Model-assisted analysis for tuning anthocyanin composition in grape berries

Yongjian Wang[1,2], Boxing Shang[1,2,3], Michel Génard[4], Ghislaine Hilbert-Masson[5], Serge Delrot[5], Eric Gomès[5], Stefano Poni[6], Markus Keller[7], Christel Renaud[5], Junhua Kong[1,2], Jinliang Chen[8], Zhenchang Liang[1, 2,3], Zhanwu Dai[1, 2,3,*]

[1]State Key Laboratory of Plant Diversity and Specialty Crops and Beijing Key Laboratory of Grape Science and Enology, Institute of Botany, the Chinese Academy of Sciences, Beijing, 100093, China; [2]China National Botanical Garden, Beijing 100093, China; [3]University of Chinese Academy of Sciences, Beijing 100049, China; [4]INRAE, UR1115, Unité Plantes et Systèmes de Culture Horticoles, Avignon, France; [5]EGFV, University of Bordeaux, Bordeaux-Sciences Agro, INRAE, ISVV, Villenave d'Ornon, France; [6]Department of Sustainable Crop Production, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; [7]Department of Viticulture and Enology, Irrigated Agriculture Research and Extension Center, Washington State University, Prosser, WA, USA; [8]Center for Agricultural Water Research in China, China Agricultural University, Beijing, 100083, China

**For correspondence: zhanwu.dai@ibcas.ac.cn*

ABSTRACT

Anthocyanin composition is responsible for the red color of grape berries and wines, and contributes to their organoleptic quality. However, anthocyanin biosynthesis is under genetic, developmental and environmental regulation, making its targeted fine-tuning challenging. We constructed a mechanistic model to simulate the dynamics of anthocyanin composition throughout grape ripening in *Vitis vinifera* L., employing a consensus anthocyanin biosynthesis pathway. The model was calibrated and validated using 6 datasets from 8 cultivars and 37 growth conditions. Tuning the transformation and degradation parameters allowed us to accurately simulate the accumulation process of each individual anthocyanin under different environmental conditions. The model parameters were robust across environments for each genotype. The coefficients of determination (R^2) for the simulated versus observed values for the 6 datasets ranged from 0.92 to 0.99, while the relative root mean square errors (RRMSEs) were between 16.8% and 42.1%. The leave-one-out cross-validation for 3 datasets showed R^2 values of 0.99, 0.96, and 0.91, and RRMSE values of 28.8%, 32.9%, and 26.4%, respectively, suggesting a high prediction quality of the model. Model analysis showed that the anthocyanin profiles of diverse genotypes are relatively stable in response to parameter perturbations. Virtual experiments further suggested that targeted anthocyanin profiles may be reached by manipulating a minimum of 3 parameters, in a genotype-dependent manner. This model presents a promising methodology for characterizing the temporal progression of anthocyanin composition, while also offering a logical foundation for bioengineering endeavors focused on precisely adjusting the anthocyanin composition of grapes.

Keywords:

anthocyanin profile, mechanistic model, metabolic pathways, biochemical decoration

INTRODUCTION

Anthocyanins are specialized metabolites essential for the coloration of plant organs, particularly fruits and flowers, whose color is an important quality criterion for their visual appeal and market value (Zhang *et al.*, 2014). Moreover, anthocyanins possess health-promoting properties, including antioxidant, anti-inflammatory, and anti-cancer activities, making them a subject of interest in medicinal and nutraceutical research (Khoo *et al.*, 2017). More than 600 anthocyanins have been identified to date, and they are synthesized from the precursor amino acid phenylalanine through the phenylpropanoid pathway (Zhang *et al.*, 2014). This considerable diversity in anthocyanins is largely due to the biochemical decorations of a common chemical backbone via hydroxylation, methylation, glycosylation, and acylation (Jaakola, 2013). These biochemical decorations have important influences on stability, bioavailability, and color hues, forming a nexus of structure and functionality for anthocyanins (Saito *et al.*, 2013; Wen *et al.*, 2020; Houghton *et al.*, 2021). Therefore, developing plants with targeted anthocyanin profiles is of great interest for both scientific and application prospects.

The anthocyanin profile varies greatly among plant species and cultivars, mainly as a result of the activities and specificities of diverse decorating enzymes involved in the biosynthesis pathway, ultimately giving rise to a broad range of color hues and patterns (Jaakola, 2013). In grape (*Vitis vinifera*) genotypes, anthocyanins include 3-monoglucosides, which can be further acylated, of 5 anthocyanidins (Mattivi *et al.*, 2006; He *et al.*, 2010), namely the delphinidin (Dp), cyanidin (Cy), petunidin (Pt), peonidin (Pn), and malvidin (Mv), while pelargonidin is rarely present because of the substrate specificity of the enzyme DFR (Dihydro Flavonol 4-Reductase) (Mattivi *et al*., 2006). Grape berry anthocyanins can be further categorized into di- or tri-hydroxylated anthocyanins based on their hydroxylation via the F3'H (Flavonoid 3'-hydroxylases) and F3'5'H (Flavonoid 3',5'-hydroxylases) enzymes (Falginella *et al.*, 2010), their methylation by AOMT1 (Anthocyanin O-methyltransferase 1) and AOMT2 (Anthocyanin O-methyltransferase 2) (Hugueney *et al.*, 2009; Fournier-Level *et al.*, 2011), and their acylation by 3AT (Anthocyanin 3-O-glucoside-6''-O-acyltransferase) (Rinaldo *et al.*, 2015). The dominance or absence of a specific category of anthocyanins is genotype-dependent in grapevine (Mattivi *et al.*, 2006). For example, most grape genotypes accumulate predominantly tri-hydroxylated anthocyanins (also known as Dp-derived anthocyanins), while genotypes such as Sangiovese synthesize dominantly di-hydroxylated anthocyanins (also known as Cy-derived anthocyanins) (Mattivi *et al.*, 2006; Pastore *et al.*, 2017); Some cultivars, including Pinot noir, only synthesize unacylated anthocyanins (Dimitrovska *et al.*, 2011; Rinaldo *et al.*, 2015). Moreover, the anthocyanins of most grape cultivars are limited to the berry skin, except in teinturier cultivars, which accumulate anthocyanins in both berry skin and pulp (Kong *et al.*, 2021). These specificities in anthocyanin composition can be considered as a cultivar identification fingerprint (Mattivi *et al.*, 2006; van Leeuwen *et al.*, 2013).

In addition to the genetic determinants, the composition of anthocyanins within a given cultivar undergoes dynamic changes throughout berry development and is influenced by environmental factors such as temperature (Sugiura *et al.*, 2018; de Rosas *et al.*, 2022), water supply (Berdeja *et al.*, 2014), nitrogen supply (Hilbert *et al.*, 2003; Olsen *et al.*, 2009), and light (Keller & Hrazdina, 1998). These complexities hinder the identification of key molecular regulators via association analysis

between genotypes and phenotypes for the determination of anthocyanin composition (Costantini *et al.*, 2015), because the traditional one-time-point phenotyping approaches cannot fully reflect the developmental and environmental dynamics of anthocyanin composition. Phenotyping time series across berry development may provide more comprehensive information for deciphering the role of genotype × development × environment interactions for anthocyanin composition. However, there has been limited research into the relative contributions of genotype, development, and environment on the dynamics of anthocyanin composition.

Mathematical models for metabolic pathway analysis (Morgan & Rhodes, 2002; Baghalian *et al.*, 2014) play an important role in experimental biology for advancing the boundaries of plant science (Marshallcolon *et al.*, 2017). These models may help to dissect complex traits, such as sugar concentrations in fruits, into simple traits by integrating time-series of phenotyping data, consequently facilitating the association analysis between genotypes and phenotypes (Génard *et al.*, 2010; Prudent *et al.*, 2011). Moreover, these models can also provide a rational analysis of biochemical pathways, identify potential limiting metabolic steps, screen candidate intervention points for bioengineering, as well as generate novel hypotheses for testing by wet-lab experiments (Rios-Estepa *et al.*, 2008; Farré *et al.*, 2014; Wang *et al.*, 2019, 2022). For example, several mathematical models haven been developed for lignin biosynthesis (Lee & Voit, 2010; Faraji & Voit, 2017; Wang *et al.*, 2018; Matthews *et al.*, 2020, 2021). Faraji et al. (2018) investigated several computational models of lignin biosynthesis in various plant species, including black cottonwood (*Populus trichocarpa*), alfalfa (*Medicago truncatula*), switchgrass (*Panicum virgatum*) and the grass *Brachypodium distachyon*. Their findings indicated that the intermediates of the lignin heteropolymer biosynthetic pathway are similar, while the enzymatic reactions of the pathway exhibit significant variation, providing valuable clues for targeted bioengineering of lignin composition (Matthews et al., 2021). Recently, a qualitative metabolic model has been constructed for the metabolic pathway of anthocyanins (Wheeler & Smith, 2019). Wheeler et al. (2019; 2020) developed a kinetic model for the anthocyanin pathway, which explained the production of blue, purple, and red anthocyanin pigments with multiple branches and substrate competition. This model was used to investigate the evolution of the anthocyanin pathway through fixed mutations and provided a theoretical framework to predict the consequences of new mutations in pigment phenotypes and pleiotropy (Wheeler & Smith, 2019). Notwithstanding its predictive value in assessing the effects of mutations on pigment production, this anthocyanin model provides mostly qualitative results and lacks quantitative outputs for a specific genotype over development under various environments.

The present study aimed at developing a dynamic model of anthocyanin composition based on the biosynthetic pathway (Boss *et al.*, 1996), which should be robust for a wide spectrum of grape genotypes and environmental conditions. The Dynamic Anthocyanin Composition Model (DACM) introduced here was calibrated and validated through the utilization of observed individual anthocyanin concentrations in berries of diverse grape cultivars under varying conditions. Calibration was achieved using five publicly available datasets and one dataset that has not been published. A global sensitivity analysis was then conducted to identify the key parameters controlling the concentration of individual anthocyanins. Model-based virtual experiments were finally utilized to explore strategies for fine tuning the anthocyanin composition with specific targets of biochemical decorations, providing possible intervention points for reorienting the metabolic fluxes within the pathway.

MATERIALS AND METHODS

Model description

The Dynamic Anthocyanin Composition Model (DACM) is a computational tool that enables the simulation of the accumulation profiles of individual anthocyanins during the development of a particular grape genotype from veraison (the onset of ripening) to maturity under varying environmental conditions (Hernández-Montes *et al.*, 2021), using one set of genotype-specific parameters. The structure of the model (Fig. 1B) is based on an explicit anthocyanin biosynthesis pathway (Boss *et al.*, 1996), which encompasses up to 15 different anthocyanins (3-glucosides) with various biochemical decorations of hydroxylation, methylation and acylation (Fig. 1A). Multiple enzymatic reactions are concatenated to simplify the model while preserving the crucial topology of the anthocyanin biosynthesis pathway (Fig. 1B).

The time-course changes in the quantity of a given anthocyanin were described as a result of 3 processes: biosynthesis, conversion to other anthocyanin forms, and degradation to non-anthocyanin metabolites. The biosynthesis of one individual anthocyanin or the conversion of one individual anthocyanin (V_{Cy}, V_{Dp}) to another was modeled as the product of a relative rate constant (r_i, i from 1 to 13) multiplying the substrate quantity, following the mass balance reaction principle. Similarly, the degradation of one anthocyanin to non-anthocyanin metabolites was modeled by multiplying the quantity of the specific anthocyanin with a relative degradation constant (*kd*). The degradation of anthocyanins is not well understood in grape berry, nor in other plant species, and the specific degradation rate for each individual anthocyanin is not available in the literature. For the sake of simplicity, we assumed that *kd* is the same for all degradation reactions, as previously proposed (Guardiola et al., 1995). To verify the robustness of this assumption, we also tested the model performance with different *kd*s together with model comparisons using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (Burnham and Anderson, 2002).

Taken together, the dynamic changes of each individual anthocyanin during berry development can then be described by the following 15 ordinary differential equations (ODEs):

$$\frac{dCy_{glc}}{dt} = V_{Cy} - (r_1 + r_{10} + r_{11} + kd)Cy_{glc}$$

$$\frac{dDp_{glc}}{dt} = V_{Dp} - (r_2 + r_3 + r_{12} + r_{13} + kd)Dp_{glc}$$

$$\frac{dPn_{glc}}{dt} = r_1 Cy_{glc} - (r_4 + r_5 + kd)Pn_{glc}$$

$$\frac{dPt_{glc}}{dt} = r_2 Dp_{glc} - (r_6 + r_7 + kd)Pt_{glc}$$

$$\frac{dMv_{glc}}{dt} = r_3 Dp_{glc} - (r_8 + r_9 + kd)Mv_{glc}$$

$$\frac{dPn_{ac}}{dt} = r_4 Pn_{glc} - kdPn_{ac}$$

$$\frac{dPn_{cou}}{dt} = r_5 Pn_{glc} - kdPn_{cou}$$

$$\frac{dPt_{ac}}{dt} = r_6 Pt_{glc} - kdPt_{ac}$$

$$\frac{dPt_{cou}}{dt} = r_7 Pt_{glc} - kdPt_{cou}$$

$$\frac{dMv_{ac}}{dt} = r_8 Mv_{glc} - kdMv_{ac}$$

$$\frac{dMv_{cou}}{dt} = r_9 Mv_{glc} - kdMv_{cou}$$

$$\frac{dCy_{ac}}{dt} = r_{10} Cy_{glc} - kdCy_{ac}$$

$$\frac{dCy_{cou}}{dt} = r_{11} Cy_{glc} - kdCy_{cou}$$

$$\frac{dDp_{ac}}{dt} = r_{12} Dp_{glc} - kdDp_{ac}$$

$$\frac{dDp_{cou}}{dt} = r_{13} Dp_{glc} - kdDp_{cou}$$

where r_i (i from 1 to 13) are the relative transformation rate constants (g anthocyanin/g whole berry/day); *kd* is the relative degradation constant (g/g/day), V_{Cy} is the influx of cyanidin-based anthocyanins (g/day), V_{Dp} is the influx of delphinidin-based anthocyanins (g/day). Cy_{glc} (g/berry) is the quantity of cyanindin-3-glucoside per berry, Dp_{glc} for delphinidin-3-glucoside, Pn_{glc} for peonidin-3-glucoside per berry; Pt_{glc} for petunidin-3-glucoside per berry; Mv_{glc} for malvidin-3-glucoside per berry; Pn_{ac} for peonidin-3-acetylglucoside per berry; Pn_{cou} for Peonidin-3-coumaroyl glucoside per berry; Pt_{ac} for petunidin-3-acetylglucoside per berry; Pt_{cou} for petunidin-3-coumaroyl glucoside per berry; Mv_{ac} for malvidin-3-acetylglucoside per berry; Mv_{cou} for malvidin-3-coumaroyl glucoside per berry; Cy_{ac} for cyanidine-3-acetylglucoside per berry; Cy_{cou} for cyanidine-3-coumaroyl glucoside per berry; Dp_{ac} for delphinidin-3-acetylglucoside per berry; Dp_{cou} for delphinidin-3-coumaroyl glucoside per berry. The quantity of each individual anthocyanin was expressed in equivalents of Mv_{glc}.

The input to the model was the total anthocyanin influx into the anthocyanin-specific pathway on each day (V_{in}, g/day), which was the sum of influxes into the two anthocyanin biosynthesis branches (V_{Cy} and V_{Dp}).

$$V_{in} = V_{Cy} + V_{Dp}$$

$$V_{Cy} = \delta \times V_{in}$$

$$V_{Dp} = (1 - \delta) \times V_{in}$$

where δ is an allocation coefficient (from 0 to 1) between the two branches of the metabolic pathway.

The V_{in} can be calculated with the following equations:

$$V_{in} = \frac{d\mathrm{TA}_{obs}}{dt} + kd\mathrm{TA}$$

$$kd = a_{kd} \times DOY + b_{kd}$$

where dTA_{obs}/dt is the newly accumulated total anthocyanin per berry at each time step. It was estimated from the first order derivative of the observed total anthocyanin (TA_{obs}) curves (Fig. S1) in the skin or pulp. The TA (g/berry) is the instantaneous total anthocyanins per berry in the simulation system. To take into account the developmental changes in anthocyanin degradation (Mori *et al.*, 2007; Movahed *et al.*, 2016), the *kd* was assumed to change as a linear function of berry development (DOY: day of the year) with parameters a_{kd} and b_{kd}. The a_{kd} (g/g/day^2) represents how fast the anthocyanin degradation constant changes as a function of berry development, while the b_{kd} (g/g/day) represents the basal degradation constant across the whole berry development (Table S1).

The integration of ODEs was utilized to determine the quantity of accumulated individual anthocyanins in each berry. The resulting amount of anthocyanin was then divided by the fresh weight of the grape berry or specific tissues (Fig. S2), such as skin or pulp, to calculate the concentration of each individual anthocyanin.

Experimental data

Six datasets were utilized in this study obtained from 6 separate experiments encompassing 8 distinct grapevine cultivars (Table S2), 2 rootstocks, and 37 growth conditions that included variations in growing seasons, leaf-to-fruit ratio, water stress, nitrogen supply, and light levels (Table 1). The datasets 1-5 were derived from previously published studies, and their respective experimental designs are briefly summarized below.

The first experiment investigated the effects of nitrogen (N) availability (0.34, 1.7, or 3.4 g N per plant as NH_4NO_3 applied at bloom) and light intensity (3 weeks at 100%, 20%, or 2% of full sunlight, modulated outdoors with shade cloth starting from veraison) on growth and fruit ripening of pot-grown cv. Cabernet Sauvignon (*V. vinifera*) vines (Keller & Hrazdina, 1998). The second experiment contained 3 nitrogen treatments applied from fruit set to leaf fall (1.4 mM N, 3.6 mM N and 7.2 mM N, denominated N1, N2, N3I, respectively) with cv. Merlot (*V. vinifera*) vines in a greenhouse (Hilbert *et al.*, 2003). The third dataset illustrated the effects of different rootstocks and water supply levels on berry growth and anthocyanins in cv. Pinot noir (*V. vinifera*), which was grafted onto either rootstock 110R (drought tolerant, medium to high vigour) or 125AA (drought sensitive, high vigour) during 3 growing seasons in the field under control or water shortage conditions (Berdeja *et al.*, 2014). The fourth experiment investigated the effects of two leaf-to-fruit ratio levels on the quality of berries of cvs. Cabernet Sauvignon and Sangiovese (*V. vinifera*), respectively (Bobeica *et al.*, 2015). The fifth experiment showed the anthocyanins in skin and pulp of grape berries collected from cvs.

Gamay, Gamay de Bouze and Gamay Freaux (*V. vinifera*) grown in a greenhouse (Kong *et al.*, 2021). The sixth experiment was conducted in the current study. Briefly, two cultivars, Cabernet Sauvignon and Tempranillo (*V. vinifera*), were grown in a common garden vineyard named 'Vitadapt' (Suter *et al.*, 2021) under standard viticultural practices. Berries were collected at 10-day intervals from veraison to maturity in two growing seasons with 3 biological replicates (30 berries of each replicate) for measuring the berry fresh weight with high-precision balance and anthocyanin composition with HPLC as described in Kong *et al.* (2021).

These datasets were chosen based on 4 criteria, including 1) there must be measurements of individual anthocyanins in grape berries with specific analytical equipment (HPLC); 2) the dataset should contain different environmental treatments, as this model aims to simulate the accumulation of anthocyanin composition under varying environmental conditions; 3) the anthocyanin composition must be measured throughout berry ripening, covering at least 4 different developmental stages; 4) the berry and/or skin fresh weight should be measured, enabling calculation of the total anthocyanin content per berry at each sampling date. Accordingly, the aforementioned 6 experiments all determined the developmental dynamics of grape growth (e.g. fresh weight, Fig. S2) and quantified the anthocyanin composition from veraison to maturity with 4-10 sampling dates (Fig. S1). Moreover, the analysis of individual anthocyanins for all 6 datasets was performed by HPLC, and the quantification was based on calibration curves with malvidin-3-glucoside as external standard, with all other individual anthocyanins being expressed as malvidin-3-glucoside equivalents.

Model resolution, parameterization and validation

The dynamic model was simulated with a one-day time step and implemented using R software (R Core Team, 2013). The ODEs of the model were numerically integrated using the Euler method.

For the parameterization, a genetic algorithm (GA) was employed to estimate all the model parameters for each cultivar in the dataset (Table S1). The GA function in R (R Core Team, 2013) was used to minimize an objective criterion, which was defined as:

$$\text{criterion} = \sum_{i=1}^{n} \frac{\sqrt{(1/m)\sum_{j=1}^{m}(y_{oj} - y_{sj})_i^2}}{(1/m)\sum_{j=1}^{m} y_{oji}} \#$$

where n is the number of conditions for each cultivar, m is the sampling number of each condition throughout berry development; y_o and y_s are the observed and simulated values of anthocyanins, respectively.

The parameterization process with GA was repeated 5 times to obtain 5 sets of parameters to assess parameter stability. The values of the set of parameters that gave the smallest criterion value was used as the best estimated parameter values for each cultivar in each dataset (Table S1 and Table S3).

During the process of parameterizing the model for a specific cultivar under diverse environmental conditions it was discovered that the allocation coefficient (δ) varied in relation to the levels of

nitrogen and light intensity. Subsequently, an equation was developed to illustrate the responses of δ to different nitrogen and light intensity conditions for the cultivar in question.

$$\delta = \frac{1}{1 + e^{-[a_\delta \times N_{rel} + b_\delta \times light_{rel}]}} + c_\delta \#$$

where a_δ and b_δ are the coefficients of nitrogen and light levels, respectively and c_δ is a constant. N_{rel} was the relative nitrogen availability under optimum growing condition (considered as 1) for the investigated cultivar, while $light_{rel}$ was the relative light intensity under full sunlight (considered as 1).

To assess the performance and predictive ability of the model, various statistical measures were used for each cultivar, including the root mean squared error (RMSE), relative root mean squared error (RRMSE), and coefficient of determination for the linear correlation (R^2). The model's prediction quality was evaluated through the leave-one-out cross-validation method (Wallach *et al.*, 2006). Specifically, this process was carried out for Cabernet Sauvignon (dataset 1), Pinot noir (dataset 2), and Merlot (dataset 3), which each contained more than 3 growth conditions. For each cultivar, a set of optimal parameters was estimated using all observed data except for one condition, and these parameters were then used to predict the anthocyanin composition under the omitted condition. This validation process was repeated multiple times, with 9 runs for Cabernet Sauvignon, 12 runs for Pinot noir, and 3 runs for Merlot. The RMSE and RRMSE were calculated for each validation condition, and averaging these values across all growing conditions provided an overall estimate of prediction quality for each cultivar (Wallach *et al.*, 2014).

Global sensitivity analysis

To determine the key metabolic steps influencing anthocyanin composition, global sensitivity analysis using the Morris method (Morris, 1991) was performed. The investigation involved the use of 8-18 parameters for different anthocyanin compositions to study their effects on model outputs. The assumption was made that all the investigated parameters were uniformly distributed within the range of 0.9 to 1.1 times of the default values. The sensitivity index (SI) and its standard deviation (ST) obtained from the Morris method were used to assess parameter sensitivity. The stability of the sensitivity analysis was ensured by verifying the convergence of sensitivity ranking based on SI of each parameter by gradually increasing the number of trajectories, and finally, 300 trajectories were applied to provide a stable sensitivity ranking. The sensitivity analysis was performed using the 'Morris' function in the 'sensitivity' package of the R language (Iooss & Lemaître, 2015).

RESULTS

Overview of the influence of genotype and environment on total anthocyanins and anthocyanin composition

The concentrations of total anthocyanins showed clear differences between cultivars and growing seasons and were increased by high light intensity (Fig. S1A) and leaf-to-fruit ratio (Fig. S1D), but were decreased by high N levels (Fig. S1A, B) and water supply (Fig. S1C). Here, we focused on the responses of anthocyanin composition related to their biochemical decorations of hydroxylation (Fig. 2), methylation (Fig. S3) and acylation (Fig. S4).

Based on their hydroxylation, individual anthocyanins can be grouped into di- and tri-hydroxylated anthocyanins (Fig. 2). The ratio of di- to tri-hydroxylated anthocyanins (RDT) represents one of the most important properties for anthocyanin composition, indicating whether the grape color is reddish or dark-blue (Castellarin *et al.*, 2006). This ratio was largely determined by genotypes: the Sangiovese and the pulp of Gamay de Bouze and Gamay Freaux had RDT larger than one (Fig. 2D and E), while the remaining genotypes, including Cabernet Sauvignon, Gamay, Merlot, Pinot noir, and Tempranillo, had RDT smaller than one (Fig. 2A, B, C, and F). These results showed that Sangiovese skin and the pulp of Gamay de Bouze and Gamay Freaux accumulated dominantly di-hydroxylated anthocyanins, while the remaining genotypes accumulated dominantly tri-hydroxylated anthocyanins. Over berry development, the RDT gradually increased in genotypes with RDT larger than 1 (Fig. 2D and E), while it gradually decreased in genotypes with RDT smaller than 1 (Fig. 2A, B, C, and F). However, the RDT varied less in response to the explored combinations of growth conditions (Fig. 2) than the concentrations of total anthocyanins (Fig. S1). The RDT in Cabernet Sauvignon was not significantly different between light and nitrogen conditions, except the extreme condition under the highest light intensity (L100) and lowest nitrogen supply (N1) (Fig. 2A). The RDT of the extreme condition (L100-N1) was more than 1 while in other conditions it was less than 1 (Fig. 2A). Moreover, the effects of nitrogen and light intensity on RDT were opposite, as the RDT responded positively to increasing light intensity and negatively to increasing N levels (Fig. 2A). A similar negative effect of nitrogen on the RDT was also observed in Merlot (Fig. 2B). The RDT of Pinot noir was hardly affected by rootstock, water stress, and growing seasons, particularly around maturity (Fig. 2C). The RDT was affected by the leaf-to-fruit ratio in a genotype-dependent manner, with the RDT of Cabernet Sauvignon being systematically decreased by lower leaf-to-fruit ratio while the RDT of Sangiovese was hardly affected by the leaf-to-fruit ratio (Fig. 2D). The developmental profiles of RDT were different in 2013 and 2014 for Tempranillo and Cabernet Sauvignon; however, they overlapped when resynchronized with veraison dates (Fig. 2F).

Based on the methylation of individual anthocyanins, they could be grouped into methylated and unmethylated anthocyanins (Fig. S3). Seven out of the 8 investigated cultivars possessed predominantly methylated anthocyanins, with a ratio of methylated to unmethylated anthocyanins ranging between 2 and 200 (Fig. S3), while only Sangiovese had more unmethylated anthocyanins at maturity with a methylated to unmethylated ratio smaller than 1 (Fig. S3D). In response to various growing conditions, this ratio was increased by conditions that reduced the total anthocyanins (Fig. S3), such as low light (Fig. S3A), high nitrogen (Fig. S3A and S3B), and low leaf-to-fruit ratio for Cabernet Sauvignon (Fig. S3D).

Based on their acylation, individual anthocyanins could be grouped into acylated and unacylated anthocyanins (Fig. S4). Pinot noir and Sangiovese do not accumulate acylated anthocyanins (Mattivi *et al.*, 2006; Rinaldo *et al.*, 2015) and thus showed a ratio of acylated to unacylated anthocyanins at 0 (Fig. S4C and S4D). For the other cultivars, this ratio ranged from 0.05 to 1.0 (Fig. S4B, S4D, S4E and S4F), with higher values under conditions that reduced the total anthocyanins, such as high nitrogen (Fig. S4B) and low leaf-to-fruit ratio (Fig. S4D).

Calibration of the dynamic anthocyanin composition model (DACM)

To simulate the developmental dynamics of anthocyanin composition in different cultivars under various growth conditions, the DACM was developed based on mass balance of biochemical reaction rules with the total anthocyanin influx as input and reaction rates as parameters. The DACM was calibrated with actual measurements of each individual anthocyanin along berry development from the 6 datasets (Table 1). With a unique set of parameter values for each cultivar in each dataset (Table S3), the model simulations were highly aligned with the observed results for most genotype x growth conditions (Figs. 3, 4, S5-S12). The performance of the DACM for each dataset is briefly described below.

For the dataset collected during the current study (dataset 6), the DACM precisely simulated the developmental dynamics of the observed concentrations of 11 individual anthocyanins in Cabernet Sauvignon in 2013 and 2014 with RRMSEs of 25.2% and 25.6%, respectively, and R^2 of 0.96 in both years (Fig. 3). The model also accurately reproduced the developmental dynamics of the observed concentrations of 9 individual anthocyanins in Tempranillo in both 2013 and 2014, with RRMSE of 23.7% and 17.0%, and R^2 of 0.96 and 0.99, respectively (Fig. 3). These results were obtained with the same set of parameter values for a given cultivar in both years (Table S3), indicating the stability of parameters across growing seasons.

Moreover, we tested the capability of the model to simulate different anthocyanin categories according to their molecular decorations, including di- vs tri-hydroxylated, and methylated vs unmethylated anthocyanins (Fig. 4). The model precisely simulated the developmental dynamics of the observed concentrations of the 4 anthocyanin categories in Cabernet Sauvignon and Tempranillo in both 2013 and 2014, with RRMSEs ranging from 10.2% to 15.4% and R^2 ranging from 0.98 to 1. These results show that the model accuracy is higher for anthocyanin categories with distinct decorations (Fig. 4) than for individual anthocyanins (Fig. 3).

Similarly, the model simulation agreed well with experimental observations for Cabernet Sauvignon under 9 growth conditions covering 3 canopy light intensities and 3 levels of soil nitrogen supply (dataset 1, Fig. S5 and S6) with RRMSE ranging from 12.8% to 44.9% and R^2 ranging from 0.82 to 0.97. In detail, the model simulation performed better in the 3 N conditions under high (light 3) and moderate light (light 2) with RRMSE ranging from 16.9 to 23.5%, than in the low light conditions (light 1) with RRMSE ranging from 28.8 to 44.9% (Fig. S5, Keller). The larger RRMSE under low light intensity (Light 1) were attributed mainly to the underestimation of the malvidin-3-glucoside (Mv_{glc}) concentration in the rapid accumulation stage, when their counterparts under higher light (light 2

and light 3) had already plateaued. On the other hand, the model performed similarly under the 3 N levels under a given light intensity in dataset 1 (Fig. S5, Keller). The second dataset (Fig. S5 and S6, Hilbert) consisted of 3 soil nitrogen supply levels in Merlot with 13 individual anthocyanins, and the model precisely reproduced the developmental dynamics of individual anthocyanins or anthocyanin categories with RRMSE ranging from 18.4 to 28.8% and R^2 ranging from 0.95 to 0.98.

The third dataset (Fig. S7 and S8) consisted of 2 rootstocks and 2 water supply conditions for Pinot noir in 3 growing seasons (2009-2011), which contained 5 individual anthocyanins with RRMSE for the simulation results ranging from 8.0 to 38.7%, and R^2 ranging from 0.82 to 1.00.

The fourth dataset (Fig. S9 and S10) consisted of 2 leaf-to-fruit ratios for Cabernet Sauvignon and Sangiovese, which contained 11 and 5 individual anthocyanins respectively with RRMSE for the simulation results ranging from 13.4 to 61.2%, and R^2 ranging from 0.86 to 0.97. Under low leaf-to-fruit ratio (3 leaves per cluster), the large RRMSE for Cabernet Sauvignon was attributed mainly to the model overestimating delphinidin-3-glucoside (Dp_{glc}) and peonidin-3-glucoside (Pn_{glc}) during berry development (Fig. S8, 3L CS), while the large RRMSE for Sangiovese was attributed mainly to the model underestimating cyanidin-3-glucoside (Cy_{glc}) in the late berry development stages (Fig. S8, 3L S).

The fifth dataset (Fig. S11 and S12) consisted of 3 cultivars (Gamay, Gamay de Bouze and Gamay Freaux) in 2 berry tissues (skin and pulp) which contained 7 individual anthocyanins with RRMSE for the simulation results ranging from 15.4 to 50.1%, and R^2 ranging from 0.94 to 0.98. The large RRMSE for pulp was attributed mainly to the model underestimating peonidin-3-glucoside (Pn_{glc}) and overestimating malvidin-3-glucoside (Mv_{glc}) during berry development (Fig. S11, pulp). Because of the close genetic relationship among the 3 cultivars (Kong *et al.*, 2021), the anthocyanins in the skin were simulated with the same set of parameter values for all cultivars (Table S3), while those in the pulp were simulated with a set of parameter values different from the skin (Table S3). These results indicated that the skin and pulp of teinturier cultivars (i.e., cultivars with anthocyanins in both the skin and pulp) need to be considered separately.

Validation of the dynamic anthocyanin composition model

To assess the prediction quality of the model, 3 datasets with more than 3 environmental conditions for each dataset (dataset 1, dataset 2 and dataset 3) were tested with the leave-one-out cross-validation (Wallach *et al.*, 2006). The model prediction quality was high under the 9 combinations of light and N levels in Cabernet Sauvignon, with RRMSEP = 26.38% and mean R^2 = 0.91 (Table 2). Similarly, the model prediction quality was high under 3 N levels in Merlot, with mean RRMSEP = 32.91% and mean R^2 = 0.96 (Table 2). The model prediction quality was also high under the 12 combinations of water stress and rootstocks in different vintages in Pinot noir, with mean RRMSEP = 28.80% and mean R^2 = 0.99 (Table 2). Overall, the leave-one-out cross-validation results (RMSEP, RRMSEP) were comparable to model calibration results (RMSE, RRMSE), indicating the DACM possessed very high prediction quality.

Sensitivity analysis of the model

To investigate the influence of different parameters on model outputs, a global sensitivity analysis was conducted using the Morris method (Morris, 1991). The model parameters were categorized into 3 groups based on their association with conversion rate (r_i), degradation rate (*kd*), or allocation coefficient (δ). Four cultivars with distinct anthocyanin composition, namely Pinot noir, Sangiovese, the skin of Gamay Freaux, and Cabernet Sauvignon from dataset 4, were chosen for the sensitivity analysis (Table S2). The results indicated that the parameter sensitivity for any given anthocyanin remained relatively stable throughout berry development under the growing conditions investigated (Figs. S13-S16). Therefore, only the most sensitive parameters for the concentrations of each individual anthocyanin at maturity were examined in detail (Figs. 5 and S17-S20).

Both Pinot noir and Sangiovese berries accumulate exclusively unacylated anthocyanins (Fig. S4C and S4D, Mattivi *et al.*, 2006; Rinaldo *et al.*, 2015), therefore their berries biosynthesize only 5 individual anthocyanins that were modeled here. Moreover, Pinot noir predominately accumulates tri-hydroxylated anthocyanins (RDT<1, Fig. 1C) and Sangiovese predominately accumulates di-hydroxylated anthocyanins (RDT>1, Fig. 1D, Table S2, Mattivi *et al.*, 2006). Therefore, we compared parameter sensitivities in these two cultivars for the 5 individual anthocyanins (Fig. 5).

In Pinot noir (Figs. 5A and S17), degradation (a_{kd}, b_{kd}) had the greatest effect on the two unmethylated anthocyanins (Cy_{glc} and Dp_{glc}), with a smaller effect from influx to the pathway (c_δ). The concentration of mono-methylated Pn_{glc} was mainly affected by biosynthesis (r_1), influx to the pathway (c_δ) and degradation (a_{kd}, b_{kd}). The concentration of mono-methylated Pt_{glc} was primarily affected by competition with Mv_{glc} biosynthesis (r_3) from the common substrate Dp_{glc}, followed by degradation (a_{kd}, b_{kd}), influx to the pathway (c_δ), and biosynthesis (r_2). The concentration of di-methylated and usually dominant Mv_{glc} was primarily affected by degradation (a_{kd}, b_{kd}), followed by influx to the pathway (c_δ), competition (r_2) and biosynthesis (r_3). The ratio of di- to tri-hydroxylated anthocyanins was only affected by influx to the pathway (c_δ), while the ratios of methylated to unmethylated anthocyanins were mainly affected by degradation (a_{kd}, b_{kd}), followed by r_3 and r_1.

In Sangiovese (Figs. 5B and S18), all 5 individual anthocyanins were mainly affected by degradation (a_{kd}, b_{kd}) and influx to the pathway (c_δ), while their biosynthesis, conversion, or competition played minor roles. The ratio of anthocyanins with 2 different decorations showed the same model parameter sensitivity as in Pinot noir.

In the berry skin of Gamay Freaux (Figs. 5C and S19), the degradation (a_{kd}, b_{kd}) had the greatest effect on the two unmethylated anthocyanins (Cy_{glc} and Dp_{glc}), with a smaller effect from biosynthesis (c_δ). The concentration of methylated and 3 unacylated anthocyanins (Pn_{glc}, Pt_{glc}, Mv_{glc}) was mainly affected by influx to the pathway (c_δ), followed by biosynthesis and competition. The concentration of acylated anthocyanins was mainly affected by degradation (a_{kd}, b_{kd}), followed by competition and biosynthesis. The ratio of di- to tri-hydroxylated anthocyanins was only affected by influx to the pathway (c_δ), while the ratios of methylated to unmethylated anthocyanins and acylated to unacylated anthocyanins were mainly affected by degradation (a_{kd}, b_{kd}).

In Cabernet Sauvignon (Figs. 5D and S20), all individual anthocyanins were mostly influenced by their biosynthesis, followed by degradation (a_{kd}, b_{kd}) and competition. The ratio of the 3 differently decorated anthocyanins showed the same model parameter sensitivity as in Gamay Freaux.

Virtual experiment for targeted tuning of anthocyanin composition

A virtual simulation (Fig. 6) was conducted to explore potential strategies aiming at tuning the anthocyanin composition for targeted objectives, which were set to increase the proportions of tri-hydroxylated and methylated anthocyanins that are more stable and provide specific color hues (He et al., 2010, Liu et al., 2022, Houghton et al., 2021). The virtual experiment was implemented in Sangiovese and Cabernet Sauvignon (Table S2) by modifying the top 7 most sensitive parameters identified via the global sensitivity analysis (Fig. 5). Modulating these 7 parameters around their default values by -10%, -5%, 0%, +5%, and +10% while keeping other initial, input, and parameter values in the default condition resulted in 78,125 possible combinations. As expected from the global sensitivity analysis, the proportions of tri-hydroxylated anthocyanins were exclusively affected by the parameter c_δ (Fig. 5A, B), which reflects the influx at the branching point of the anthocyanin metabolism pathway (Fig. 1B). In contrast, the proportions of methylated anthocyanins were influenced by all 7 parameters with different intensities (Fig. 6A, B). Interestingly, the same variation range (±10%) in c_δ led to different magnitudes of changes in the tri-hydroxylated anthocyanins in Sangiovese and Cabernet Sauvignon, with variations of ±22% in Sangiovese but only ±1.3% in Cabernet Sauvignon (Fig. 6A, B). Similar differences were also observed in the proportion of methylated anthocyanins, where Sangiovese showed a greater variation range (-25~55%) than those in Cabernet Sauvignon (only about ±6%) (Fig. 6A, B).

The global optimal combinations for the highest proportions of tri-hydroxylated and methylated anthocyanins were obtained when all 7 parameters were simultaneously set to their largest boundaries (namely ±10%, the max7 in Fig. 6E, F). These parameter combinations resulted in a 22% increase in the proportion of tri-hydroxylated anthocyanins and 58% increase in the proportion of methylated anthocyanins in Sangiovese, while the increases were only 1.3% and 5.5%, respectively, in Cabernet Sauvignon (Fig. 6A, B). However, modifying 7 parameters that may represent 7 enzymatic steps will be almost unfeasible via current bioengineering technologies, which are more suitable to handle 1-3 genes or enzymes (Noda *et al.*, 2017; Zhu *et al.*, 2021). Therefore, we further explored what might be the minimum number of adjusted parameters to reach about 90% of the improvements in the global optimal combinations (Fig. 6C, D). It was found that the targets were approached when changing 1 to 3 parameters, after which further increasing the number of adjusted parameters brought minor improvements. In fact, changing only 3 parameters (max3) yielded a performance of 92.5% of the global optimal combination (max7) in Sangiovese and 79.4% in Cabernet Sauvignon (Fig. 6E, F). Moreover, these local optimal combinations (max3) were reached by adjusting different sets of 3 parameters in Sangiovese and Cabernet Sauvignon, namely the c_δ (-10%), a_{kd} (-10%), b_{kd} (-10%) for Sangiovese, and c_δ (-10%), a_{kd} (-10%), r_3 (+10%) for Cabernet Sauvignon, respectively (Fig. 6E, F).

We then verified which/how individual anthocyanins were modified for increasing the proportions of tri-hydroxylated and methylated anthocyanins in the simulations. Under different parameter combinations, the changes in proportions of individual anthocyanins in Sangiovese were mainly attributed to a significant decrease in the unmethylated anthocyanin Cy_{glc}, while its downstream product Pn_{glc}, which is a methylated anthocyanin, increased significantly (Fig. 6G). In Cabernet Sauvignon, which has a broader spectrum of individual anthocyanins, the changes were relatively

small compared to Sangiovese. Specifically, the contents of the unmethylated anthocyanins Cy_{glc} and Dp_{glc}, and the methylated anthocyanin Pt_{glc}, all decreased slightly, while the methylated anthocyanins Pn_{glc} and Mv_{glc} and their downstream products all increased slightly in Cabernet Sauvignon (Fig. 6H).

DISCUSSION

The present study parameterized and tested a mechanistic model that accurately simulated the dynamic accumulation of individual anthocyanins in ripening grape berries. To calibrate and validate the model, 6 datasets with 8 *V. vinifera* cultivars and 37 environmental conditions were utilized. The model parameters remained consistent and reliable across varying environmental conditions within each dataset for a given cultivar. This genotype-dependent but environment-independent property of model parameters enables the current model to serve as a novel phenotyping tool to dissect the complex traits of dynamic anthocyanin profiles into simple traits (Génard *et al.*, 2010; Bertin *et al.*, 2010; Dai *et al.*, 2017). Combining modeled traits with QTL/GWAS analysis could help us to unveil metabolic steps responsible for fine-tuning anthocyanin composition in grape berries and, eventually, other crops. Such a tool would facilitate genotype-to-phenotype analysis and prediction (Chenu *et al.*, 2018).

The biosynthesis and subsequent decorations of anthocyanins, as for most specialized metabolites, are arranged in metabolic pathways with various topological structures, including linear, cyclical, branched, or 3-dimensional grids (Farré *et al.*, 2014). For example, the anthocyanin decorations in Arabidopsis seem to be arranged in a highly branched 3-dimensional grid (Saito *et al.*, 2013), while the anthocyanin decorations follow strict orders in a linear way with several metabolic branches in *Petunia spp*. (Provenzano *et al.*, 2014), as well as in grapes (*V. vinifera*) (Fig.1, Ford *et al.*, 1998; Hugueney *et al.*, 2009). As a result, attempts to modify anthocyanin composition, as well as other specialized metabolites, often suffer from high uncertainties due to pathway complexity (Zhang *et al.*, 2014; Noda *et al.*, 2017; Zhu *et al.*, 2021; Wang *et al.*, 2022). Rational analysis of a metabolic pathway with mathematic models may aid in overcoming such difficulties by evaluating flux distributions and identifying candidate intervention points for bioengineering in order to enrich desirable compounds (Farré *et al.*, 2014; Faraji & Voit, 2017; Wang *et al.*, 2019). To this end, we first developed a dynamic anthocyanin composition model (DACM) and then used it in a virtual experiment to explore possible strategies to enrich tri-hydroxylated and methylated anthocyanins in different genetic backgrounds (Fig. 6). The variation in the relative proportions of di- to tri-hydroxylated anthocyanins was primarily caused by the allocation coefficient (δ) between the two anthocyanin biosynthesis branches, in agreement with the suggestion that anthocyanin hydroxylation is mainly determined by the relative expression of *VvF3'H* and *VvF3'5'H* in grapevine (Castellarin *et al.*, 2007). On the other hand, fine-tuning methylated anthocyanins might be more challenging, because the proportion of methylated anthocyanins exhibited more complex reactions in a genetic background dependent manner. This seems mainly due to the higher number of parameters regulating anthocyanin methylation than hydroxylation and due to the fact that the cumulative effects of parameters produce greater changes for the ratio of methylated-to-unmethylated anthocyanins. Interestingly, genotypes with complex anthocyanin profiles, such as Cabernet Sauvignon with at least 14 individual anthocyanins, exhibited greater composition stability

in response to parameter perturbations than those with simple anthocyanin profiles, such as Sangiovese, with 5 individual anthocyanins. These results highlight the trade-offs that may occur in complex anthocyanin biosynthesis networks, because of the competition for shared substrates and/or enzymes (Wheeler & Smith, 2019). With these trade-offs, the shifts towards producing one type of pigment can result in a reduction in the production of other pigments due to limited substrates or modified substrate specificity of enzymes (Wheeler and Smith, 2019). Within a more complex anthocyanin biosynthesis network, the number of metabolic steps/branches with competing substrates and/or enzymes will increase, and consequently increase the probabilities to mitigate the effects of exogenous perturbations. Keeping this in mind, manipulating the anthocyanin composition to a specific target through genetic engineering may be easier in cultivars with simpler anthocyanin profiles (Zhu et al., 2021). These cultivars can offer reduced regulatory complexity, making them easier to achieve desired alterations without unintended effects, and provide more predictability for the outcomes (Shimada et al., 2001; Lin-Wang et al., 2014).

In theory, the enrichment of a desirable metabolite in a pathway may be reached by increasing its biosynthesis precursor and catalytic enzyme activities while decreasing its degradation and/or competitive pathway (Farré *et al.*, 2014; Manela *et al.*, 2015; Wang *et al.*, 2021).However, it is not always straightforward to predict which steps or combinations of steps are the most pertinent strategy to reconstruct or reorient a metabolic pathway for producing targeted products. Our virtual experiments showed that the target, which was set to simultaneously increase the proportions of tri-hydroxylated and methylated anthocyanins, was optimally achieved by collectively modifying up to 7 parameters. Considering the feasibility in most circumstances for plant-based bioengineering, we further explored the minimal set of parameters required to reach about 90% of the optimal achievement. We found that 3 parameters will largely fulfill the target, but the exact 3 parameters differed between the two tested genotypes, namely Cabernet Sauvignon and Sangiovese. For Cabernet Sauvignon, the best parameter combination constituted the influx at branching points of the pathway and the degradation of anthocyanins. For Sangiovese, the two first parameters were the same as in Cabernet Sauvignon, but the third parameter (r_3) was related to reducing a competitive branch. Interestingly, the parameter r_3 may represent the step of anthocyanin methylation, primarily involving two O-methyltransferases (AOMTs) in grape berries. The VvAOMT1 is known to preferentially catalyze the 3', 5' methylation (Hugueney et al., 2009), while VvAOMT2 prefers 3' methylation (Fournier-Level et al., 2011). In a study mimicking an increase in r_3 through the over-expression of *VvAOMT1* in *Petunia spp*., a higher percentage of methylated anthocyanins was observed (Provenzano et al., 2014), in agreement with model predictions. These results offer a significant direction for the development of more efficient strategies for fine tuning anthocyanin composition, and highlight again the importance of the genetic background of the host plants.

Both our model analysis and virtual experiments showed that the anthocyanin degradation related parameters (a_{kd}, b_{kd}) played an important role in the model for different anthocyanin compositions (Fig. 6). Similarly, previous studies suggested that reducing the *in vivo* process of anthocyanin degradation could increase crop pigmentation and prevent color degradation (Oren-Shamir, 2009; Zipor *et al.*, 2014; Liu *et al.*, 2018). In fact, anthocyanin degradation can be induced by spontaneous

reactions, enzymatic activity, or both (Oren-Shamir, 2009). Peroxidases were discovered to be involved in anthocyanin degradation in grape (Calderon *et al.*, 1992) and *Brunfelsia calycina* flowers (Zipor *et al.*, 2014). Moreover, the *VviPrx31* peroxidase may act as a candidate gene involved in anthocyanin degradation in ripening grape berries under high temperature (Movahed *et al.*, 2016). However, the nature of anthocyanin degradation in grape and other plants is far from being fully understood. To further explore this point, we tested the model performance using two sets of kd values for the di- and tri-hydroxylated anthocyanins. The results showed that the models with two sets of kd values generally provided slightly better reproductions of the observations (mean RRMSE=29.85%, Table S4) than models with one set of kd values (mean RRMSE=31.7%, Table S4). However, when simultaneously considering model precision and model complexity, the two-kd model performed worse than the one-kd model (Table S4), as indicated by the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (Burnham and Anderson, 2002). Following the Occam's razor principle (Forster, 2000), we opted for a simpler model with a single set of kd. Though this point remains to be tested experimentally, this analysis suggests that the degradation constant might be similar for distinct anthocyanins and the one-kd assumption seems reliable. Moreover, the only study that conducted substrate specificity analysis for one anthocyanin degradation enzyme, BcPrx01 from *Brunfelsia calycina*, also showed that the rates of peroxidase-catalyzed degradation were similar for all individual anthocyanins (Zipor *et al.*, 2015). These findings underscore the complexity of anthocyanin degradation and signal the need for further investigation into this process in grapes and other plants.

Despite the robustness of our dynamic anthocyanin composition model, the model may not fully account for alterations in anthocyanin composition under extreme treatment conditions, such as dark or severe nutrient stress. These severe stresses may result in significant variability in the pattern of anthocyanin composition across different cultivars. This suggests that the model parameter values may vary with the extreme environmental perturbations and that this type of response should be studied to improve the model prediction. Moreover, given the laborious measurement of anthocyanin concentrations, which serve as the input for the current model, further improvements to model usability are currently being explored. In particular, the development of a total anthocyanin prediction model, which incorporates cultivar, sugars, light, temperature (Sugiura *et al.*, 2018), and other factors as inputs, may enhance the predictive power of the model. Overall, the model holds promise as a useful tool in phenotyping time-series of anthocyanin measurements, as well as a rationale for bioengineering applications aiming to fine tune anthocyanin composition.

Acknowledgments

This research was supported partly by the National Key R&D Program of China (2021YFE0109500), National Natural Science Foundation of China (U20A2041), Agricultural Breeding Project of Ningxia Hui Autonomous Region (NXNYYZ202101), and CAS Youth Interdisciplinary Team (JCTD-2022-06). Research conducted as part of the LIA INNOGRAPE International Associated Laboratory.

Author contributions

The study was designed by YW and ZD; BS, MG, GH, SD, EG, SP, MK, CR and ZD contributed to the datasets collection; YW, JC and ZD constructed the model and wrote the simulation code; YW undertook the model testing and refinement. YW, BS, and ZD contributed to analyzing the simulation results; YW wrote the draft and all authors contributed to revising the paper; all authors approved the final manuscript.

References

Baghalian K, Hajirezaei M-R, Schreiber F, 2014. Plant metabolic modeling: achieving new insight into metabolism and metabolic engineering. *The Plant Cell* **26**, 3847-3866.

Berdeja M, Hilbert G, Dai ZW, Lafontaine M, Stoll M, Schultz HR, Delrot S, 2014. Effect of water stress and rootstock genotype on Pinot Noir berry composition: Rootstock and water stress affect berry quality. *Australian Journal of Grape and Wine Research* **20**, 409-421.

Bertin N, Martre P, Génard M, Quilot B, Salon C, 2010. Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. *Journal of Experimental Botany* **61**, 955-967.

Bobeica N, Poni S, Hilbert G, Renaud C, GomÃ¨s E, Delrot S, Dai Z, 2015. Differential responses of sugar, organic acids and anthocyanins to source-sink modulation in Cabernet Sauvignon and Sangiovese grapevines. *Frontiers in Plant Science* **6**, 382.

Boss PK, Davies C, Robinson SP, 1996. Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv shiraz grape berries and the implications for pathway regulation. *Plant Physiology* **111**, 1059-1066.

Burnham KP, Anderson DR (eds.), 2002. Model Selection and Multimodel Inference: A Practical Information-theoretic Approach. New York, NY, USA: *Springer*.

Calderon AA, Garcfa-Florenciano E, Munoz R, Ros Barcelo A, 1992. Gamay grapevine peroxidase: lts role in vacuolar anthocyani(di)n degradation. *Vitis* **31**, 139-147.

Castellarin SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, Paillard S, Adam-Blondon A-F, Testolin R, 2006. Colour variation in red grapevines (*Vitis vinifera* L.): genomic organisation, expression of flavonoid 3'-hydroxylase, flavonoid 3',5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. *BMC Genomics* **7**, 12.

Castellarin SD, Matthews MA, Di Gaspero G, Gambetta GA, 2007. Water deficits accelerate ripening and induce changes in gene expression regulating flavonoid biosynthesis in grape berries. *Planta* **227**, 101-112.

Chenu K, Van Oosterom EJ, McLean G, Deifel KS, Fletcher A, Geetika G, Tirfessa A, Mace ES, Jordan DR, Sulman R, Hammer GL, 2018. Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. *Journal of Experimental Botany* **69**, 3181-3194.

Costantini L, Malacarne G, Lorenzi S, Troggio M, Mattivi F, Moser C, Grando MS, 2015. New candidate genes for the fine regulation of the colour of grapes. *Journal of Experimental Botany* **66**, 4427-4440.

Dai Z, Hilbert G, Gomès E, Bobeica N, Poni S, Génard M, Delrot S,, 2017. Mathematic model for simulating anthocyanin composition during grape ripening:

another way of phenotyping. *Acta Horticulturae*, 375-380.

Dimitrovska M, Bocevska M, Dimitrovski D, Murkovic M, 2011. Anthocyanin composition of Vranec, Cabernet Sauvignon, Merlot and Pinot Noir grapes as indicator of their varietal differentiation. *European Food Research and Technology* **232**, 591-600.

Falginella L, Castellarin SD, Testolin R, Gambetta GA, Morgante M, Di Gaspero G, 2010. Expansion and subfunctionalisation of flavonoid 3',5'-hydroxylases in the grapevine lineage. *BMC Genomics* **11**, 562-562.

Faraji M, Fonseca LL, Escamilla-Treviño L, Barros-Rios J, Engle N, Yang ZK, Tschaplinski TJ, Dixon RA, Voit EO, 2018. Mathematical models of lignin biosynthesis. *Biotechnology for Biofuels* **11**, 34.

Faraji M, Voit EO, 2017. Improving bioenergy crops through dynamic metabolic modeling. *Processes* **5**, 61.

Farré G, Blancquaert D, Capell T, Van Der Straeten D, Christou P, Zhu C, 2014. Engineering complex metabolic pathways in plants. *Annual Review of Plant Biology* **65**, 187-223.

Ford CM, Boss PK, Høj PB, 1998. Cloning and characterization of vitis vinifera UDP-glucose: flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize bronze-1 locus that may primarily serve to glucosylate anthocyanidins in vivo. *Journal of Biological Chemistry* **273**, 9224-9233.

Forster MR, 2000. Key Concepts in Model Selection: Performance and Generalizability. *Journal of Mathematical Psychology* **44**, 205-231.

Fournier-Level A, Hugueney P, Verriès C, This P, Ageorges A, 2011. Genetic mechanisms underlying the methylation level of anthocyanins in grape (*Vitis vinifera* L.). *BMC Plant Biology* **11**, 179.

Génard M, Bertin N, Gautier H, Lescourret F, Quilot B, 2010. Virtual profiling: a new way to analyse phenotypes. *The Plant Journal* **62**, 344-355.

Guardiola J, Iborra JL, Cánovas M., 1995. A model that links growth and secondary metabolite production in plant cell suspension cultures. *Biotechnology and Bioengineering* **46**: 291-297.

He F, Mu L, Yan G-L, Liang N-N, Pan Q-H, Wang J, Reeves MJ, Duan C-Q, 2010. Biosynthesis of anthocyanins and their regulation in colored grapes. *Molecules* **15**, 9057-9091.

Hernández-Montes E, Zhang Y, Chang B-M, Shcherbatyuk N, Keller M, 2021. Soft, sweet, and colorful: Stratified sampling reveals sequence of events at the onset of grape ripening. *American Journal of Enology and Viticulture* **72**, 137-151.

Hilbert G, Soyer JP, Molot C, Giraudon J, Milin S, Gaudillere JP, 2003. Effects of nitrogen supply on must quality and anthocyanin accumulation in berries of cv. Merlot. *Vitis* **42**, 69-76.

Houghton A, Appelhagen I, Martin C, 2021. Natural blues: structure meets

function in anthocyanins. *Plants* **10**, 726.

Hugueney P, Provenzano S, Verriès C, Ferrandino A, Meudec E, Batelli G, Merdinoglu D, Cheynier V, Schubert A, Ageorges A, 2009. A novel cation-dependent O-methyltransferase involved in anthocyanin methylation in grapevine. *Plant Physiology* **150**, 2057-2070.

Iooss B, Lemaître P, 2015. A review on global sensitivity analysis methods. In: Dellino G,, Meloni C, eds. *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*. Boston, MA: Springer, 101–122.

Jaakola L, 2013. New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science* **18**, 477-483.

Keller M, Arnink KJ, Hrazdina G, 1998. Interaction of nitrogen availability during bloom and light intensity during veraison. I. Effects on grapevine growth, fruit development, and ripening. *American Journal of Enology and Viticulture* **49**, 333–340.

Keller M, Hrazdina G, 1998. Interaction of nitrogen availability during bloom and light intensity during veraison. II. Effects on anthocyanin and phenolic development during grape ripening. *American Journal of Enology and Viticulture* **49**, 341–349.

Khoo HE, Azlan A, Tang ST, Lim SM, 2017. Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food & Nutrition Research* **61**, 1361779.

Kong J, Wu J, Guan L, Hilbert G, Delrot S, Fan P, Liang Z, Wu B, Matus JT, Gomès E, Dai Z, 2021. Metabolite analysis reveals distinct spatio-temporal accumulation of anthocyanins in two teinturier variants of cv. 'Gamay' grapevines (*Vitis vinifera* L.). *Planta* **253**, 84.

Lee Y, Voit EO, 2010. Mathematical modeling of monolignol biosynthesis in *Populus* xylem. *Mathematical Biosciences* **228**, 78-89.

van Leeuwen C, Roby J-P, Alonso-Villaverde V, Gindro K, 2013. Impact of clonal variability in *Vitis vinifera* Cabernet franc on grape composition, wine quality, leaf blade stilbene content, and downy mildew resistance. *Journal of Agricultural and Food Chemistry* **61**, 19-24.

Lin-Wang K, McGhie TK, Wang M, Liu Y, Warren B, Storey R, Espley RV, Allan AC, 2014. Engineering the anthocyanin regulatory complex of strawberry (*Fragaria vesca*). *Frontiers in Plant Science* **5**, 651.

Liu Y, Tikunov Y, Schouten RE, Marcelis LFM, Visser RGF, Bovy A, 2018. Anthocyanin biosynthesis and degradation mechanisms in solanaceous vegetables: a review. *Frontiers in Chemistry* **6**, 52.

Liu Y, Qian J, Li J, Xing M, Grierson D, Sun C, Xu C, Li X, Chen K., 2022. Hydroxylation decoration patterns of flavonoids in horticultural crops: chemistry, bioactivity and biosynthesis. *Horticulture Research* **9**, uhab068.

Manela N, Oliva M, Ovadia R, Sikron-Persi N, Ayenew B, Fait A, Galili G, Perl A, Weiss D, Oren-Shamir M, 2015. Phenylalanine and tyrosine levels are rate-limiting factors in production of health promoting metabolites in *Vitis vinifera* cv. Gamay Red cell suspension. *Frontiers in Plant Science* **6**, 538-538.

Marshallcolon A, Long SP, Allen DK, Allen G, Beard DA, Benes B, Von Caemmerer S, Christensen AJ, Cox DJ, Hart JC, 2017. Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform. *Frontiers in Plant Science* **8**, 786.

Matthews ML, Wang JP, Sederoff R, Chiang VL, Williams CM, 2020. Modeling cross-regulatory influences on monolignol transcripts and proteins under single and combinatorial gene knockdowns in *Populus trichocarpa*. *PLoS Computational Biology* **16**, e1007197-e1007197.

Matthews ML, Wang JP, Sederoff R, Chiang VL, Williams CM, 2021. A multiscale model of lignin biosynthesis for predicting bioenergy traits in *Populus trichocarpa*. *Computational and Structural Biotechnology Journal* **19**, 168-182.

Mattivi F, Guzzon R, Vrhovsek U, Stefanini M, Velasco R, 2006. Metabolite profiling of grape: flavonols and anthocyanins. *Journal of Agricultural and Food Chemistry* **54**, 7692-7702.

Morgan JA, Rhodes D, 2002. Mathematical modeling of plant metabolic pathways. *Metabolic Engineering* **4**, 80-89.

Mori K, Goto-Yamamoto N, Kitayama M, Hashizume K, 2007. Loss of anthocyanins in red-wine grape under high temperature. *Journal of Experimental Botany* **58**, 1935-1945.

Morris MD, 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* **33**, 161-174.

Movahed N, Pastore C, Cellini A, Allegro G, Valentini G, Zenoni S, Cavallini E, D'Incà E, Tornielli GB, Filippetti I, 2016. The grapevine VviPrx31 peroxidase as a candidate gene involved in anthocyanin degradation in ripening berries under high temperature. *Journal of Plant Research* **129**, 513-526.

Noda N, Yoshioka S, Kishimoto S, Nakayama M, Douzono M, Tanaka Y, Aida R, 2017. Generation of blue chrysanthemums by anthocyanin B-ring hydroxylation and glucosylation and its coloration mechanism. *Science Advances* **3**, e1602785-e1602785.

Olsen KM, Slimestad R, Lea US, Brede C, Løvdal T, Ruoff P, Verheul M, Lillo C, 2009. Temperature and nitrogen effects on regulators and products of the flavonoid pathway: experimental and kinetic model studies. *Plant, Cell & Environment* **32**, 286-299.

Oren-Shamir M, 2009. Does anthocyanin degradation play a significant role in determining pigment concentration in plants? *Plant Science* **177**, 310-316.

Pastore C, Allegro G, Valentini G, Muzzi E, Filippetti I, 2017. Anthocyanin and flavonol composition response to veraison leaf removal on Cabernet Sauvignon, Nero d'Avola, Raboso Piave and Sangiovese *Vitis vinifera* L. cultivars. *Scientia*

Horticulturae **218**, 147-155.

Provenzano S, Spelt C, Hosokawa S, Nakamura N, Brugliera F, Demelis L, Geerke DP, Schubert A, Tanaka Y, Quattrocchio F, Koes R, 2014. Genetic control and evolution of anthocyanin methylation. *Plant physiology* **165**, 962-977.

Prudent M, Lecomte A, Bouchet J-P, Bertin N, Causse M, Génard M, 2011. Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. *Journal of Experimental Botany* **62**, 907-919.

R Core Team, 2013. R: A language and environment for statistical computing.

Rinaldo AR, Cavallini E, Jia Y *et al.*, 2015. A grapevine anthocyanin acyltransferase, transcriptionally regulated by VvMYBA, can Produce most acylated anthocyanins present in grape skins. *Plant Physiology* **169**, 1897-1916.

Rios-Estepa R, Turner GW, Lee JM, Croteau RB, Lange BM, 2008. A systems biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2818-2823.

de Rosas I, Deis L, Baldo Y, Cavagnaro JB, Cavagnaro PF, 2022. High temperature alters anthocyanin concentration and composition in grape berries of Malbec, Merlot, and Pinot Noir in a cultivar-dependent manner. *Plants* **11**, 926.

Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR, 2013. The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. *Plant Physiology and Biochemistry* **72**, 21-34.

Shimada Y, Ohbayashi M, Nakano-Shimada R, Okinaka Y, Kiyokawa S, Kikuchi Y, 2001. Genetic engineering of the anthocyanin biosynthetic pathway with flavonoid-3',5'-hydroxylase: specific switching of the pathway in petunia. *Plant Cell Reports* **20**, 456-462.

Sugiura T, Shiraishi M, Konno S, Sato A, 2018. Prediction of skin coloration of grape berries from air temperature. *The Horticulture Journal* **87**, 18-25.

Suter B, Destrac Irvine A, Gowdy M, Dai Z, van Leeuwen C, 2021. Adapting wine grape ripening to global change requires a multi-trait approach. *Frontiers in Plant Science* **12**, 624867.

Wallach D, Makowski D, Jones JW (Eds.), 2006. *Working with dynamic crop models: evaluation, analysis, parameterization, and applications*. Amsterdam: Elsevier.

Wallach D, Makowski D, Jones JW, Brun F, 2014. Model evaluation. *Working with Dynamic Crop Models*, 345-406.

Wang R, Lenka SK, Kumar V, Sikron-Persi N, Dynkin I, Weiss D, Perl A, Fait A, Oren-Shamir M, 2021. A synchronized increase of stilbenes and flavonoids in metabolically engineered *Vitis vinifera* cv. Gamay red cell culture. *Journal of Agricultural and Food Chemistry* **69**, 7922-7931.

Wang Y, Li PM, Yao LN, Shang YY, Liu S, Meng JX, ZHANG SY, Li HH, 2022. Advances in the application of biosynthesis and metabolic engineering of flavonoids

in plants. *Biologia Plantarum* **66**, 163-171.

Wang JP, Matthews ML, Naik PP, Williams CM, Ducoste JJ, Sederoff RR, Chiang VL, 2019. Flux modeling for monolignol biosynthesis. *Current Opinion in Biotechnology* **56**, 187-192.

Wang JP, Matthews ML, Williams CM, Shi R, Yang C, Tunlaya-Anukit S, Chen H-C, Li Q, Liu J, Lin C-Y, Naik P, Sun Y-H, Loziuk PL, Yeh T-F, Kim H, Gjersing E, Shollenberger T, Shuford CM, Song J, Miller Z, Huang Y-Y, Edmunds CW, Liu B, Sun Y, Lin Y-CJ, Li W, Chen H, Peszlen I, Ducoste JJ, Ralph J, Chang H-M, Muddiman DC, Davis MF, Smith C, Isik F, Sederoff R, Chiang VL, 2018. Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis. *Nature Communications* **9**, 1579-1579.

Wen W, Alseekh S, Fernie AR, 2020. Conservation and diversification of flavonoid metabolism in the plant kingdom. *Current Opinion in Plant Biology* **55**, 100-108.

Wheeler LC, Smith SD, 2019. Computational modeling of anthocyanin pathway evolution: biases, hotspots, and trade-offs. *Integrative and Comparative Biology* **59**, 585-598.

Wheeler LC, Wing BA, Smith SD, 2020. Structure and contingency determine mutational hotspots for flower color evolution. *Evolution letters* **5**, 61-74.

Zhang Y, Butelli E, Martin C, 2014. Engineering anthocyanin biosynthesis in plants. *Current Opinion in Plant Biology* **19**, 81-90.

Zhu X, Liu X, Liu T, Wang Y, Ahmed N, Li Z, Jiang H, 2021. Synthetic biology of plant natural products: From pathway elucidation to engineered biosynthesis in plant cells. *Plant Communications* **2**, 100229-100229.

Zipor G, Duarte P, Carqueijeiro I, Shahar L, Ovadia R, Teper□Bamnolker P, Eshel D, Levin Y, Doron□Faigenboim A, Sottomayor M, Oren-Shamir M, 2014. In planta anthocyanin degradation by a vacuolar class III peroxidase in *Brunfelsia calycina* flowers. *New Phytologist* **205**, 653-665.

Table 1. List of datasets that were used to develop the Dynamic Anthocyanin Composition model.

Dataset ID	Cultivar	Vintage	Conditions	Reference	Site
1	Cabernet Sauvignon	1995	9 conditions s: 3 nitrogen levels × 3 light levels	Keller et al. 1998	Field
2	Merlot	2001	3 conditions: 3 nitrogen levels	Hilbert et al. 2003	Greenhouse
3	Pinot noir	2009 2010 2011	12 conditions: 3 years × 2 rootstocks × 2 water supply levels	Berdeja et al. 2014	Field
4	Cabernet Sauvignon Sangiovese	2012 2013	4 conditions: 2 cultivars × 2 leaf-to-fruit ratio levels	Bobeica et al. 2015	Field & Greenhouse
5	Gamay Gamay de Bouze Gamay Freaux	2013	5 conditions: 2 cultivars × 2 tissues + 1 cultivars × 1 tissue	Kong et al. 2021	Greenhouse
6	Cabernet Sauvignon Tempranillo	2013 2014	4 conditions: 2 years× 2 cultivars	This research	Field

Table 2. Leave-one-out cross-validation results of dataset 1, 2 and 3

Dataset	Cultivar	Selected condition	Individual anthocyanins	RMSE	RMSEP	RRMSE	RRMSEP	R^2
1	Cabernet Sauvignon	L3_N1	5	14.3	17.9	23.5	29.4	0.77
1	Cabernet Sauvignon	L3_N2	5	15.3	16.7	23.3	25.4	0.84
1	Cabernet Sauvignon	L3_N3	5	10.9	10.7	19.6	19.3	0.92
1	Cabernet Sauvignon	L2_N1	5	9.5	10.4	16.9	18.4	0.93
1	Cabernet Sauvignon	L2_N2	5	8.4	7.6	18.2	16.4	0.96
1	Cabernet Sauvignon	L2_N3	5	7.8	8.3	20	21.3	0.94
1	Cabernet Sauvignon	L3_N1	5	9.2	7.6	28.8	23.9	0.96
1	Cabernet Sauvignon	L3_N2	5	11.1	10.7	43.2	41.5	0.91
1	Cabernet Sauvignon	L3_N3	5	9	8.5	44.9	41.9	0.92
mean	**Cabernet Sauvignon**		**5**	**10.6**	**10.9**	**26.5**	**26.4**	**0.91**
2	Merlot	N1	13	7.1	8.4	18.4	24.8	0.97
2	Merlot	N2	13	7	9.8	28.8	46.4	0.93
2	Merlot	N3	13	5.3	5.7	22.4	27.6	0.97
mean	**Merlot**		**13**	**6.5**	**8.0**	**23.2**	**32.9**	**0.96**
3	Pinot noir	2009_110R_WS	5	43.5	54.4	24.6	30.8	0.98
3	Pinot noir	2009_110R_CK	5	22.9	32.3	20.2	28.5	0.98
3	Pinot noir	2009_125AA_WS	5	40.4	45.5	25.9	29.2	0.98
3	Pinot noir	2009_125AA_CK	5	37.9	46.3	31.7	38.8	0.98
3	Pinot noir	2010_110R_WS	5	75.8	80.7	29.4	31.3	0.98
3	Pinot noir	2010_110R_CK	5	21.5	22.1	13.8	14.2	0.99
3	Pinot noir	2010_125AA_WS	5	26.1	24.8	12.8	12.2	0.99

3	Pinot noir	2010_125AA_CK	5	17.9	18.0	15.6	15.7	0.99
3	Pinot noir	2011_110R_WS	5	39.5	36.5	29.4	27.2	0.99
3	Pinot noir	2011_110R_CK	5	37.8	49.4	33.6	43.8	0.98
3	Pinot noir	2011_125AA_WS	5	41.7	41.6	38.7	38.6	0.98
3	Pinot noir	2011_125AA_CK	5	32.9	32.0	36.3	35.3	0.99
mean	**Pinot noir**		**5**	**36.5**	**40.3**	**26.0**	**28.8**	**0.99**

Note of Table 2: Dataset 1 included 9 conditions resulting from the combinations of 3 soil nitrogen supply levels (from low to high: N1, N2 and N3) and 3 canopy light levels (from low to high: L1, L2, and L3). Dataset 2 included 3 soil nitrogen supply levels (from low to high: N1, N2 and N3). Dataset 3 included 12 conditions resulting from the combinations of 2 rootstocks (110R and 125AA), 2 water supply levels (CK: rainfed condition, WS: water stress) and 3 vintages (2009, 2010, 2011). RMSE, RRMSE: root mean square error and relative root mean square error of all anthocyanins in model calibration. RMSEP, RRMSEP: root mean square error and relative root mean square error of each individual anthocyanin in model validation.

Fig. 1 A: Biochemical decoration patterns of hydroxylation, methylation and acylation of anthocyanins in grape berry (*Vitis vinifera* L.). This figure was adapted from Fournier-Level *et al.* (2011). B: Schematic diagram of the anthocyanin biosynthesis pathway in grapevine. Red color represents the metabolic branch leading to cyanidin-based anthocyanins; purple color represents the metabolic branch leading to delphinidin-based anthocyanins. Arrows and boxes represent the anthocyanin fluxes and individual anthocyanins, respectively. Solid and dashed lines with arrows represent anthocyanin conversion and degradation, respectively. V_{in} is the total anthocyanin influx and is the sum of two branch influxes (V_{Cy} and V_{Dp}). ri (i from 1 to 13) is the relative rate of transformation from one anthocyanin to the arrowhead anthocyanin. *kd* represents the relative degradation rate of anthocyanins. The set-up of each parameter is explained in Table S1.

Fig. 2. The ratio of di- to tri-hydroxylated (RDT) anthocyanins during grape ripening in various genotypes and environments. A-F correspond to datasets 1-6, respectively. Line colors, symbols and line types represent different cultivars and growth conditions in each dataset. The first dataset (A) shows the effects of nitrogen availability (0.34, 1.7, or 3.4 g N per plant named N1, N5, N10, respectively) and light intensity (100%, 20%, or 2% of full sunlight named L100, L20, L2 respectively) in Cabernet Sauvignon. The second dataset (B) shows 3 nitrogen treatments at 1.4 mM, 3.6 mM and 7.2 mM N (denominated N1, N2, N3, respectively), applied from fruit set to leaf fall in Merlot. The third dataset (C) shows Pinot noir grafted on either rootstock 110R (drought tolerant, edium to high vigour) or 125AA (drought sensitive, high vigour) during 3 growing seasons (2009-2011) in the field under normal rainfall (CK) or water shortage (WS). The fourth dataset (D) shows two leaf-to-fruit ratio levels (3L: 3 leaves per cluster, 12L: 12 leaves per cluster) to berries of Cabernet Sauvignon and Sangiovese. The fifth dataset (E) shows the skin and pulp of Gamay (G), Gamay de Bouze (GB) and Gamay Freaux (GF) berries collected from vines grown in a greenhouse. The sixth dataset (F) shows Cabernet Sauvignon and Tempranillo, in two growing seasons (2013, 2014). The inserts in (D) and (E) are zoom-in of genotypes with RDT < 1.

Fig. 3. Comparison between observed and simulated concentrations of individual anthocyanins for 2 cultivars in 2 vintages (dataset 6). For each condition, two types of figures are used to compare the observed and simulated results: one shows the developmental profiles of each individual anthocyanin, with symbols for the observed and lines for simulated values; the other shows the correlation between the observed and simulated concentrations of all individual anthocyanins with the 1:1 line, as well as the goodness-of-fit criteria RMSE, RRMSE and R^2.

Symbol and line colors represent different anthocyanins; cyanidin-based anthocyanins are represented by warm colors and delphinidin-based anthocyanins are represented by cool colors. Abbreviations of individual anthocyanins are the same as in Fig. 1A. Each point represents the mean of 3 biological replicates.

Fig. 4. Comparison between observed and simulated concentrations of 4 anthocyanin types (di: di-hydroxylated, tri: tri-hydroxylated, meth: methylated and nometh: unmethylated) for 2 cultivars in 2 vintages (dataset 6). For each condition, two types of figures were used to compare observed and simulated results: one shows the developmental profiles of each anthocyanin type, with symbols for the observed and lines for simulated values; the other shows the correlation between the observed and simulated concentrations of the 4 anthocyanin types with the 1:1 line, as well as the goodness-of-fit criteria RMSE, RRMSE and R^2.

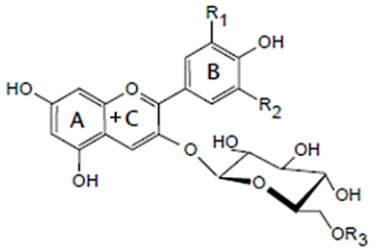
Fig. 5. Model parameter sensitivity for each individual anthocyanin and anthocyanin decoration categories at maturity in 4 grape cultivars (A Pinot noir, B Sangiovese, C Gamay Freaux skin, D Cabernet Sauvignon). Sensitivity index is represented by color range (white: 0, red:1) as indicated in the color key. For the x-axis labels, parameters related to degradation (a_{kd}, b_{kd}), allocation coefficient (a_δ, b_δ, c_δ) and conversion rate (r_i) are highlighted by red, blue and black, respectively. Abbreviations of individual anthocyanins are as in Fig. 2. The ratios of di- to tri-hydroxylated anthocyanins, methylated to unmethylated anthocyanins, acylated to unacylated anthocyanins are represented by 'h_ratio', 'm_ratio' and 'ac_ratio', respectively.

Fig. 6. The impact of manipulating model parameters on the proportion of tri-hydroxylated and methylated anthocyanins was explored in Sangiovese and Cabernet Sauvignon grapes. The sensitivity analysis identified c_δ as the sole sensitive parameter for the proportion of tri-hydroxylated anthocyanins in both cultivars. Additionally, the top 7 sensitive parameters, namely a_{kd}, b_{kd}, b_δ, c_δ, r_1, r_2, and r_3 for Sangiovese, and a_{kd}, b_{kd}, c_δ, r_1, r_3, r_{11}, and r_{12} for Cabernet Sauvignon, were identified as influencing the proportion of methylated anthocyanins. Modulating these parameters around their default values by -10%, -5%, 0%, +5%, and +10% while keeping other initial values, inputs, and parameters in the default condition resulted in 78,125 possible combinations of the 7 parameters. The simulated proportions of tri-hydroxylated and methylated anthocyanins are expressed as the percentage variation from the default values. The distributions of alterations in the proportions of tri-hydroxylated and methylated anthocyanins are depicted with 'jitter plot', which means adding a minor random 'noise' around

the 6 values of the proportions of tri-hydroxylated anthocyanins to improve the visualization of the proportions of methylated anthocyanins (A, B). The jitter plots (C, D) show the variations in the proportion of methylated anthocyanins as a function of the number of altered parameters for all combinations (black points) and the combinations with optimal proportion of tri-hydroxylated anthocyanins (orange points). The trends of the optimal solution as a function of the number of altered parameters for all combinations (black line) and the combinations with optimal proportion of tri-hydroxylated anthocyanins (orange line) are shown (C, D). Moreover, 7 local optimal combinations (labeled as max1 to max7) were selected by simultaneously considering the number of altered parameters, high proportion of tri-hydroxylated anthocyanins, and high proportion of methylated anthocyanins (C, D). The details of these 7 combinations are shown for their parameters and their h_ratio and m_ratio (E, F). The proportions of individual anthocyanins under the 7 selected parameter combinations are also illustrated (G, H).

Figure 1

A



R_3 = ① -H

② -CO-CH_3 (-acetyl)

③ -CO-CH=CH-  OH (-*p*-coumaroyl)

Anthocyanidin	R_1	R_2	R_3	Abbreviation	Hydroxylation	Methylation	Acylation
Cyanidin	OH	H	①	Cy_{glc}	dihydroxylated	unmethylated	unacylated
Delphinidin	OH	OH	①	Dp_{glc}	trihydroxylated	unmethylated	unacylated
Peonidin	OCH_3	H	①	Pn_{glc}	dihydroxylated	methylated	unacylated
Petunidin	OCH_3	OH	①	Pt_{glc}	trihydroxylated	methylated	unacylated
Malvidin	OCH_3	OCH_3	①	Mv_{glc}	trihydroxylated	methylated	unacylated
Cyanidin	OH	H	②	Cy_{ac}	dihydroxylated	unmethylated	acylated
Delphinidin	OH	OH	②	Dp_{ac}	trihydroxylated	unmethylated	acylated
Peonidin	OCH_3	H	②	Pn_{ac}	dihydroxylated	methylated	acylated
Petunidin	OCH_3	OH	②	Pt_{ac}	trihydroxylated	methylated	acylated
Malvidin	OCH_3	OCH_3	②	Mv_{ac}	trihydroxylated	methylated	acylated
Cyanidin	OH	H	③	Cy_{cou}	dihydroxylated	unmethylated	acylated
Delphinidin	OH	OH	③	Dp_{cou}	trihydroxylated	unmethylated	acylated
Peonidin	OCH_3	H	③	Pn_{cou}	dihydroxylated	methylated	acylated
Petunidin	OCH_3	OH	③	Pt_{cou}	trihydroxylated	methylated	acylated
Malvidin	OCH_3	OCH_3	③	Mv_{cou}	trihydroxylated	methylated	acylated

B

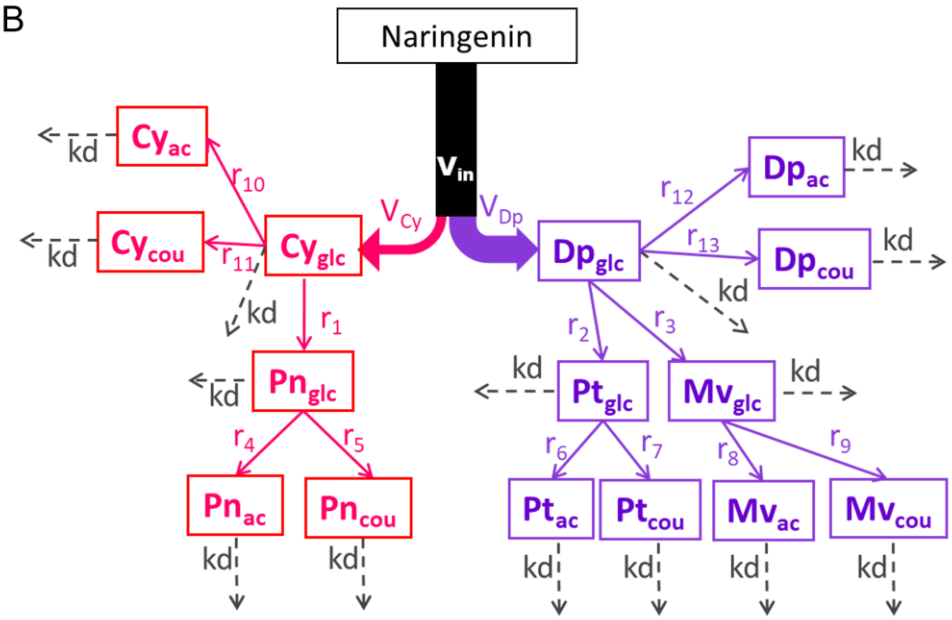


Figure 2

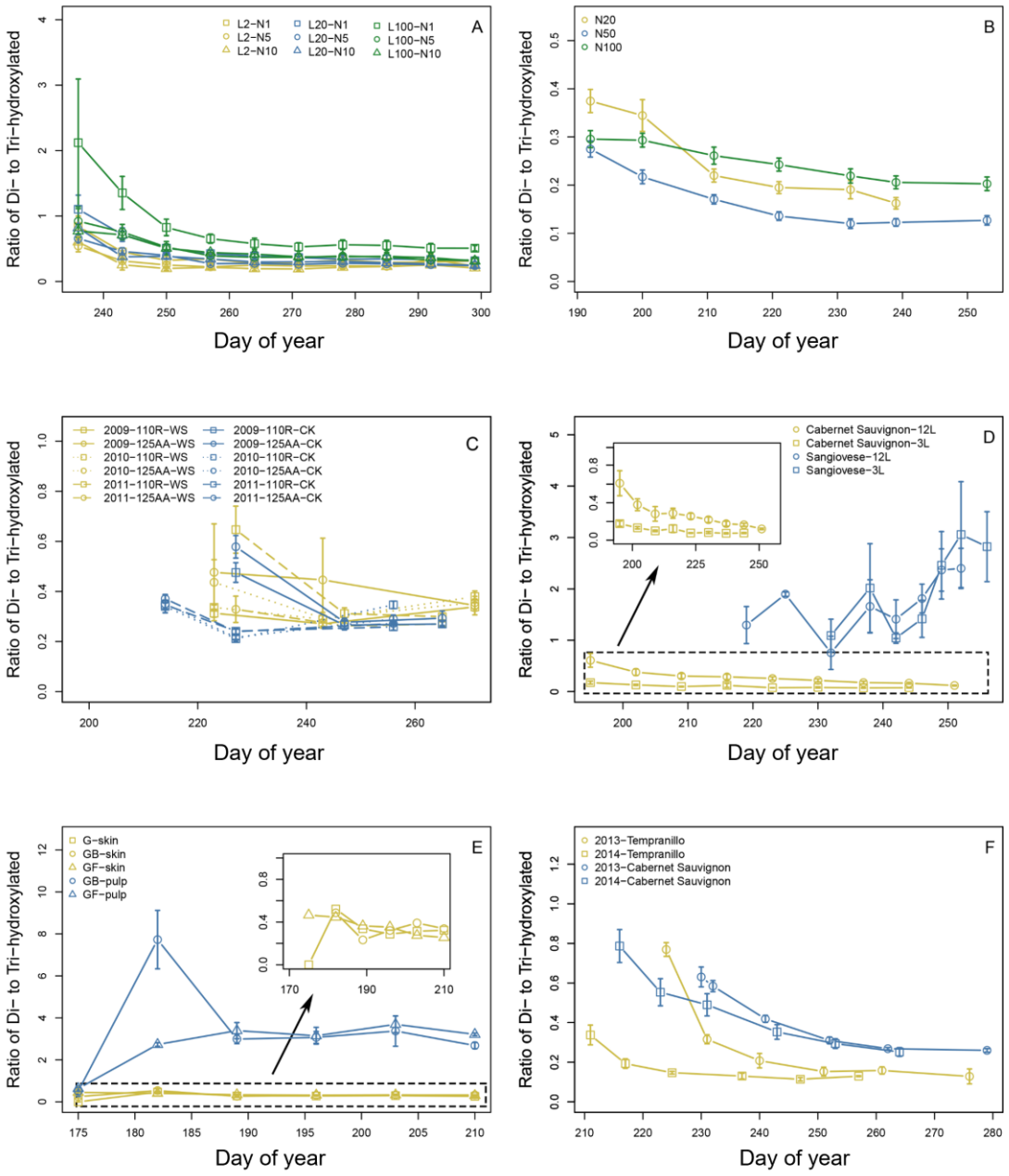


Figure 3

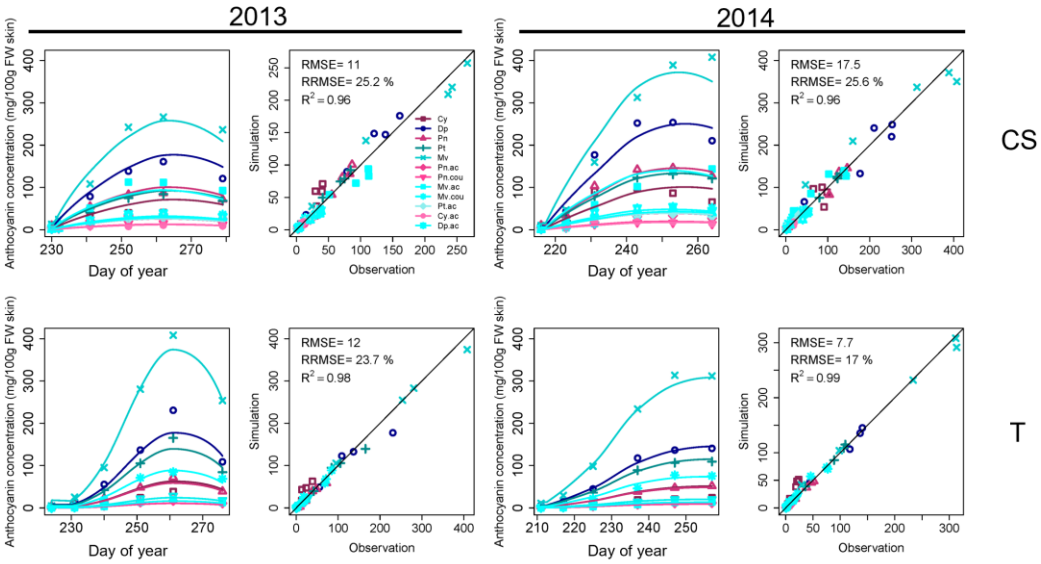


Figure 4

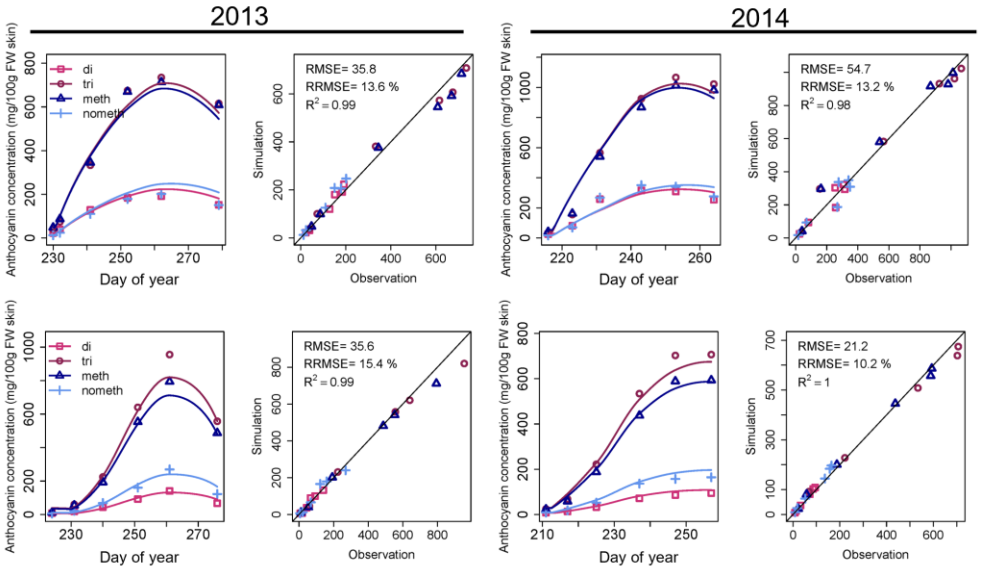


Figure 5

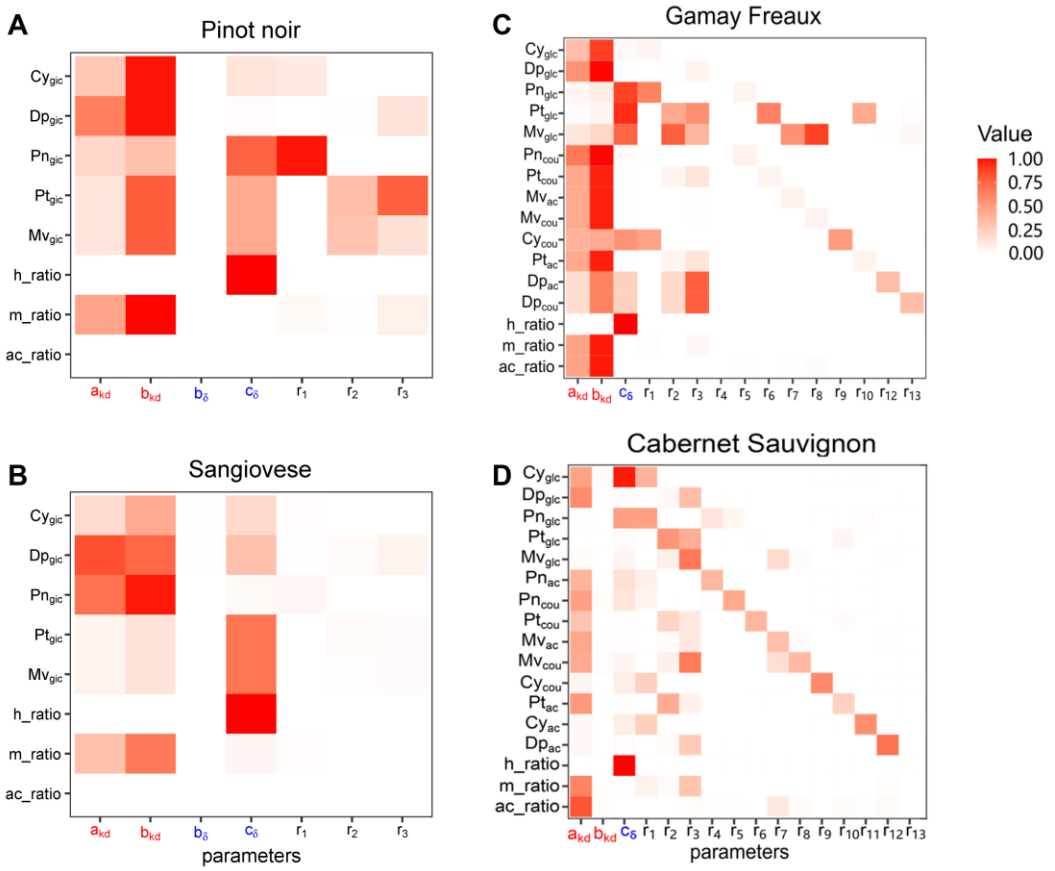
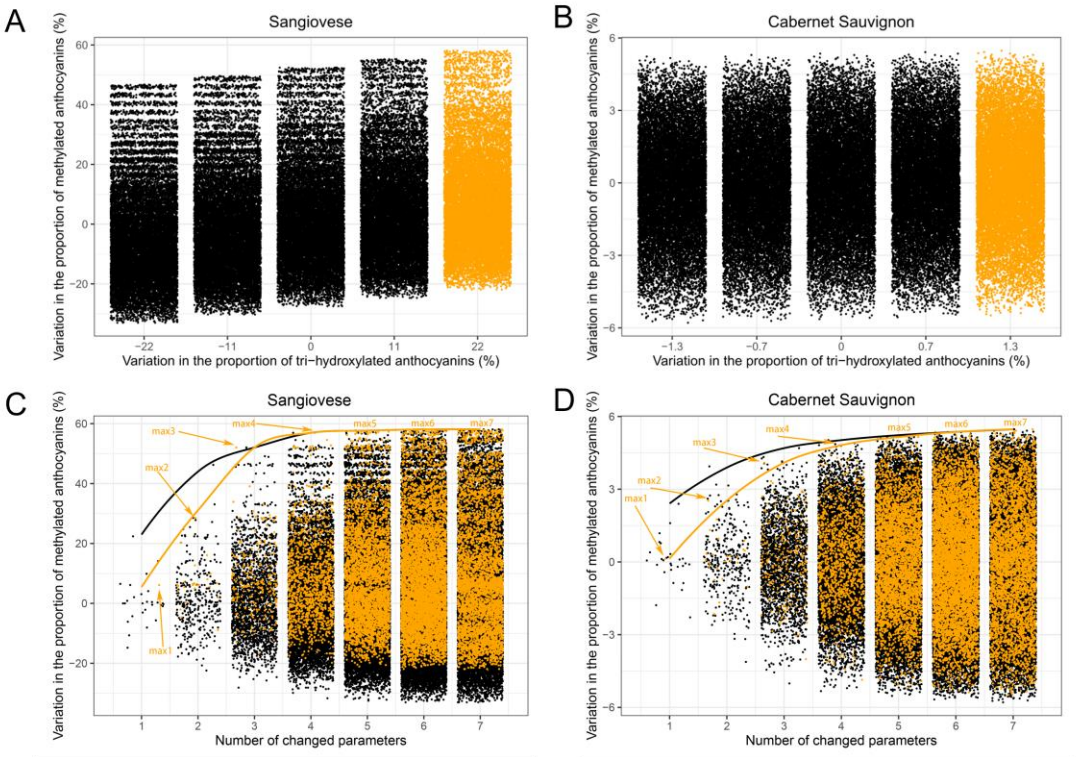


Figure 6



E

	max1	max2	max3	max4	max5	max6	max7
a_{kd}	0%	0%	-10%	-10%	-10%	-10%	-10%
b_{kd}	0%	-10%	-10%	-10%	-10%	-10%	-10%
b_{δ}	0%	0%	0%	0%	0%	0%	-10%
c_{δ}	-10%	-10%	-10%	-10%	-10%	-10%	-10%
r_1	0%	0%	0%	+10%	+10%	+10%	+10%
r_2	0%	0%	0%	0%	0%	+10%	+10%
r_3	0%	0%	0%	0%	+10%	+10%	+10%
h_ratio	+22%	+22%	+22%	+22%	+22%	+22%	+22%
m_ratio	+6%	+29%	+52%	+57%	+58%	+58%	+58%

F

	max1	max2	max3	max4	max5	max6	max7
a_{kd}	0%	-10%	-10%	-10%	-10%	-10%	-10%
b_{kd}	0%	0%	0%	0%	0%	-10%	-10%
c_{δ}	-10%	-10%	-10%	-10%	-10%	-10%	-10%
r_1	0%	0%	0%	+10%	+10%	+10%	+10%
r_3	0%	0%	+10%	+10%	+10%	+10%	+10%
r_{11}	0%	0%	0%	0%	0%	0%	-10%
r_{12}	0%	0%	0%	0%	-10%	-10%	-10%
h_ratio	+1.3%	+1.3%	+1.3%	+1.3%	+1.3%	+1.3%	+1.3%
m_ratio	+0.1%	+2.5%	+4.1%	+4.8%	+5.1%	+5.4%	+5.5%