Towards the definition of a detailed transcriptomic map of berry development

Marianna Fasoli[1,*], *Chandra L.* Richter[1], *Sara* Zenoni[2], *Marco* Sandri[2], *Paola* Zuccolotto[3], *Silvia* Dal Santo[2], *Mario* Pezzotti[2], *Nick* Dokoozlian[1], and *Giovanni Battista* Tornielli[2]

[1]E&J Gallo Winery, Modesto, CA 95353, USA
[2]Department of Biotechnology, University of Verona, 37134 Verona, Italy
[3]Big & Open Data Innovation Laboratory, University of Brescia, 25123 Brescia, Italy

Abstract. The progress of the grapevine genomics and the development of high-throughput technologies for gene expression analysis stimulated the investigation of the physical, biochemical and physiological changes of grape berry growth and maturation at transcriptomic level. The molecular information generated in the last decade is however still fragmented since it relies upon detailed analysis of few stages and thus lacks continuity over grape development. To identify the molecular events associated with berry development at a higher temporal resolution and define a transcriptomic map, we performed RNA-seq analysis of berry samples collected every week from fruit-set to maturity in Pinot noir and Cabernet Sauvignon for three consecutive years, resulting in 219 samples. Using the most variable portion of the transcriptome, we built a preliminary transcriptomic model of berry development based on the Cabernet Sauvignon samples. The Pinot noir samples were then aligned onto this preliminary ripening map to investigate its performance in describing the development of another grape variety. A further step for testing the model was the projection of RNA-seq samples of fruit development of five red-skin Italian cultivars. For all these surveys, the transcriptomic route allowed a precise definition of the progression of berry development during both formation and ripening phases.

1 Introduction

The identification of growth stages of grape berries is critical to viticulture research and vineyard operations. The growth stage of the grape berry is normally defined by visual traits that refer to grapevine phenological schemes or, particularly during ripening phase, by means of easily measurable compositional parameters like sugar concentration in juice. The Modified E-L and the Extended BBCH systems [1, 2] are the phenology scale systems most adopted by viticulturists. These systems describe grape berry development from fruitset to maturity and number the main developmental stages by increasing order. The use of phenological scales is extremely helpful to describe and compare the timing of grape development and ripening, which can vary greatly by vintage and/or cultivation site. However, although some stages are well defined and clearly mark developmental events from a physiological point of view (e.g. fruitset, veraison), defining a comparable developmental stage for grapes of the same cultivar when grown in different conditions or for grapes of different varieties can very likely generate mistakes. For example, the stage of 'pea size' or of 'beginning of bunch closure' (berries touching), respectively reported as stages 31 and 32 in the Modified E-L scale, may correspond to different developmental stages for berries of different grape varieties. The definition of the berry developmental stage becomes even more complicated after the onset of ripening. The ripening degree is normally described by the sugar content that represents a major technological parameter for winemaking. However, it is well known that this parameter is highly influenced by several factors like climate, water availability, canopy management practices and crop load. The uncoupling of sugar concentration from other ripening parameters such as berry acidity or skin anthocyanin content have been reported [3,4], evidencing that sugar content may not solely denote the physiological stage of berry ripening progression.

In the last years the application of genomic tools to the analysis of gene expression during grape berry development has generated a huge amount of transcriptomic data from different varieties and growing conditions. These data have been fundamental to understand the molecular basis of crucial developmental and metabolic rearrangements occurring in grape berries during its formation and ripening [5, 6]. It is now clear that some of the transcriptomic changes occurring during development are widely dependent on the cultivar and/or on the different growing conditions. However, the variations of a portion of the berry transcriptome (the

---

[*] Corresponding author: marianna.fasoli@ejgallo.com

core transcriptome) seem to be conserved across cultivars and growing condition of grapevines, and thus may be used to describe the developmental stage of berry development [6-8].

In this work we explore the possibility of using the transcriptomic data generated from grape berries sampled weekly from Cabernet Sauvignon and Pinot noir vines grown in the same location over three consecutive vintages to map the development of the grape berry. We used the most variable portion of the transcriptome to build a preliminary transcriptomic model of berry development, which allowed a precisely defined progression of berry development during both formation and ripening phases.

2 Results and discussion

2.1 Berry development and ripening

Berry samples were collected from fruit set to full maturity from Cabernet Sauvignon and Pinot noir vines grown in the same location over three consecutive vintages. Samples were collected every 7–10 days using a randomized block approach to account for field variability. This resulted in 219 samples. The recording of heat accumulation (growing degree days) showed that the 2013 and 2014 seasons were warmer than 2012 during March to August (data not shown). Veraison therefore occurred 10–20 days earlier in Pinot noir and 4–11 days earlier in Cabernet Sauvignon. Total Soluble Solids (TSS by °Brix) were measured from the sampling time point following 50% veraison to harvest (target °Brix value for commercial harvest was 24.5). Pinot noir berries accumulated sugar more rapidly than Cabernet Sauvignon berries, resulting in a shorter development time and an earlier harvest in all three vintages (Table 1). In 2014 the estimation of veraison stage may have been slightly anticipated and veraison samples (time point 3) likely collected at an actual earlier stage for both varieties compared to the previous vintages, as suggested by the lower sugar content of the first post-veraison samples (time point 4). Later maturation stages occasionally showed TSS fluctuation, in particular for Cabernet Sauvignon in 2012, which made harvest prediction particularly challenging. At this developmental phase, sugar accumulation is considerably slowed down, and the irrigation schedule is the most responsible for the evidenced variation in TSS. Alternation of irrigation and dry days is an approach used to manage the harvest schedule of a vineyard block and adjust the TSS to the desired value. This is a clear example suggesting that sugar content may not be a suitable parameter—when used alone—to determine the stage of grape berry development during the ripening period.

2.2 Creation of the transcriptomic model

RNA-Seq was used to monitor the expression of all grapevine genes (http://srs.ebi.ac.uk). Cabernet Sauvignon showed a longer development time and thus the preliminary transcriptomic grape ripening model was developed on this variety. To focus on the most meaningful genes, 9,819 genes were selected over non-

or poorly-expressed genes across the Cabernet Sauvignon 120 samples.

Principal component analysis (PCA) was used to evaluate changes in gene expression during berry development, using the average expression value of the sample triplicates (Figure 1).

Table 1. Sample collection details.

Variety	Year	Sampling Time Point	Day Of the Year (DOY)	Total Soluble Solids (°Brix)
Cabernet Sauvignon	2012	0	172	
		1	181	
		2	191	
		3	200	
		4	**209**	
		5	220	12.8
		6	230	16.0
		7	241	19.1
		8	251	19.5
		9	262	23.3
		10	272	21.5
		11	283	27.4
		12	289	24.6
	2013	0	178	
		1	184	
		2	191	
		3	198	
		4	**204**	
		5	212	12.0
		6	219	15.7
		7	226	17.8
		8	233	21.1
		9	240	22.5
		10	247	23.8
		11	253	24.0
		12	260	24.3
		13	267	27.1
	2014	0	174	
		1	183	
		2	190	
		3	**197**	
		4	205	7.7
		5	211	13.3
		6	218	15.2
		7	225	17.5
		8	232	21.2
		9	239	20.9
		10	246	23.2
		11	253	23.3
		12	260	25.0
Pinot noir	2012	0	172	
		1	180	
		2	190	
		3	199	
		4	**208**	
		5	219	14.7
		6	229	16.2
		7	240	20.6
		8	250	26.5
		9	261	26.7
	2013	0	176	
		1	182	
		2	189	
		3	**196**	
		4	203	11.9
		5	210	15.6
		6	217	17.7
		7	224	18.9
		8	231	22.0
		9	238	22.9
		10	246	28.8
	2014	0	168	
		1	174	
		2	181	
		3	**188**	
		4	195	8.4
		5	202	12.2
		6	209	14.4
		7	216	18.5
		8	223	19.9
		9	230	23.9
		10	237	25.7
		11	240	25.6

* bold font corresponds to 50% veraison

The first principal component (Dim 1, 66.8%) separated samples according to extent of development and/or time. The second principal component (Dim 2, 19.3%) added resolution in separating samples from time points 4 to 13. The third component (Dim 3, 5.6 %) distinguished early time point samples (from 0 to 3). The PCA did not reveal sample separation by vintage.
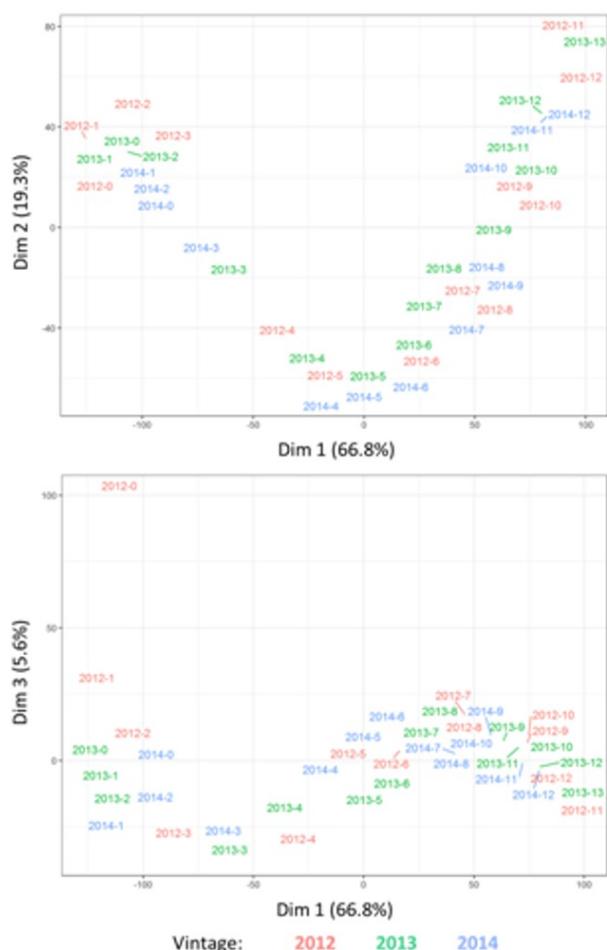
Fig. 1. Principal component analysis showing the distribution of gene expression in Cabernet Sauvignon berries for all three vintages. Dim 1 and Dim 2 explain 86.1% of the variance (top plot). Dim 1 and Dim 3 explain 72.4% of the variance (bottom plot). Samples represent averaged biological triplicates.

The investigation of the genes explaining each principal component revealed different specific expression trends: gradual upward or downward profiles over time for genes correlated with the first component; up or downregulation peak at mid-development for genes associated with the second; while genes correlated to the third component evidenced expression differences at the first sampling time points (data not shown).

The transcriptomic route of the Cabernet Sauvignon berry during its development could be summarized by the two plots showed in Figure 1. Pinot noir RNA-seq triplicates were averaged and then projected onto the Cabernet Sauvignon transcriptomic maps (Dim 1-Dim 2, Dim 1-Dim 3; Figure 2).
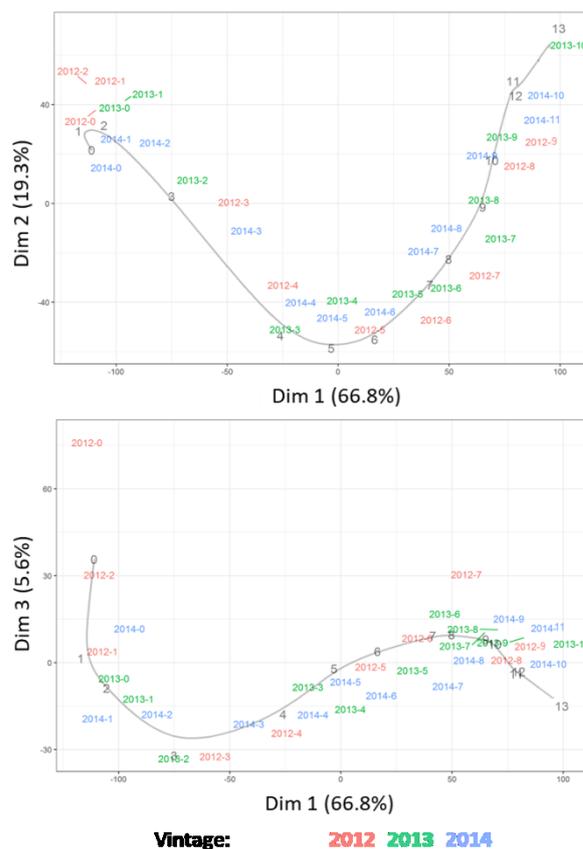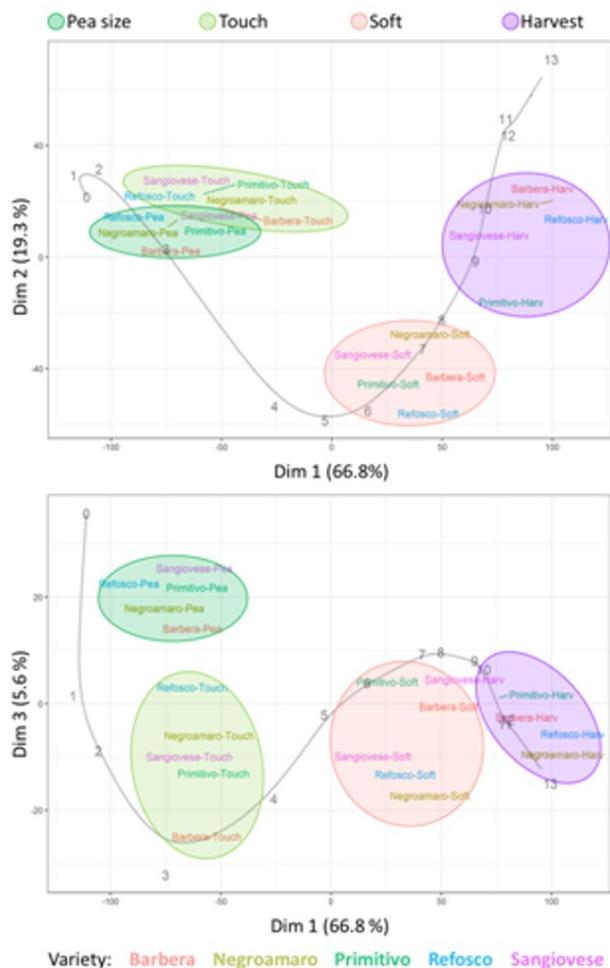

Fig. 2. Principal component analysis showing the distribution of gene expression in Pinot noir berries (all three vintages) onto the Cabernet Sauvignon transcriptional map (Dim 1-Dim 2, top plot; Dim 1-Dim 3, bottom plot). The transcriptomic route modeled on Cabernet Sauvignon berry development is represented by gray line and numbers. Pinot noir samples represent averaged biological triplicates.

The alignment of the Pinot noir samples on the Cabernet Sauvignon transcriptomic route evidenced the suitability of the model in comparing grape developmental data of different varieties. The Dim 1-Dim 2 plot highlighted that the Pinot noir samples distribution matched over the entire route built with the Cabernet Sauvignon dataset despite the two varieties featuring a different berry development timing as described before (Table 1). The Dim 1-Dim 3 plot pointed out variations among the first sampling time points, evidencing some vintage variability, probably due to the timing of sampling.

2.3 Model performance with five red-skin Italian varieties

To verify the performance of the Cabernet Sauvignon transcriptomic model for other varieties and different growing regions, we used a RNA-Seq dataset describing berry development transcriptome profiles of five grapevine red-berry varieties (Sangiovese, Barbera, Negroamaro, Refosco, and Primitivo) cultivated during 2011 in the same experimental vineyard (Conegliano, Veneto region, Italy) [6]. This dataset was generated from berries collected at four phenological stages: pea-size (BBCH 75) at 20 days after flowering (Pea), beginning to touch (BBCH 77) just prior to veraison

(Touch), softening (BBCH 85) at the end of veraison (Soft), and ripe (BBCH 89) (Harv). After averaging the biological triplicates, the twenty RNA-seq averaged samples were plotted onto the Cabernet Sauvignon transcriptomic maps (Dim 1-Dim 2, Dim 1-Dim 3; Figure 3).

Fig. 3. Principal component analysis showing the distribution of gene expression in five red-skin Italian varieties during berry ripening onto the Cabernet Sauvignon transcriptional map (Dim 1-Dim 2, top plot; Dim 1-Dim 3, bottom plot). The transcriptomic route modeled on Cabernet Sauvignon berry development is represented by gray line and numbers. Berry samples of Barbera, Negroamaro, Primitivo, Refosco, and Sangiovese represent averaged biological triplicates, as reported by Massonnet and co-authors [6]. Stages of berry development and ripening: Pea = pea size berries (BBCH 75); Touch = beginning of bunch closure (BBCH 77); Soft = berry softening (BBCH 85); Harvest = ripe berries (BBCH 89).

The five-variety samples aligned on the Cabernet Sauvignon maps and clustered by phenological stage rather than cultivar, as has been previously shown by Massonnet et al [6]. As expected, the Dim 1-Dim 2 plot separated the green-phase and the maturation stages, and Soft and Harvest stages from each other (Figure 3, top plot). In particular, at berry Soft stage the five varieties lay between time point 6 and 8 of the Cabernet Sauvignon transcriptomic route, whereas at harvest between 9 and 10. This alignment seemed to mirror the grapes maturation degree in terms of TSS accumulated

by each variety at the specific stage. Indeed, the berry samples collected from the Italian varieties never reached TSS content as high as the Cabernet Sauvignon grapes (24.5 Brix or higher). Barbera, Negroamaro and Refosco were harvested at ~23 Brix, which indeed corresponded, on average, to time point 10 in Cabernet Sauvignon. Samples of Sangiovese and Primitivo were, instead, collected at lower sugar levels, and in fact were slightly separated and aligned closer to time point 9. The Dim 1-Dim 3 plot confirmed the potentiality of resolving the phenological stages of the pre-veraison phase (Figure 3, bottom plot). Pea-size and Touch berry samples well separated along Dim 3, also suggesting some variability by cultivar at the beginning-to-touch stage. Interestingly, while Soft samples confirmed their position along the model, the Harvest ones showed a shift towards the later stages, distributing over time point 10 to 13. This could correspond to a specific transcriptomic state of the berry at late ripening stages that is independent on the TSS level.

3 Conclusions

In this work we used the transcriptomic changes occurring during the entire development of grape berry (cv Cabernet Sauvignon) to define a model of "molecular phenology" that can be useful to map the ontogenetic development of the fruit with high precision. We showed that this approach could be successful to align the developmental stage of berries of the same cultivar sampled in different vintages and of different cultivars characterized by different phenology and/or grown in different sites. To facilitate its use in viticultural research, the model performances will be evaluated for the smallest number of key genes that, with their expression changes during the season, can be sufficient to precisely define the developmental stage of the grape berries.

4 Materials and Methods

4.1 Vineyard features and plant material growth conditions

Vitis vinifera cultivars Cabernet Sauvignon (clone FPS 8 grafted on 5C rootstock and planted in 1997) and Pinot noir (clone FPS 23 grafted on Freedom rootstock and planted in 2001) were used in this study. Both varieties were grown in sandy clay loam in an east–west orientation with 10-foot row spacing and 5-foot vine spacing.

4.2 Sampling strategy

Berries were collected at 10-day intervals in 2012, and weekly in 2013 and 2014, beginning at fruit-set and continuing until harvest (24.5 °Brix). Veraison was defined as 50% colored berries per cluster. All samples were collected at the same time of day (8:00 am) in randomized block designs for each cultivar: eight-vine blocks for Pinot noir and six-vine blocks for Cabernet Sauvignon. The block designs were replicated along three rows for each cultivar, to allow the collection of biological triplicates. We therefore collected 219 samples in total: 120 for Cabernet Sauvignon (39, 42 and

39 during vintages 2012, 2013 and 2014, respectively) and 99 for Pinot noir (30, 33 and 36 during vintages 2012, 2013 and 2014, respectively). Each sample replicate comprised 26 clusters of berries from each vine block.

4.3 RNA extraction

Sixty berries, from six isolated clusters randomly selected from the vine blocks, were ground under liquid nitrogen. Seeds were removed before grinding. Frozen powder was divided into 400-mg aliquots for RNA extraction. Total RNA was isolated using the Spectrum Plant Total RNA kit (Sigma-Aldrich, St Louis, MO, USA) following the manufacturer's instructions with some modifications [5]. RNA quality and quantity were determined using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and a Bioanalyzer Chip RNA 7500 series II (Agilent, Santa Clara, CA, USA).

4.4 Library preparation and RNA sequencing analysis

We prepared 219 non-directional cDNA libraries from 2.5 μg total RNA using the Illumina TruSeq RNA Sample preparation protocol (Illumina Inc., San Diego, CA, USA). Library quality was determined using the Agilent High Sensitivity DNA kit on the Agilent 2100 Bioanalyzer, and the quantity was determined by quantitative PCR using the KAPA Library Quantification kit (Kapa Biosystems, Roche Diagnostics, Basel, Switzerland). Single-end reads of 100 nucleotides were obtained using an Illumina HiSeq 1000 sequencer and sequencing data were generated using the base-calling software Illumina Casava v1.8 ($32,027,722 \pm 7,628,415$ per sample). The reads were aligned to the grapevine 12x reference genome PN40024 [9] using TopHat v2.0.6 with default parameters [10]. An average of 86.7% of reads was mapped for each sample. Mapped reads were used to reconstruct the transcripts using Cufflinks v2.0.2 [11] and the reference genome annotation V1 (http://genomes.cribi.unipd.it/DATA). The normalized expression of each transcript was calculated as RPKM (reads per kilobase of transcript per million mapped reads) for each sample.

4.5 Multivariate analysis and model creation

PCA was carried out using R Studio software (https://www.rstudio.com/) and the Spearman statistical metric was chosen to create the correlation matrix.

To focus on the most meaningful genes, filtering steps were applied to the Cabernet Sauvignon and Pinot noir dataset, and the red-skin Italian varieties survey [6]. In addition to eliminating genes with no expression across all samples, we targeted and removed genes showing low expression (RPKM < 1; [12]) at each time point except one in at least one year and one cultivar. A second filter was applied to remove genes showing low variability (almost constant expression) across the experimental conditions. A final polish aimed at removing genes with overall low expression level (RPKM < 1) and random isolated peaks considered expression outliers. This multistep approach allowed concentrating the creation of the transcriptomic model on 9,819 meaningful genes.

References

1. B.G. Coombe, Australian Journal of Grape and Wine Research **1**, 2 (1995)

2. D.H. Lorenz, K.W. Eichhorn, H. Bleiholder, R. Klose, U. Meier, and E. Weber, Australian Journal of Grape and Wine Research **1**, 2 (1995)

3. V.O. Sadras and M.A. Moran, Australian Journal of Grape and Wine Research **18**, 2 (2012)

4. N. Bobeica, S. Poni, G. Hilbert, C. Renaud, E. Gomes, S. Delrot, and Z. Dai, Front Plant Sci **6**, (2015)

5. M. Fasoli, S. Dal Santo, S. Zenoni, G.B. Tornielli, L. Farina, A. Zamboni, A. Porceddu, L. Venturini, M. Bicego, V. Murino, A. Ferrarini, M. Delledonne, and M. Pezzotti, Plant Cell **24**, 9 (2012)

6. M. Massonnet, M. Fasoli, G.B. Tornielli, M. Altieri, M. Sandri, P. Zuccolotto, P. Paci, M. Gardiman, S. Zenoni, and M. Pezzotti, Plant Physiol **174**, 4 (2017)

7. S. Dal Santo, G.B. Tornielli, S. Zenoni, M. Fasoli, L. Farina, A. Anesi, F. Guzzo, M. Delledonne, and M. Pezzotti, Genome Biol **14**, 6 (2013)

8. S. Dal Santo, et al., Plant J **93**, 6 (2018)

9. O. Jaillon, et al., Nature **449**, 7161 (2007)

10. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg, Genome Biol **14**, 4 (2013)

11. A. Roberts, H. Pimentel, C. Trapnell, and L. Pachter, Bioinformatics **27**, 17 (2011)

12. A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, and B. Wold, Nat Methods **5**, 7 (2008)