR, Olszewski NE (2004) SPINDLY and GIGANTEA interact and act in Arabidopsis thaliana pathways involved in light responses, flowering, and rhythms in cotyledon movements. The Plant Cell 16: 1550-1563.

70. Upadhyay A, Jogaiah S, Maske SR, Kadoo NY, Gupta VS (2015) Expression of stable reference genes and SPINDLY gene in response to gibberellic acid application at different stages of grapevine development. Biologia Plantarum 59: 436-444.

71. Cheng C, Jiao C, Singer SD, Gao M, Xu X, et al. (2015) Gibberellin-induced changes in the transcriptome of grapevine (Vitis labrusca × V. vinifera) cv. Kyoho flowers. BMC Genomics 16.

72. Fleet CM, Sun TP (2005) A DELLAcate balance: the role of gibberellin in plant morphogenesis. Current Opinion in Plant Biology 8: 77-85.

73. Khalil-Ur-Rehman M, Sun L, Li CX, Faheem M, Wang W, et al. (2017) Comparative RNA-seq based transcriptomic analysis of bud dormancy in grape. BMC Plant Biology 17: 18.

74. Doyle MR, Bizzell CM, Keller MR, Michaels SD, Song J, et al. (2005) HUA2 is required for the expression of floral repressors in Arabidopsis thaliana. The Plant Journal 41: 376-385.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

963 75. Douglas SJ, Riggs CD (2005) Pedicel development in Arabidopsis thaliana: contribution of
964 vascular positioning and the role of the BREVIPEDICELLUS and ERECTA genes. Developmental
965 Biology 284: 451-463.

966 76. Venglat SP, Dumonceaux T, Rozwadowski K, Parnell L, Babic V, et al. (2002) The homeobox
967 gene BREVIPEDICELLUS is a key regulator of inflorescence architecture in Arabidopsis.
968 Proceedings of the National Academy of Sciences of the United States of America 99: 4730-4735.

969 77. Douglas SJ, Chuck G, Dengler RE, Pelecanda L, Riggs CD (2002) KNAT1 and ERECTA
970 Regulate Inflorescence Architecture in Arabidopsis. The Plant Cell 14: 547-558.

971 78. Shpak ED, Berthiaume CT, Hill EJ, Torii KU (2004) Synergistic interaction of three ERECTA-
972 family receptor-like kinases controls Arabidopsis organ growth and flower development by
973 promoting cell proliferation. Development 131: 1491-1501.

974 79. Yamasaki K, Kigawa T, Inoue M, Yamasaki T, Yabuki T, et al. (2006) An Arabidopsis SBP-
975 domain fragment with a disrupted C-terminal zinc-binding site retains its tertiary structure. FEBS
976 Letters 580: 2109-2116.

977 80. Bellaoui M, Pidkowich MS, Samach A, Kushalappa K, Kohalmi SE, et al. (2001) The
978 Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved
979 between plants and animals. The Plant Cell 13: 2455-2470.

980 81. Liu H, Yu X, Li K, Klejnot J, Yang H, et al. (2008) Photoexcited CRY2 interacts with CIB1 to
981 regulate transcription and floral initiation in Arabidopsis. Science 322: 1535-1539.

982 82. Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, Rühl M, et al. (2009) Arabidopsis DOF
983 transcription factors act redundantly to reduce CONSTANS expression and are essential for a
984 photoperiodic flowering response. Developmental Cell 17: 75-86.

985 83. Doyle MR, Davis SJ, Bastow RM, McWatters HG, Kozma-Bognár L, et al. (2002) The ELF4
986 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. Nature 419: 74-77.

987 84. Kikis EA, Khanna R, Quail PH (2005) ELF4 is a phytochrome-regulated component of a
988 negative-feedback loop involving the central oscillator components CCA1 and LHY. The Plant
989 Journal 44: 300-313.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

85. Sreekantan L, Mathiason K, Grimplet J, Schlauch K, Dickerson JA, et al. (2010) Differential floral development and gene expression in grapevines during long and short photoperiods suggests a role for floral genes in dormancy transitioning. Plant Molecular Biology 73: 191-205.

86. Peng M, Cui Y, Bi YM, Rothstein SJ (2006) AtMBD9: a protein with a methyl-CpG-binding domain regulates flowering time and shoot branching in Arabidopsis. The Plant Journal 46: 282-296.

87. Huang X, Ouyang X, Yang P, Lau OS, Chen L, et al. (2013) Conversion from CUL4-based COP1-SPA E3 apparatus to UVR8-COP1-SPA complexes underlies a distinct biochemical function of COP1 under UV-B. Proceedings of the National Academy of Sciences of the United States of America 110: 16669-16674.

88. Srivastava AK, Senapati D, Srivastava A, Chakraborty M, Gangappa SN, et al. (2015) Short Hypocotyl in White Light1 Interacts with Elongated Hypocotyl5 (HY5) and Constitutive Photomorphogenic1 (COP1) and Promotes COP1-Mediated Degradation of HY5 during Arabidopsis Seedling Development. Plant Physiology 169: 2922-2934.

Bibliography

Supplementary Figures:

S1 Figure: Illustration of the reproductive developmental cycle of grapevine showing the stages of flowering and berry development ([according to 1]). UP: uncommitted primordia.

S2 Figure: Flowering periods in days after January 1st in the population GF.GA-47-42 x ‘Villard Blanc’ in the years 1999 and 2010 – 2016.

S3 Figure: Expression profile of the three B-class floral homeotic genes *VvAP3, VvTM6* and *VvPI* over consecutive developmental stages of bud- and inflorescence development in GF.GA-47-42. The last three time points refer to developing stages of visible inflorescence structures.

S1 Table: F_1 individuals of the mapping population Gf.Ga-47-42 x Villard blanc and days until full bloom after January 1st in the years 1999, 2009, and 2011-2016.

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

1017 S2 Table: Median flowering time of the years 1999 and 2010-2016 for amplicon sequenced F_1
1018 individuals.

1019 S3 Table: Nomenclature system for the alleles of genes.

1020 S4 Table: FTC candidate genes in Vitis.

1021 S5 Table: Amplicon sequenced FTC target genes; genomic positions and length of phased intervals.

1022 S6 Table: Molecular marker information.

1023 S7 Table: Correlation analysis with molecular markers for F1 mapping population.

1024 S8 Table: Inflorescence-specific FTC candidate genes.

1025

1026 **Tables:**

1027 Table 1: Dates of flowering periods of the mapping population GF.GA-47-42 x 'Villard Blanc' and
1028 the amount of global radiation at the location of the vineyards (Geilweilerhof) insofar available.

1029 Table 2: Samples collected from grapevine cultivar GF.GA-47-42 for gene expression analysis, the
1030 developmental stage and the corresponding BBCH code.

1031 Table 3: P-values of the correlation between the E1 allele distribution of *WNK6* and *VvTM6* in
1032 relation to different sets of phenotypic data using 35 amplicon sequenced F_1 individuals.

1033 Table 4: Comparison of the expected and observed allele sizes (bp) and segregation patterns of
1034 several FTC target genes.

1035 Table 5: P-values analysis from both the allele phasing workflow and marker analysis.

Data preparation

amplification of amplicons by PCR

sequencing of amplicons on Illumina-MiSeq platform

bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

Preprocessing of data

analysis of read quality (FastQC)

trimming of raw data (trimmomatic 0.36)

Mapping and processing of mappings

map reads to reference (BWA-MEM 0.7.13)

sort and index alignments (Samtools 1.2)

remove duplicated reads (MarkDuplicates - Picard Tools 1.119)

add sample specific readgroups (AddOrReplaceReadGroups - Picard Tool 1.119)

realign reads around InDels (RealignerTargetCreator, IndelRealigner - GATK 3.5)

recalibrate base quality scores (BaseQualityScoreRecalibrator, PrintReads - GATK 3.5)

Allele phasing

variant calling (HaplotypeCaller - GATK 3.6)

merge records of population samples (GenotypeGVCFs - GATK 3.6)

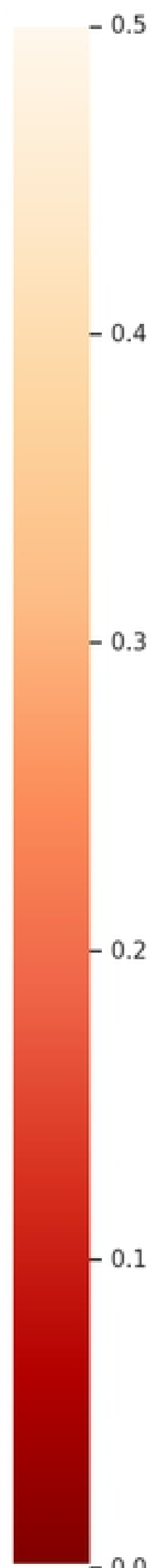
filter variants by quality (SelectVariants, FilterVariant, CombineVariants - GATK 3.6)

physical phasing of alleles (extractHAIRS, HAPCUT - HapCUT 0.7)

create alternate allele sequences (python scripts, Blast2Sam - jvarkit, FastaAlternateReferenceMaker - GATK 3.6)

Median 1999-2016

Chr	Gene	Allele constitution: GF.GA-47-42	Allele constitution: Villard Blanc	Allele counts and phenotype of F1: E1	E2	L1	L2	P-values: E-alleles	P-values: L-alleles
1	VvSEP4	E1/N2	L1/N2	12 (0.435)	13 (0.6)	12 (0.48)	13 (0.69)	0.57	**0.044**
1	VvBS2	E1/E2	L1/L2	17 (0.46)	18 (0.61)	14 (0.41)	21 (0.61)	0.23	**0.019**
1	VvHUA2a	E1/E2	L1/L2	14 (0.41)	12 (0.665)	11 (0.41)	15 (0.66)	0.076	**0.04**
1	VvRAV1b	E0/E0	L1/L2	/	/	5 (0.41)	12 (0.665)	n.d.	**0.035**
1	VvPFT1a	E1/E2	L1/L2	13 (0.41)	14 (0.665)	12 (0.48)	15 (0.66)	0.077	0.12
1	VvCOL10	E1/E2	L1/L2	12 (0.41)	15 (0.67)	12 (0.48)	15 (0.66)	**0.018**	0.097
1	VvGAI1	E1/N	N/L2	15 (0.41)	10 (0.72)	14 (0.41)	11 (0.8)	0.071	**0.003**
1	VvBRIa	E1/E2	L0/L0	15 (0.41)	17 (0.61)	/	/	**0.05**	n.d.
1	VvMBD9a	E1/E2	L1/L2	18 (0.525)	17 (0.6)	19 (0.6)	16 (0.475)	0.45	0.2
1	VvTCP15b	N1/E2	N1/L2	17 (0.6)	16 (0.595)	18 (0.605)	15 (0.49)	0.66	0.66
1	VvCDF2a	E1/E2	L1/L2	17 (0.6)	12 (0.575)	18 (0.605)	14 (0.52)	0.7	0.22
1	VvbHLH49	E1/E2	L1/L2	13 (0.49)	14 (0.63)	12 (0.595)	15 (0.49)	0.44	0.77
1	VvSCL21a	E1/E2	L1/L2	13 (0.49)	13 (0.66)	12 (0.595)	8 (0.45)	**0.049**	0.14
1	VvFLC1	N1/E2	N1/L2	14 (0.41)	19 (0.61)	19 (0.61)	14 (0.475)	0.053	0.49
4	VvTM8	E1/N2	L1/N2	[illegible]	[illegible]	18 (0.605)	13 (0.59)	**0.007**	0.6
4	VvWNK6	E1/N2	L1/N2	25 (0.55)	10 (0.725)	19 (0.61)	16 (0.575)	**0.007**	0.7
5	VvFPA	E1/E2	L1/L2	7 (0.55)	28 (0.6)	20 (0.6)	15 (0.55)	0.54	0.26
6	VvRAP2a	Na/Na	Na/L2	/	/	15 (0.59)	20 (0.6)	n.d.	0.51
6	VvGAMYBc	E1/E2	L1/L2	16 (0.57)	19 (0.61)	27 (0.6)	8 (0.625)	0.69	0.68
7	VvCUL1c	N1/N2	N1/N2	/	/	/	/	n.d.	n.d.
7	VvSPL1b	E1/E2	L1/L2	21 (0.6)	14 (0.55)	35 (0.6)	0	0.46	n.d.
8	VvCDF3b	Nb/Nb	L1/Nb	/	/	17 (0.61)	18 (0.505)	n.d.	0.22
8	ATHB51	Na/Na	Na/L2	/	/	21 (0.56)	14 (0.6)	n.d.	0.74
10	VvGESa	E0/E0	L1/L2	/	/	19 (0.61)	16 (0.555)	n.d.	0.31
11	VvCOL5	E1/E2	L1/L2	15 (0.66)	19 (0.6)	19 (0.59)	15 (0.6)	0.45	0.57
12	VvSPB1	E1/E2	L1/L2	12 (0.58)	22 (0.605)	14 (0.635)	20 (0.575)	0.46	0.44
13	VvPPR37b	E1/E2	L1/L2	14 (0.595)	14 (0.61)	9 (0.6)	19 (0.59)	0.73	0.43
13	VvTOE3	E1/E2	L1/L2	19 (0.56)	15 (0.61)	12 (0.605)	22 (0.57)	0.23	0.94
14	VvGID1Ba	E1/E2	L1/L2	6 (0.735)	17 (0.56)	11 (0.69)	12 (0.575)	0.13	0.33
14	VvFLKa	E1/E2	L1/L2	7 (0.67)	19 (0.49)	15 (0.6)	11 (0.55)	0.069	0.39
14	VvGAIb	E1/E2	L1/L2	12 (0.665)	16 (0.41)	13 (0.69)	15 (0.46)	**0.027**	0.059
14	VvFUL-L	E1/E2	L0/L0	12 (0.725)	13 (0.41)	/	/	**0.003**	n.d.
14	VvSEP1	N1/E2	N1/L2	12 (0.725)	11 (0.41)	13 (0.69)	10 (0.435)	**0.034**	0.11
14	VvFLC2	E1/E2	L1/L2	14 (0.635)	14 (0.4)	15 (0.6)	13 (0.46)	**0.024**	0.68
15	VvAML3	E0/E0	L1/L2	/	/	16 (0.6)	18 (0.55)	n.d.	0.31
16	VvEMF2a	N1/E2	N1/L2	14 (0.635)	22 (0.61)	16 (0.6)	12 (0.61)	0.6	0.69
17	VvFL	N0/N0	N0/N0	/	/	/	/	n.d.	n.d.
17	VvMFTa	E1/E2	L1/L2	13 (0.61)	21 (0.55)	20 (0.595)	14 (0.58)	0.17	0.78
17	VvTOC1	N1/E2	N1/L2	13 (0.61)	19 (0.55)	20 (0.595)	22 (0.6)	0.15	0.52
17	VvWERb	N1/E2	N1/L2	13 (0.61)	17 (0.6)	20 (0.595)	14 (0.52)	0.28	0.27
17	VvGAIa	E1/E2	L1/L2	16 (0.6)	18 (0.555)	14 (0.575)	19 (0.6)	0.68	0.82
18	VvSVP2	E1/E2	L1/L2	17 (0.49)	17 (0.6)	15 (0.41)	19 (0.61)	**0.05**	0.064
18	VvCSTF64	E1/N2	L1/N2	16 (0.6)	9 (0.56)	10 (0.435)	15 (0.66)	0.32	**0.023**
Un	VvMFTb	E1/N2	L1/N2	1 (0.84)	23 (0.59)	15 (0.46)	9 (0.69)	0.19	0.074
Un	VvEMF2e	N1/E2	N1/L2	1 (0.84)	25 (0.6)	15 (0.46)	13 (0.6)	0.44	0.47



bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.

F1-individuals: 35

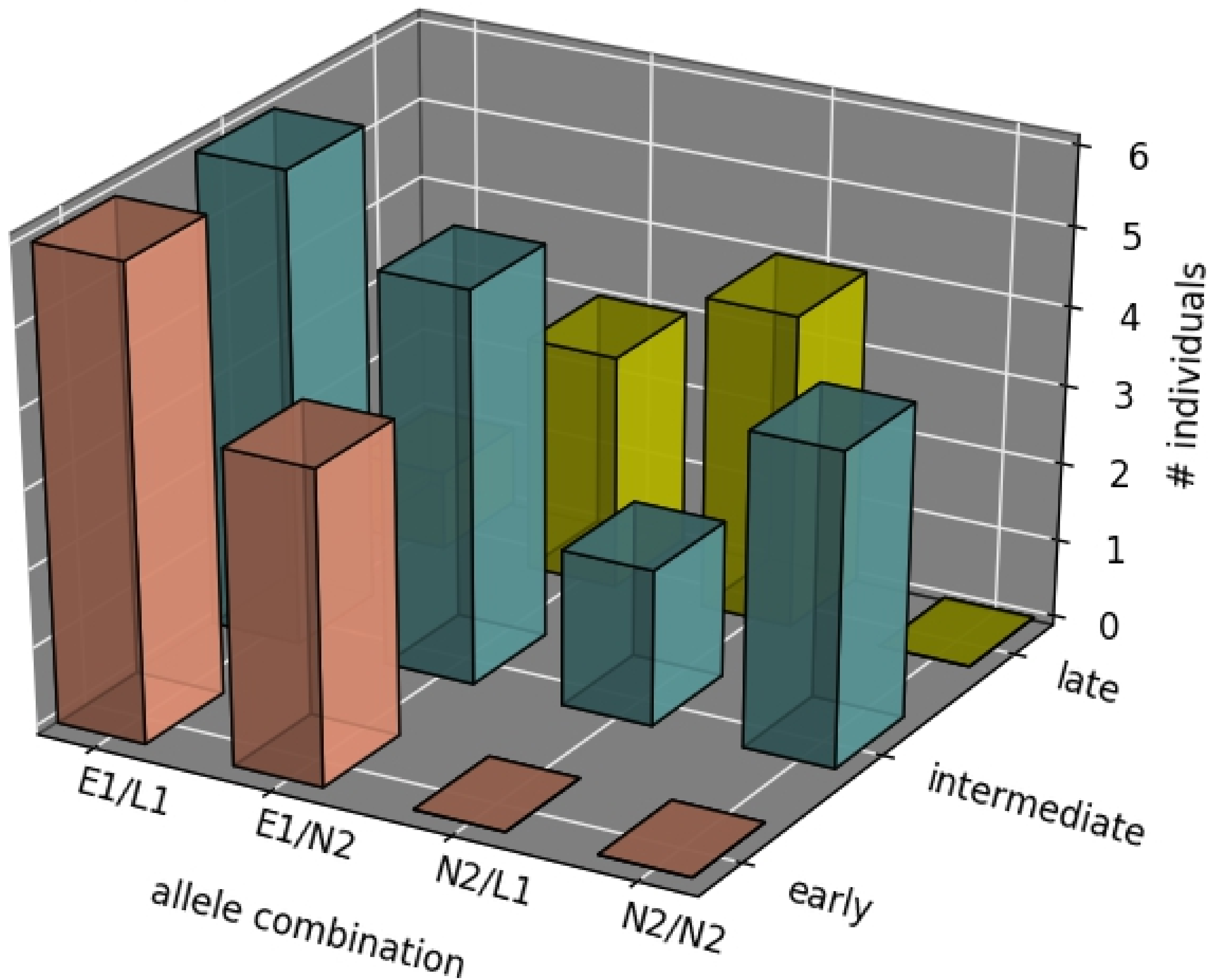
E1/L1:13, E1/N2:12, N2/L1:6, N2/N2:4

E1:25, L1:19, N2:26

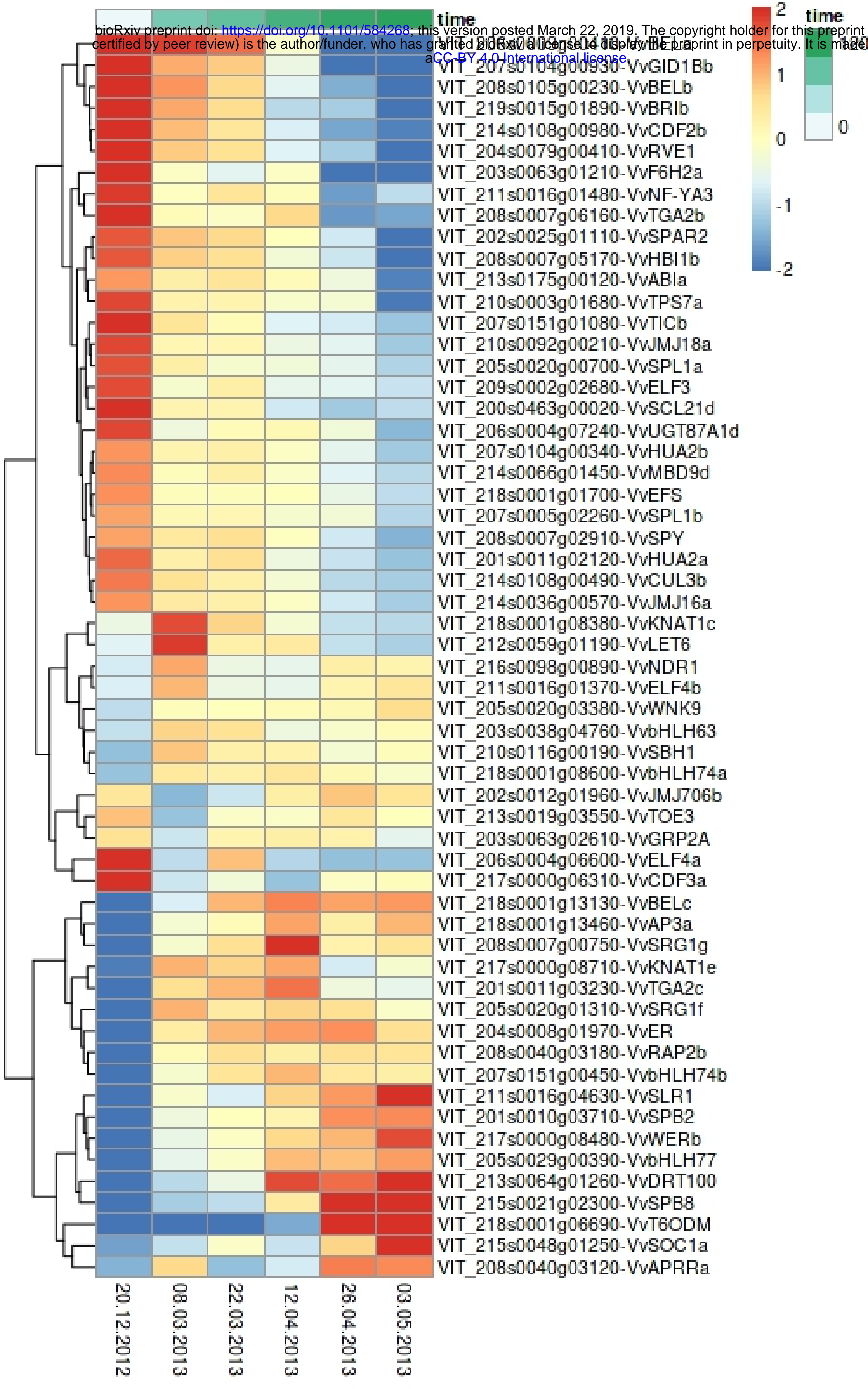
early:10

intermediate:17

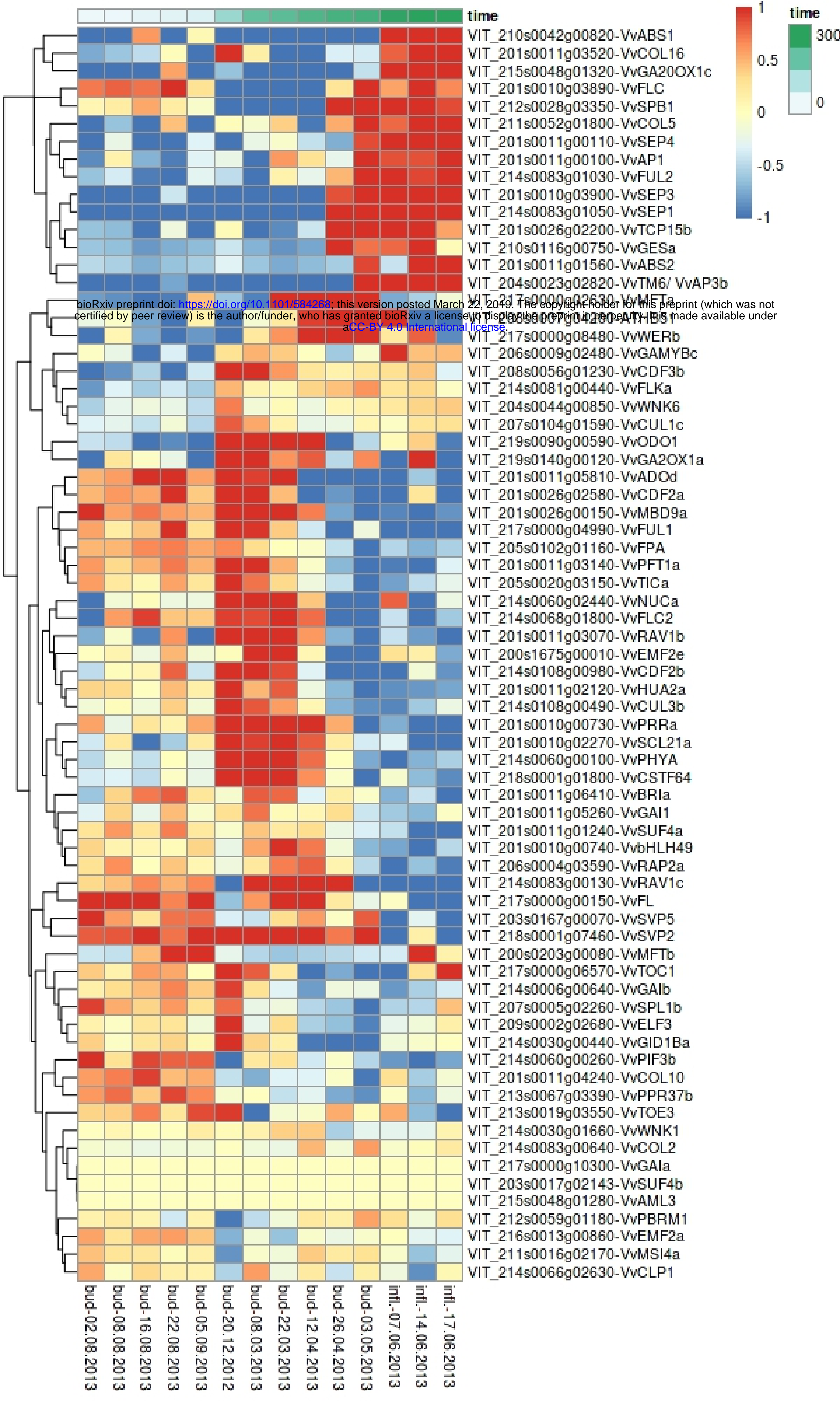
late:8

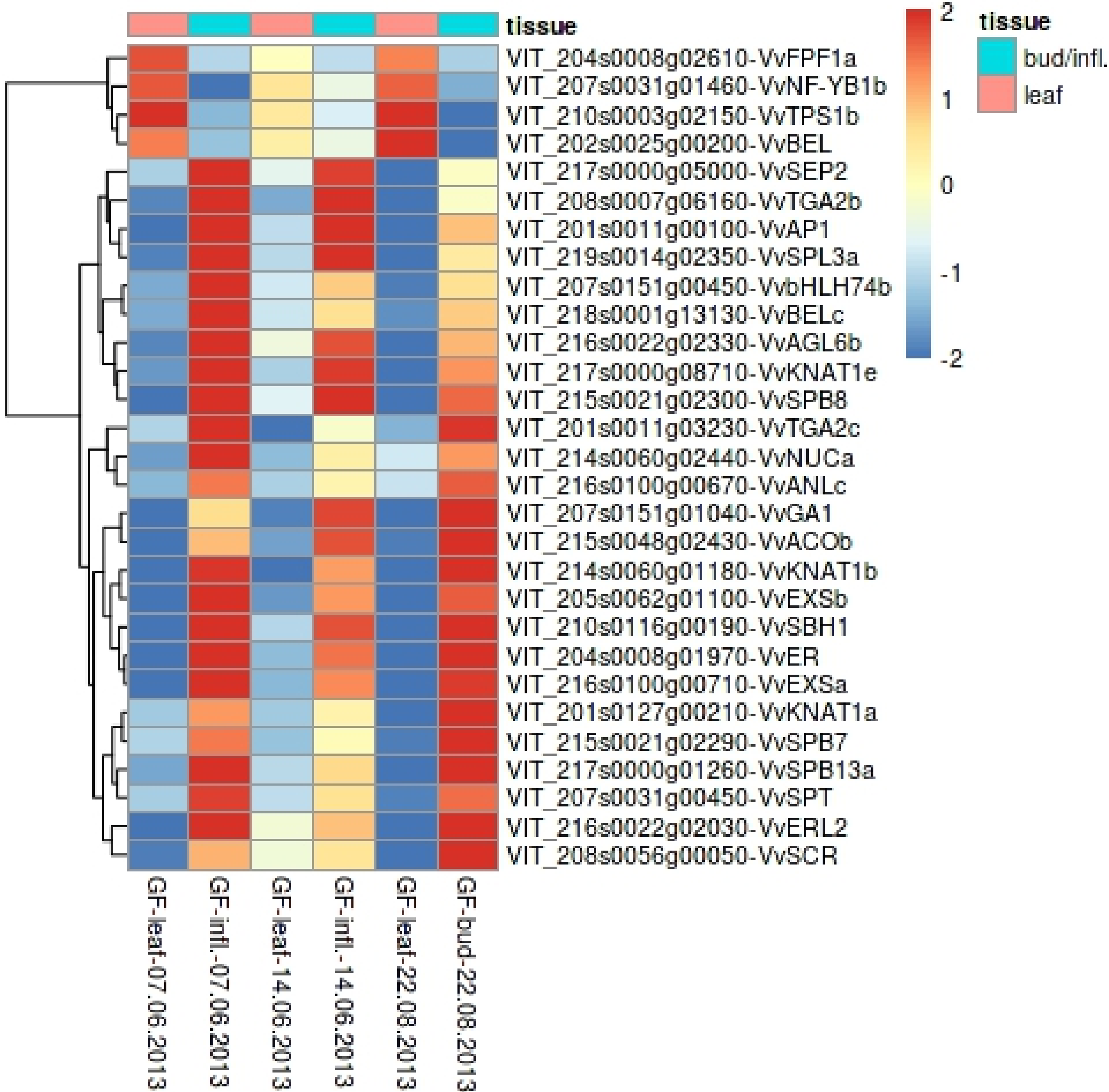


bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/584268; this version posted March 22, 2019. 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 4.0 International license.




tissue
VIT_204s0008g02610-VvFPF1a
VIT_207s0031g01460-VvNF-YB1b
VIT_210s0003g02150-VvTPS1b
VIT_202s0025g00200-VvBEL
VIT_217s0000g05000-VvSEP2
VIT_208s0007g06160-VvTGA2b
VIT_201s0011g00100-VvAP1
VIT_219s0014g02350-VvSPL3a
VIT_207s0151g00450-VvbHLH74b
VIT_218s0001g13130-VvBELc
VIT_216s0022g02330-VvAGL6b
VIT_217s0000g08710-VvKNAT1e
VIT_215s0021g02300-VvSPB8
VIT_201s0011g03230-VvTGA2c
VIT_214s0060g02440-VvNUCa
VIT_216s0100g00670-VvANLc
VIT_207s0151g01040-VvGA1
VIT_215s0048g02430-VvACOb
VIT_214s0060g01180-VvKNAT1b
VIT_205s0062g01100-VvEXSb
VIT_210s0116g00190-VvSBH1
VIT_204s0008g01970-VvER
VIT_216s0100g00710-VvEXSa
VIT_201s0127g00210-VvKNAT1a
VIT_215s0021g02290-VvSPB7
VIT_217s0000g01260-VvSPB13a
VIT_207s0031g00450-VvSPT
VIT_216s0022g02030-VvERL2
VIT_208s0056g00050-VvSCR
GF-leaf-07.06.2013
GF-infl.-07.06.2013
GF-leaf-14.06.2013
GF-infl.-14.06.2013
GF-leaf-22.08.2013
GF-bud-22.08.2013
2
1
0
-1
-2
tissue
bud/infl.
leaf