LARGE-SCALE BIOLOGY ARTICLE

A diel multi-tissue genome-scale metabolic model of *Vitis vinifera*

Marta Sampaio[1*], Miguel Rocha[1,2], Oscar Dias[1,2*]

[1]Centre of Biological Engineering, University of Minho, Campus of Gualtar, 4710-057, Braga, Portugal and

[2]LABBELS, Associate Laboratory, Braga/Guimarães, Portugal

∗Corresponding author: msampaio@ceb.uminho.pt; odias@ceb.uminho.pt

Abstract

Vitis vinifera, also known as grapevine, is widely cultivated and commercialized, particularly to produce wine. As wine quality is directly linked to fruit quality, studying grapevine metabolism is important to understand the processes underlying grape composition. Genome-scale metabolic models (GSMMs) have been used for the study of plant metabolism and advances have been made, allowing the integration of omics datasets with GSMMs. On the other hand, Machine learning (ML) has been used to analyze omics data, and while the combination of ML with GSMMs has shown promising results, it is still scarcely used to study plants. Here, the first GSSM of *V. vinifera* was reconstructed and validated, comprising 7199 genes, 5399 reactions, and 5141 metabolites across 8 compartments. Tissue-specific models for stem, leaf, and berry of the Cabernet Sauvignon cultivar were generated from the original model, through the integration of RNA-Seq data. These models have been merged into diel multi-tissue models to study the interactions between tissues at light and dark phases. The potential of combining ML with GSMMs was explored by using ML to analyze the fluxomics data generated by green and mature grape GSMMs, helping to understand the factors influencing grape quality at different developmental stages.

Introduction

Vitis vinifera is one of the major fruit crops in the world. It is cultivated worldwide and has high economic value, mainly due to wine production. In 2022, the world vineyard surface area was estimated to be 7.3 million hectares and world wine production reached 258 million hectoliters. In the same year, and despite inflation, wine exports reached a value of 37.6 billion euros (International Organisation of Vine and Wine, 2023). In addition, grapes have other purposes, being marketed as fresh and dried fruits, and used for juice production. The grape pulp contains high levels of sugars and phenolic compounds, like flavonoids and stilbenes, with potential health benefits, such as antioxidant and anti-inflammatory activities, and cardiovascular protection (Saad et al., 2020), thus being currently studied for possible pharmaceutical and cosmetic applications. Therefore, as grapevines have high economic interest and fruit quality is intrinsically linked to metabolism, the study of grapevine metabolism is essential for understanding its responses to different environmental conditions that may affect grape metabolic composition.

Genome-scale metabolic models (GSMMs) represent all metabolic reactions taking place within an organism. These models are reconstructed from the genome and allow performing phenotype predictions under different environmental or genetic conditions (Feist et al., 2008). Although GSMMs have been extensively used for the metabolic engineering of prokaryotes, several GSMMs are available for plants (Gu et al., 2019), mainly *Arabidopsis thaliana* (Poolman et al., 2009; Dal'Molin et al., 2010; Saha et al., 2011; Cheung et al., 2013; Maurice Cheung et al., 2014; de Oliveira Dal'Molin et al., 2015; Shaw and Cheung, 2018), *Zea mays* (Saha et al., 2011; Simons et al., 2014; Bogart and Myers, 2016), and *Oryza sativa* (Poolman et al., 2013; Lakshmanan et al., 2015; Chatterjee et al., 2017). Currently, the reconstruction of plant GSMMs is still very challenging and time-consuming due to the high number of gaps in genome annotations, the large diversity of metabolites, and the extensive compartmentalization of plant cells (Collakova et al., 2012; Sweetlove and George Ratcliffe, 2011; Sampaio et al., 2022). Despite the obstacles, many plant GSMMs have emerged recently, and new approaches have been developed to reconstruct more realistic models that include different plant tissues, through the integration of omics data, as well as the day-night

cycle (Maurice Cheung et al., 2014; Gomes de Oliveira Dal'Molin and Nielsen, 2018; Shaw and Cheung, 2020). These models allow for differentiating the metabolism of each tissue and analyzing the metabolic interactions between tissues and the light and dark phases.

Despite the existence of several methods for integrating omics into GSMMs, this is still a challenging and inefficient task. As omics datasets are complex and heterogeneous, Machine Learning (ML) has been used to process and integrate different types of omics to extract biological knowledge from data. Recently, ML and GSMM approaches have been combined to improve the model's predictions and interpretability, and this strategy has shown promising results (Zampieri et al., 2017; Rana et al., 2020; Antonakoudis et al., 2020; Sampaio et al., 2022; Kim et al., 2021). ML can be used to extract knowledge from the fluxomics data generated by the models or to integrate the predicted fluxomics data with experimental omics. Thus far, these studies have mainly been applied to bacteria, yeast, and human cells, but not to plants.

In this manuscript, we pioneer *V. vinifera* research with the introduction of iMS7199, the first GSMM for the grapevine, developed using the most recent genome version, PN40024.v4 (Velt et al., 2023). In addition to the overarching model, tissue-specific models for the leaf, stem, and grape were developed by incorporating RNA-Seq data from these distinct tissues. Furthermore, to capture the dynamic changes in grape metabolism, we created two separate models representing the grape in both its green and mature states. These tissue-specific models were then integrated to construct diel multi-tissue GSMMs, enabling the simulation of grapevine metabolism across the day-night cycle and facilitating the study of inter-tissue metabolic interactions. Utilizing this comprehensive model, we investigated the metabolic responses of the grapevine under varying concentrations of sulfate and nitrate.

Also, simulated fluxomics data were generated from GSMMs of grapes in the green and mature state and analyzed by ML to identify the reactions that most contribute to the model's predictions of the grape developmental phase.

Therefore, this diel multi-tissue GSMM emerges as a useful tool for exploring the metabolic behavior of *V. vinifera* under various conditions, offering insights into factors

influencing grape quality and phenolic content. In addition, the analysis of generated data from GSMMs by ML represents the first effort to apply this strategy in the study of plant metabolism.

Results and Discussion

The iplants repository

To collect and organize all the relevant data for the model reconstruction efforts, a repository with the metabolic information of PlantCyc 14.0 (Zhang et al., 2010) and MetaCyc 26.1 (Caspi et al., 2016) databases, and Universal Protein Resource (UniProt) (The UniProt Consortium, 2017) sequence data was created. In total, the repository includes 24333 metabolites, 205128 reactions, 3519 pathways, and 22433 enzymes, 72% of which have a protein sequence. The Neo4j database includes the relationships between the metabolic entities, while MongoDB includes all the metadata that characterizes the entities. Details on how data is organized in the iplants repository are available in Supplementary Material and Supplementary File 1.

In addition to data from the metabolic databases, nine plant metabolic models were integrated into the iplants repository, namely *Arabidopsis thaliana* (Poolman et al., 2009; Cheung et al., 2013), *Zea mays* (Bogart and Myers, 2016), *Oryza sativa* (Poolman et al., 2013; Chatterjee et al., 2017), *Solanum lycopersicum* (Yuan et al., 2016), *Medicago truncatula* (Pfau et al., 2018), *Glycine max* (Moreira et al., 2019), and *Setaria viridis* (Shaw and Maurice Cheung, 2019). These models have PlantCyc and MetaCyc identifiers for metabolites and reactions, which facilitated the integration. In total, 3815 metabolites and 4197 reactions from the models were successfully integrated. However, around 395 metabolites and 1498 reactions from the metabolic models did not match any entry in our database and were added to it. These can include biomass and transporter reactions, whose identifiers are not standardized, or entities with deprecated identifiers that were already removed from the PlantCyc and MetaCyc databases.

iplants repository can be accessed through an application programming interface (API) created with Django and Django REST framework for both database systems, using

Mongoengine and Neomodel Python packages. Several views were defined to allow the extraction of the data needed for the reconstruction of GSMMs and to save the data of the model under reconstruction. The API included views to get all objects in the repository, details of an object, and to create a new metabolic model object and link it to reactions, metabolites, and enzymes in the database (link to API).

Model properties

A GSMM for *V. vinifera* was reconstructed from the PN40024.v4 genome (annotation version 1) (Velt et al., 2023). DIAMOND similarity searches (Buchfink et al., 2014) against iplants resulted in 10840 protein matches, representing 26% of the 41160 proteins in the genome, which is in line with the percentage of metabolic genes described for the *A. thaliana*'s genome (between 25-30%) (Kaul et al., 2000).

The reconstructed generic model, iMS7199, includes 5399 reactions (1624 transporters and 244 exchanges), and 5141 metabolites, across eight compartments: cytosol, chloroplast, mitochondria, endoplasmic reticulum, peroxisome, Golgi apparatus, vacuole, and extracellular space. In this model, the Gene-Protein-Reaction (GPR) rules were defined using the genome protein identifiers instead of genes as genome annotation was performed using protein sequences. As genes can encode more than one protein, the model includes 7199 protein identifiers that represent the 6018 genes of the *V. vinifera* genome.

This model is mass-balanced and can simulate growth in phototrophic and heterotrophic conditions, by setting the photon and carbon dioxide or sucrose as the only energy or carbon source, respectively. In addition, it requires the uptake of nitrate, phosphate, sulfate, iron, magnesium, and water to produce biomass.

The statistics of the *V. vinifera* model, as well as other relevant plant models, are presented in Table 1. Analyzing the table, only the *Quercus suber* model (Cunha et al., 2023b) has more genes, reactions, and metabolites than the *V. vinifera* model. The other models are much smaller, even the *G. max* model, which has a high number of genes.

Table 1. Statistics of the *V. vinifera* model and other eight plant GSMMs.

	Reactions	Metabolites	Genes	Compartments
A. thaliana (Cheung et al., 2013)	2769	2739	2857	5
Z. mays (Bogart et al., 2016)	1268	1121	2140	8
O. sativa (Chatterjee et al., 2017)	1136	1330	3602	4
S. lycopersicum (Yuan et al., 2015)	2143	1998	3410	5
M. truncatula (Pfau et al., 2018)	2909	2780	3403	8
G. max (Moreira et al., 2019)	3001	2814	6127	5
S. viridis (Shaw et al. 2019)	2473	2429	3376	5
Q. suber (Cunha et al., 2021)	6230	6481	7871	8
V. vinifera (this work)	5399	5141	7199	8

The reactions of *V. vinifera* were compared with those from the other models, except for *Q. suber* which has different model identifiers. Drains, transporters, biomass pseudo-reactions, and compartments were not considered, resulting in 2769 reactions of the iMS7199 model (Figure 1).

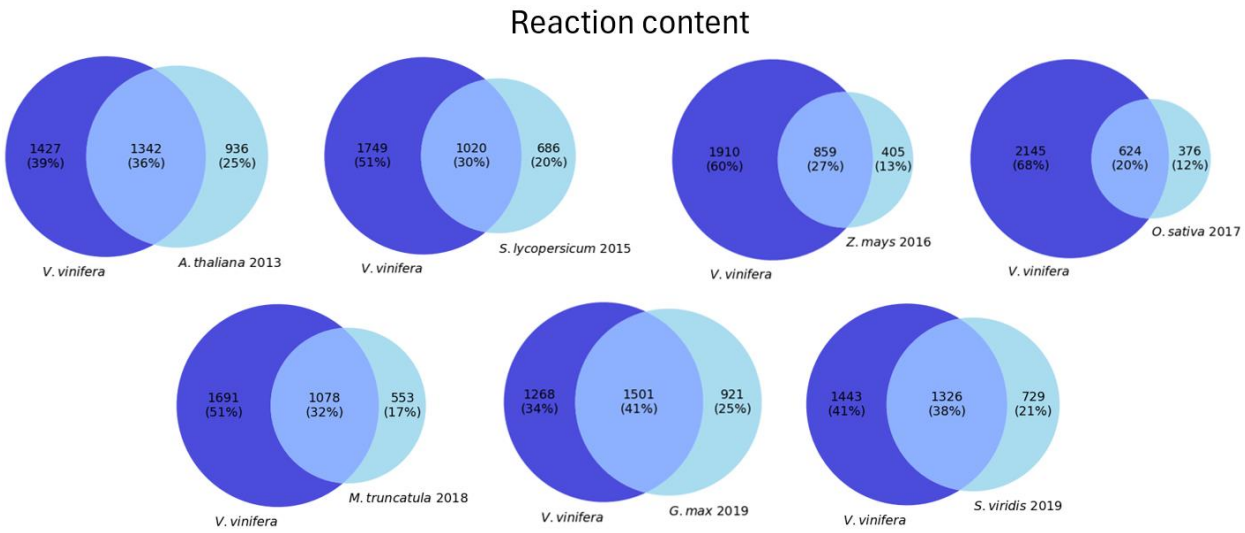


Figure 1. Venn diagrams comparing the reaction content of *V. vinifera* model with other seven plant models.

The *G. max* model shares the highest number of reactions with the *V. vinifera* model, corresponding to around 41% of all reactions from both models. This model is followed by the ones of *S. viridis and A. thaliana*, which share 1326 (38%) and 1342 (36%) reactions with iMS7199, respectively. The most distant model is the one from *O. sativa*, sharing only 624 reactions (20%).

In total, *V. vinifera* has 785 reactions that are not present in any other model. These reactions were analyzed to identify the associated pathways and gene annotation. The pathways with more unique reactions are presented in Figure 2. Reactions without pathway associations were not considered.

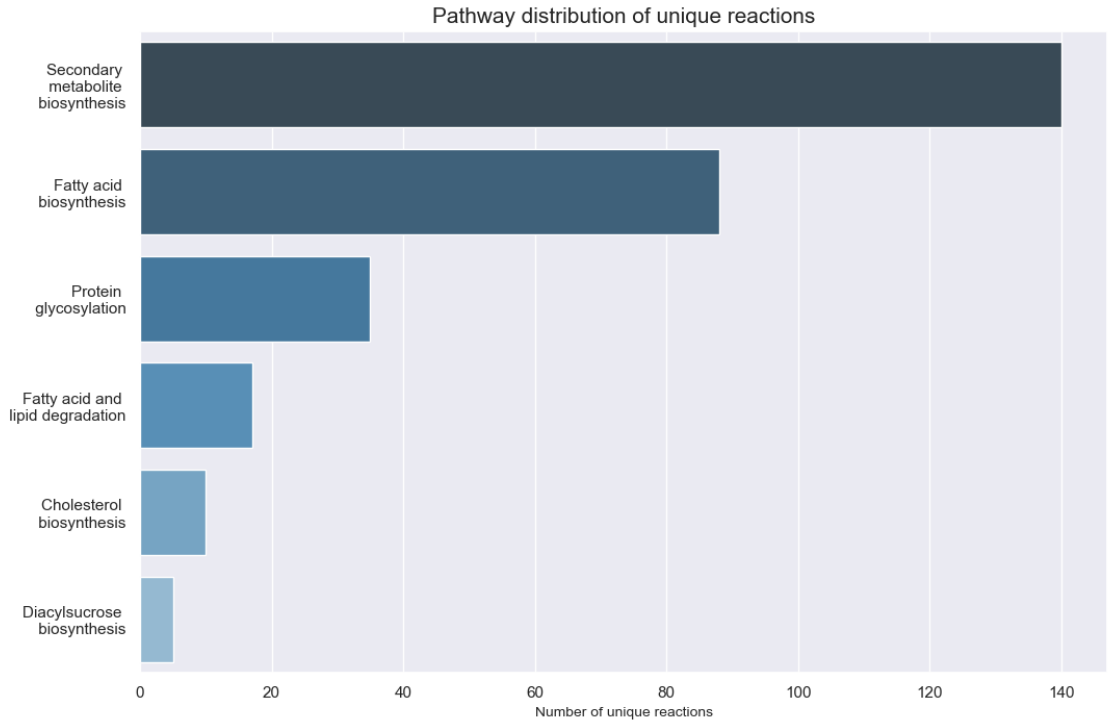


Figure 2. Pathway distribution of reactions included in the *V. vinifera* model and not in the other plant models analyzed.

As shown in Figure 2, the biosynthesis of secondary metabolites is the pathway class associated with more unique reactions, around 140, followed by fatty acid biosynthesis (88 reactions), protein glycosylation (35 reactions), and fatty acid and lipid degradation (17 reactions). These pathway classes comprise several specific pathways. Other specific pathways with more than four unique reactions include cholesterol biosynthesis

and diacylsucrose biosynthesis. Hence, the *V. vinifera* model represents a great advance compared to the previous plant models, as it comprises new reactions, especially for the secondary metabolism, which is often underrepresented in plant models.

On the other hand, the other plant models include 661 secondary pathway reactions not available in iMS7199. This number can be explained by the fact that 380 reactions are not associated with GPR rules in the models. Moreover, 93 are sub-reactions of others that are included in the *V. vinifera* model. Of the 188 reactions that have GPR rules, 155 have a corresponding enzyme sequence in the iplants database, which had a match in the DIAMOND annotation but were not the first hit for any query protein (Supplementary File 2). This data is made available and may be used in the future to improve the model by performing further manual curation. Regarding genome annotation, 27% of the proteins that catalyze these unique reactions were annotated based on the genome of *A. thaliana*. These proteins or reactions were probably not in metabolic databases when the *A. thaliana* models were reconstructed, which can explain why they are missing from these models. Besides *A. thaliana*, 12% of the unique proteins matched human proteins, and around 27% were annotated based on proteins from more than 100 different plant species, including *S. lycopersicum*, *Solanum tuberosum*, *Catharanthus roseus*, *Petunia x hybrida*, *M. truncatula*, *G. max*, and *V. vinifera*.

As *A. thaliana* is a reference organism for plants, there are several GSMMs for this organism (Poolman et al., 2009; Dal'Molin et al., 2010; Saha et al., 2011; Cheung et al., 2013; Maurice Cheung et al., 2014; de Oliveira Dal'Molin et al., 2015; Shaw and Cheung, 2018) and much enzymatic and metabolic information of *A. thaliana* is available in databases, like PlantCyc and UniProt. On the other hand, data for more complex plants is scarce. Therefore, it was expected that a large percentage of *V. vinifera* proteins would be annotated based on homologous proteins from *A. thaliana*. However, as *V. vinifera* is a much more complex plant, gene annotations can be wrong or missing, and the consequent validation process helps to limit these errors. In addition, several proteins were similar to human proteins, which was also expected as

various pathways, mainly related to lipid metabolism, are better characterized in humans than in plants.

Specialized metabolic pathways

Secondary metabolites are economically very important as they have many relevant applications. However, the pathways that produce them are very complex and diverse, and the knowledge in this subject is still limited (Collakova et al., 2012). Figures Figure ***3*** and Figure ***4*** schematize the production of the main secondary metabolites in the model, phenylpropanoids and terpenoids, respectively.

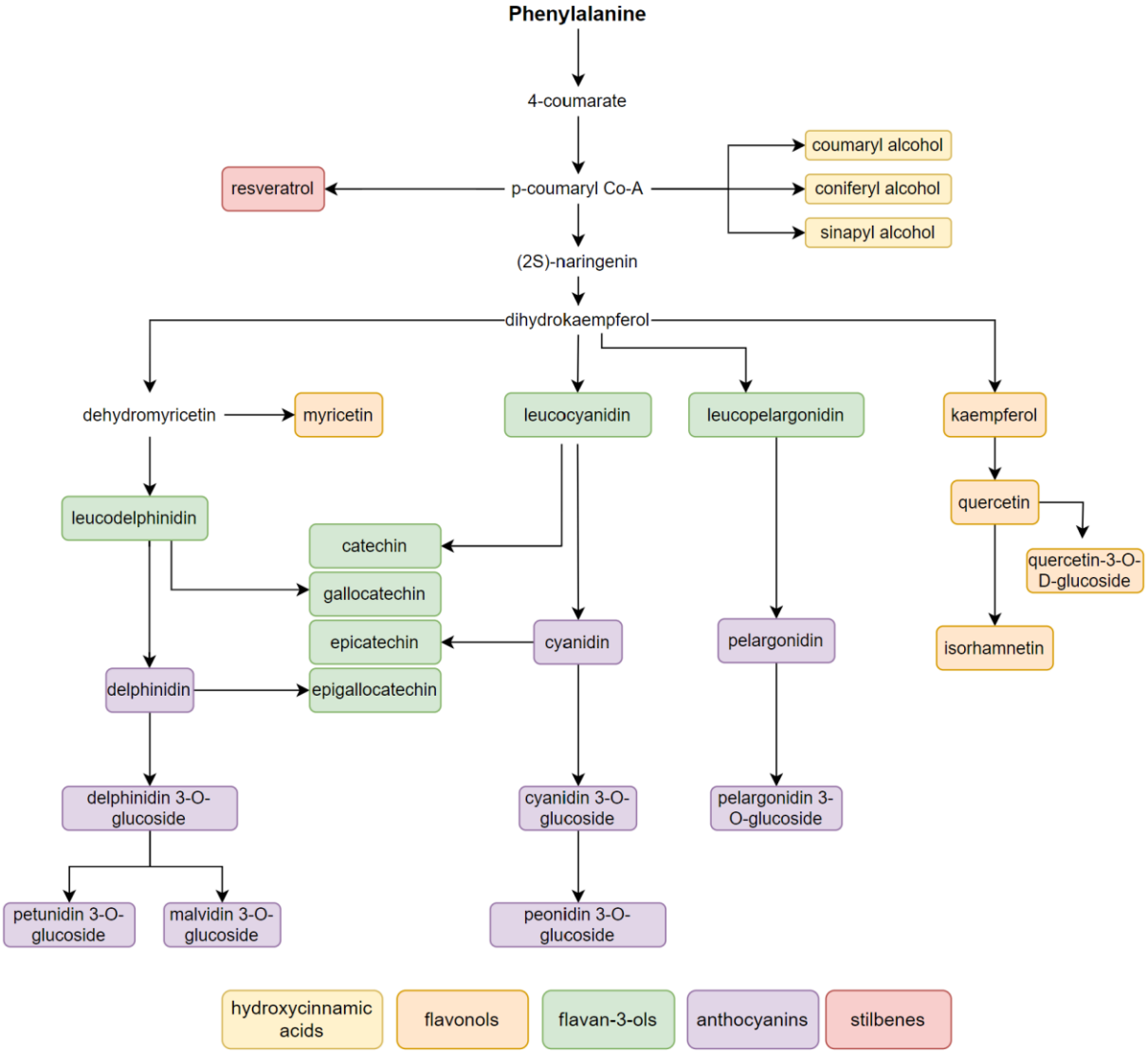


Figure 3. Simplified schema of the production of the main secondary metabolites in the *V. vinifera* model. This illustrates the phenylpropanoid pathway which starts with the amino acid phenylalanine that is converted to p-Coumaryl Coenzyme A. This metabolite can be used to produce hydroxycinnamic acids, and stilbenes, such as resveratrol, or to start the flavonoid biosynthesis pathway to produce different types of flavonoids, such as flavonols, flavan-3-ols, and anthocyanins. Compounds are colored based on the compound class they belong to.

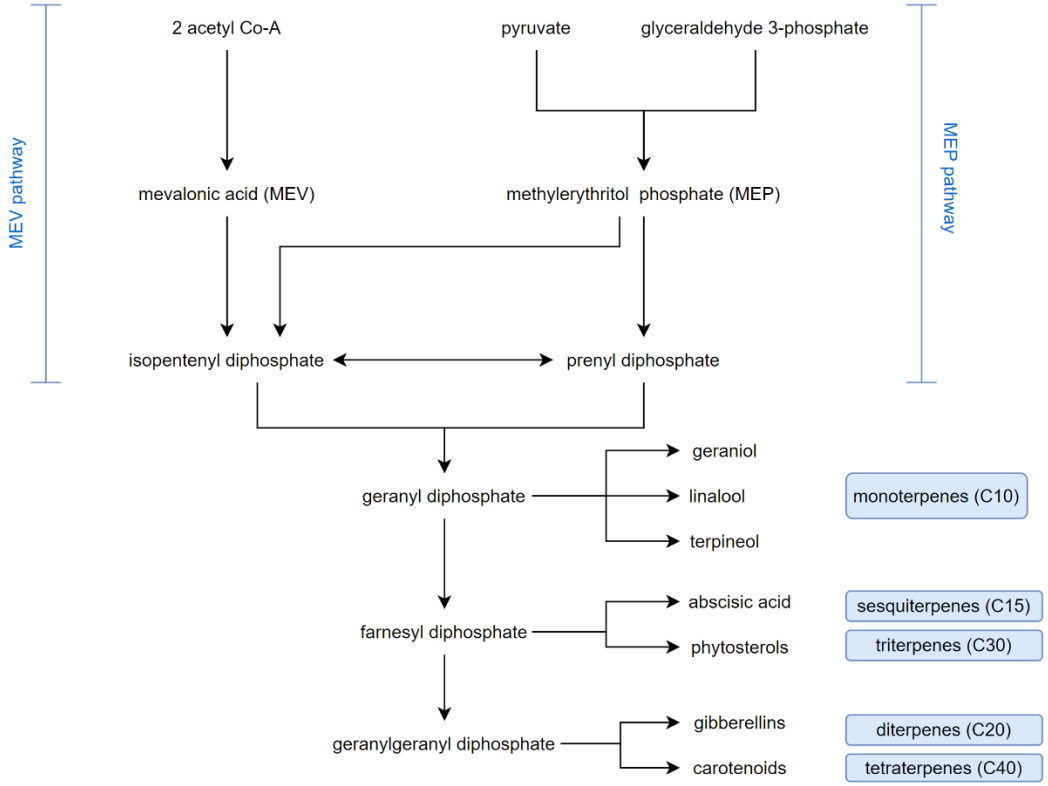


Figure 4. Simplified schema of the terpenoid biosynthesis pathway in the *V. vinifera* model. Isopentenyl diphosphate (IPP) and prenyl diphosphate are the precursors for terpenoids and can be produced from the mevalonic acid (MEV) or methylerythritol phosphate (MEP) pathways. These originate all terpenoids including monoterpenes, sesquiterpenes, triterpenes, diterpenes, and tetraterpenes.

Grapes are known to have a high content of phenolic compounds and different grape varieties usually have different phenolic compositions, which leads to different wine flavors and aromas (Singh et al., 2016). The reconstructed *V. vinifera* model contains complete pathways for the biosynthesis of several terpenoids and phenylpropanoids, which include flavonoids, such as quercetin, myricetin, kaempferol (and derivatives), and anthocyanins, like malvidin and peonidin. Anthocyanins usually accumulate during

grape maturation and are responsible for the grape color in red grapevine varieties, being absent in white varieties (Massonnet et al., 2017a).

Another important group of phenylpropanoids in grapes are stilbenes, such as resveratrol, which protects grapes from intense UV light. As resveratrol is an antioxidant agent, it has high economic importance, being used in the pharmaceutical and cosmetic industry (Saad et al., 2020). The complete pathway of resveratrol biosynthesis is described and included in the model, but there are many gaps in the biosynthetic pathways of resveratrol derivatives, such as viniferins, which are not included in the model but are important for wine flavor and aroma. However, this model also contains complete pathways for the biosynthesis of other aroma compounds, such as linalool, 1,3,5-trimethoxybenzene (TMB), and 3,5-dimethoxytoluene, the latter two being only described for *Rosa chinensis*.

In addition, complete secondary pathways for the biosynthesis of plant hormones, such as jasmonates, cytokinins, gibberellins, ethylene, and auxins, are available in the model. For instance, jasmonates are known to regulate seed germination and flower and fruit development, as well as to defend plants against some pathogens (Wasternack and Song, 2017). Cytokinins usually control cell growth and differentiation (Kieber and Schaller, 2014). Although the reconstructed GSMM does not represent the action of these hormones, it can show the metabolic potential of the network to produce them.

Thus, the reconstructed model of *V. vinifera* represents an important source of secondary metabolic data. Further curation is still necessary to fill the existing gaps and increase the number of secondary metabolites in the model, as new knowledge on these pathways becomes available.

Tissue-specific models

Tissue-specific models were reconstructed to represent the metabolic differences between tissues. This was accomplished by integrating RNA-Seq data with the iMS7199 model.

RNA-Seq Data

The RNA-Seq data of *V. vinifera* Cabernet Sauvignon was retrieved from the GREAT database (Velt and Rustenholz, 2023) for leaf, stem, and berry. In total, the RNA-Seq dataset contained the expression of the 6018 genes (matching the 7199 proteins in the model) across 162 samples. The time-point metadata for berry samples was discretized into two developmental stages, green and mature. The sample distribution and the T-distributed Stochastic Neighbor Embedding (t-SNE) for the RNA-Seq data are shown in Figure 5.

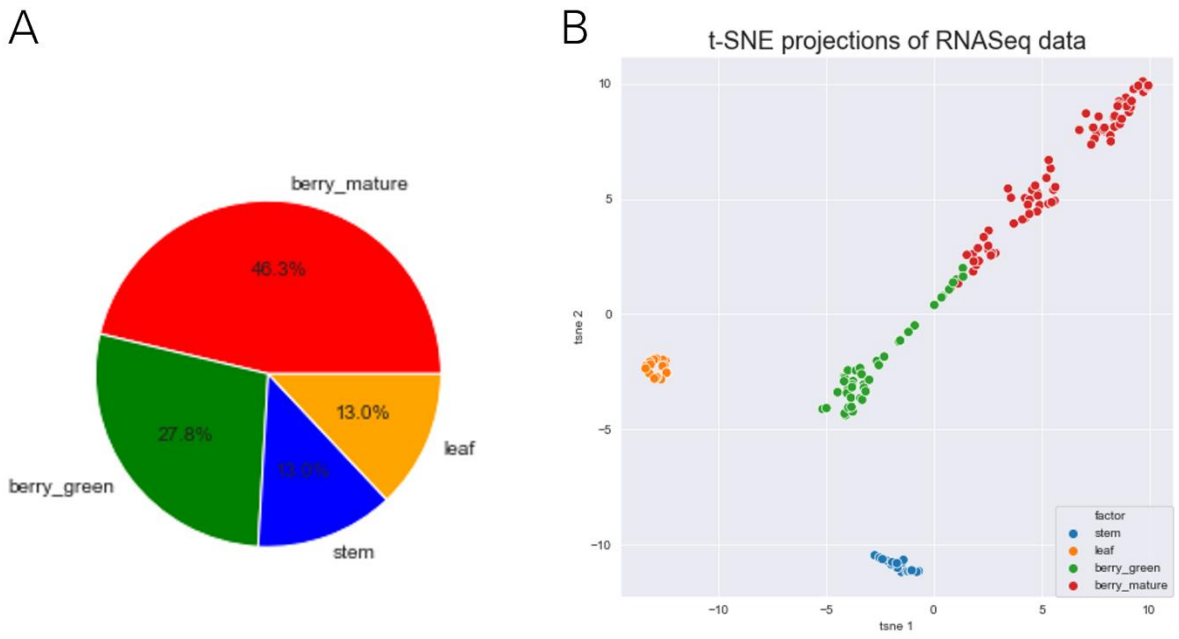


Figure 5.RNA-Seq data for all tissues: leaf, stem, and berry in a green and mature state. A. Distribution of samples across the different tissues. B. t-SNE visualization of the RNA-Seq data for all tissues: leaf (orange dots), stem (blue dots), and berry in green (green dots) and mature (red dots) states. Data was retrieved from the GREAT database, including the expression of the *V. vinifera* genes in the model across 162 samples.

Mature berry is the most represented tissue in the dataset with 46% of the samples (75 samples), 28% of the samples are from green berries (45 samples), while stem and leaf represent 13% of the samples each (21 samples).

Analyzing the t-SNE plot, the data grouped well by tissue: leaf samples are the most well-grouped, followed by stem samples, while berry samples are more scattered.

Green berry samples partially overlap mature berry samples, and some separation exists between some mature berry samples. This was expected as these samples were retrieved every week from fruit set to maturity. Thus, samples a week before and a week after veraison, which is when the maturation phase starts, may be similar in metabolism.

Biomass composition

The biomass composition of the different tissues is represented in Figure 6. Details on biomass compositions are described in the Materials and Methods section and Supplementary File 3. The biomass of leaf and green berry was considered to be the same, and it was used as a reference to define the biomass composition of the other tissues. According to other plant models (*Q. suber* (Cunha et al., 2023b) and *A. thaliana* (de Oliveira Dal'Molin et al., 2015)), the stem is expected to have a higher cell wall and carbohydrate content and lower protein and lipid levels. According to the literature, the mature berry is expected to have higher sugar and amino acid content (Cheng et al., 2016). Therefore, in the model, leaf and green berries present high levels of carbohydrates and proteins, the stem is mainly composed of carbohydrates and cell wall precursors, and mature berries present high amounts of sugars and proteins but fewer organic acids.

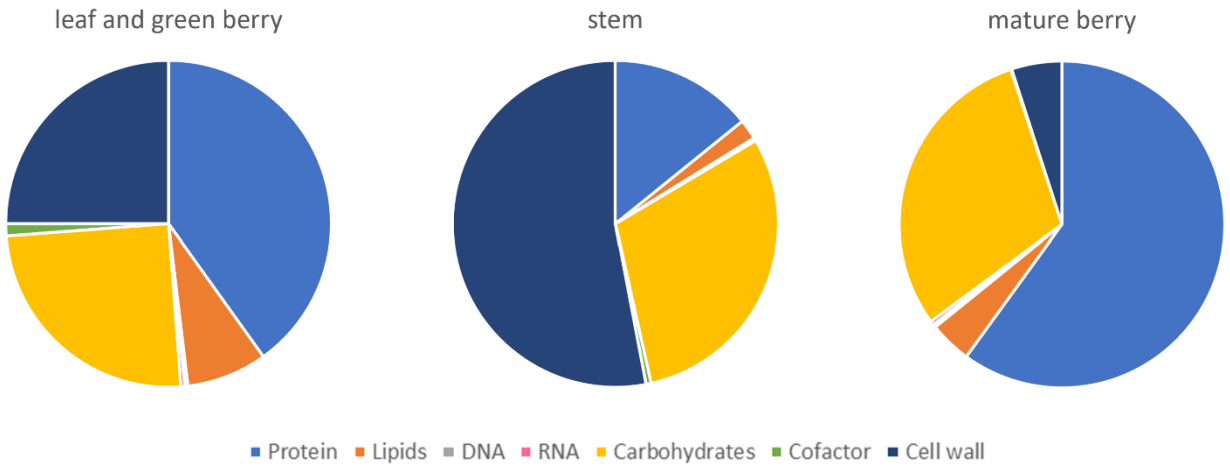


Figure 6. Biomass composition of the leaf and green berry, stem, and mature berry. These values were adapted from available plant models and literature.

Models

The FASTCORE algorithm (Vlassis et al., 2014) was used to create tissue-specific models (see Materials and Methods). The statistics of the reconstructed generic and tissue-specific models are shown in Table 2. All models have the same 244 exchange reactions.

The number of reactions is similar across all tissue-specific models. Even so, the mature berry model is the smallest one, while the leaf model is the largest, having a higher number of reactions in the chloroplast, as well as more unique reactions. At the pathway level, no significant differences were found between models (Supplementary File 4).

Table 2. Statistics of the generic and tissue-specific GSMMs of *V. vinifera*.

	Generic model	Leaf	Stem	Green berry	Mature berry
Genes	7199	6701	6602	6657	6312
Metabolites	5141	4456	4310	4399	4181
Reactions	5399	4510	4384	4495	4272
Transport	1624	1295	1315	1324	1305
Unique reactions	-	124	97	26	19
Metabolic reactions	3531	2971	2825	2927	2723
Cytosol	1434	1154	1092	1113	1059
Chloroplast	793	745	701	725	684
Mitochondria	335	313	318	320	314
Endoplasmic reticulum	568	410	370	417	353
Peroxisome	165	158	152	153	151
Vacuole	52	49	47	48	32
Golgi complex	54	41	41	50	50
Extracellular	130	101	104	101	80

The tissue-specific models were simulated using parsimonious Flux Balance Analysis (pFBA) (Lewis et al., 2010), following the first strategy defined in the Materials and Methods section of keeping biomass rate at 0.11 h^{-1} and minimizing the uptake of photons and sucrose.

A summary of the phenotype predictions is presented in Table 3 and full results are available in Supplementary file 5. The leaf tissue was simulated for all processes: photosynthesis, photorespiration, and respiration, while the other tissues were

simulated for respiration only. Photosynthesis can also occur in green berries, but not at significant levels. Hence, in this analysis, the leaf was considered to be the only photosynthetic tissue.

Table 3. Summary of the net conversions obtained from the phenotype predictions of the tissue-specific models of leaf, stem, green berry, and mature berry for photosynthesis, photorespiration (leaf only), and respiration, minimizing the uptake of photons or sucrose and fixing biomass rate at 0.11h^{-1}. This table shows the metabolites that are consumed and produced by the models. The fluxes of the metabolites are in mmol.gDW^{-1}.h^{-1} while biomass fluxes are in h^{-1}.

	photosynthesis	**photorespiration**	**respiration**			
metabolite	**leaf**			**stem**	**berry green**	**berry mature**
Uptake						
SUCROSE	-	-	0.61	0.49	0.61	0.72
Light	32.09	43.76	-	-	-	-
CARBON-DIOXIDE	4.41	4.41	-	-	-	-
NITRATE	0.37	0.37	0.37	-	0.37	-
OXYGEN-MOLECULE	-	-	1.74	1.71	1.74	1.84
PROTON	6.64	6.64	3.73	4.01	3.73	3.71
SULFATE	0.02	0.02	0.02	0.01	0.02	0.02
WATER	3.21	3.21	-	0.11	-	-
Production						
OXYGEN-MOLECULE	5.54	5.54	-	-	-	-
NITRATE	-	-	-	-	-	0.70
AMMONIUM	-	-	-	0.04	-	-
HCO3	-	-	2.86	1.98	2.86	3.47
PPI	0.05	0.05	-	-	-	-
Pi	-	-	0.09	0.21	0.09	-
WATER	-	-	0.55	-	0.55	1.25
e-Biomass	0.11	0.11	0.11	0.11	0.11	0.11

In photosynthesis and photorespiration, the leaf uptakes light, carbon dioxide, water, nitrate, sulfate, and protons to produce biomass, and releases oxygen and phosphate, as expected. Iron II and magnesium (Mg) are also captured but with very low fluxes (less than 1e-5 mmmol.gDW^{-1}.h^{-1}). Light uptake is significantly higher in

photorespiration than in photosynthesis (43.75 mmmol.gDW^{-1}.h^{-1} *vs* 32.09 mmmol.gDW^{-1}.h^{-1}), which was also expected as the latter process is known to be more efficient for energy production. In both cases, the pathways of the primary metabolism are the most active: photosynthesis, Calvin cycle, glycolysis, starch and amino acid biosynthesis, and oxidative phosphorylation. Also, the photorespiration pathway is only active under photorespiration conditions.

During photosynthetic conditions, the tricarboxylic acid (TCA) cycle is incomplete: citrate is converted to isocitrate, and this is converted to α-ketoglutarate, which is used for the biosynthesis of glutamate and glutamine instead of being used to produce succinate. Fumarate is produced from arginine biosynthesis, instead of being produced from succinate, and enters the cycle. This result is consistent with the results observed for other plant models under light conditions (Maurice Cheung et al., 2014; Cunha et al., 2023b) and with isotope labeling experiments, which stated that a cyclic TCA only happens when the demand for ATP is high. The photosynthetic ATP production reduces that demand (Sweetlove et al., 2010; Williams et al., 2008).

In respiration, the leaf uptakes sucrose, nitrate, sulfate, oxygen, and protons, and releases hydrogencarbonate, water, and phosphate. The main active pathways include glycolysis, the TCA cycle, starch and amino acid biosynthesis, and oxidative phosphorylation. The respiration results were similar across tissues. The stem uptakes water and releases ammonium, and the mature berry has a slightly higher demand for sucrose to produce the same biomass flux.

In summary, the integration of omics data into the generic GSMM created tissue-specific models that try to reflect the differences in gene expression between tissues. However, the number of reactions and the phenotype predictions are not very different between models; thus, a complementary analysis based on differential flux predictions was performed to understand the metabolic differences between the tissues.

Differential flux analysis

The ACHR sampler (Bordel et al., 2010) was used to generate 10000 sample fluxes for all reactions from the different tissue-specific models. Then, these data were used to identify the reactions with differential fluxes between models (see Materials and Methods). In total, 764 reactions were found to have altered fluxes between at least two models. The sampled flux data is shown in the t-SNE of Figure 7.

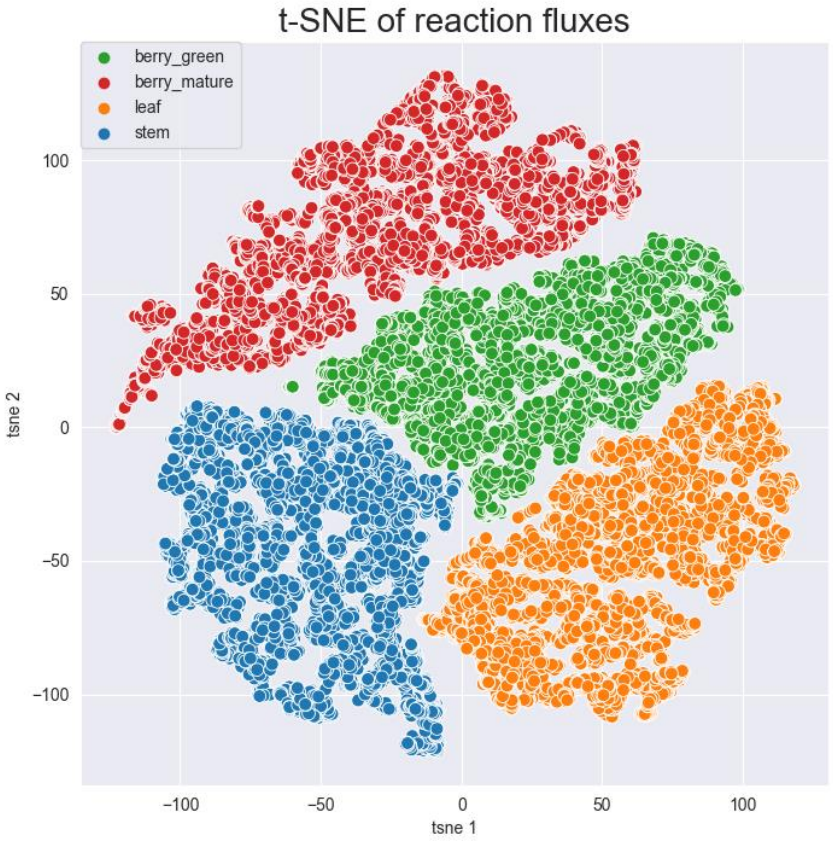


Figure 7. t-SNE visualization of the sampled reaction fluxes of the tissue-specific models. Data was generated by the ACHR sampler, filtered by the reactions with differential flux between models, and scaled. Green dots represent the reaction fluxes from the green berry model, red dots the mature berry, orange dots the leaf, and blue dots the stem.

As observed previously, despite the higher number of samples, t-SNE was able to separate well the fluxes from the different tissues as no overlap is evident between samples of different tissues. In addition, there is no group where flux samples group better nor are more separated from the other groups. These results are not fully in line with those observed for the expression data: leaf flux samples do not appear to group better than the samples from the other tissues, and there is no evident overlap between

green and mature berry samples. This may be because only reactions with differential flux between at least two tissue models were considered in the analysis. Overall, flux data seems to separate the tissues better than gene expression data.

Hypergeometric enrichment tests were used to identify the pathways that presented significantly differential flux between each pair of models. These results are available in the Supplementary File 6. Analyzing the results, it was clear that smaller pathways were not selected even when only one reaction was not identified as having differential flux. Therefore, this method seems to be more suitable for analyzing pathways with a large number of reactions. For this reason, the complete list of reactions with differential flux between the models was also analyzed.

Comparing the green and mature berry models, reactions from glycolysis, TCA cycle, and related to nucleotide biosynthesis were identified as having differential flux. In addition, anthocyanin biosynthesis exhibited more flux in the mature berry, as well as some reactions involved in the biosynthesis of quercetin and derivatives. This was expected as the mature berry has anthocyanins and a higher content of sugars in its biomass composition while demanding a lower content of nucleotides.

Comparisons between the other models are available in Supplementary Material. In summary, it was expected that the primary metabolic pathways would be identified as having differential flux between tissues, as tissue models have different demands for biomass precursors, and produce energy by different processes: the leaf performs photosynthesis, while the others perform aerobic respiration. Besides these, no relevant pathways were found to characterize the specific metabolism of each tissue.

Diel multi-tissue models

Diel multi-tissue models were created to analyze the metabolic interactions between the leaf, stem, and berry of *V. vinifera* in the light (day) and dark (night) phases of a diel cycle. Two models were created, one using the green berry tissue and the other using the mature berry. The resulting diel multi-tissue models include 32391 and 31999 reactions, and 29064 and 28710 metabolites for green and mature berries, respectively.

The structure of the multi-tissue diel models is schematized in Figure 8, showing the different tissues, the diel phases, and the connections between them.

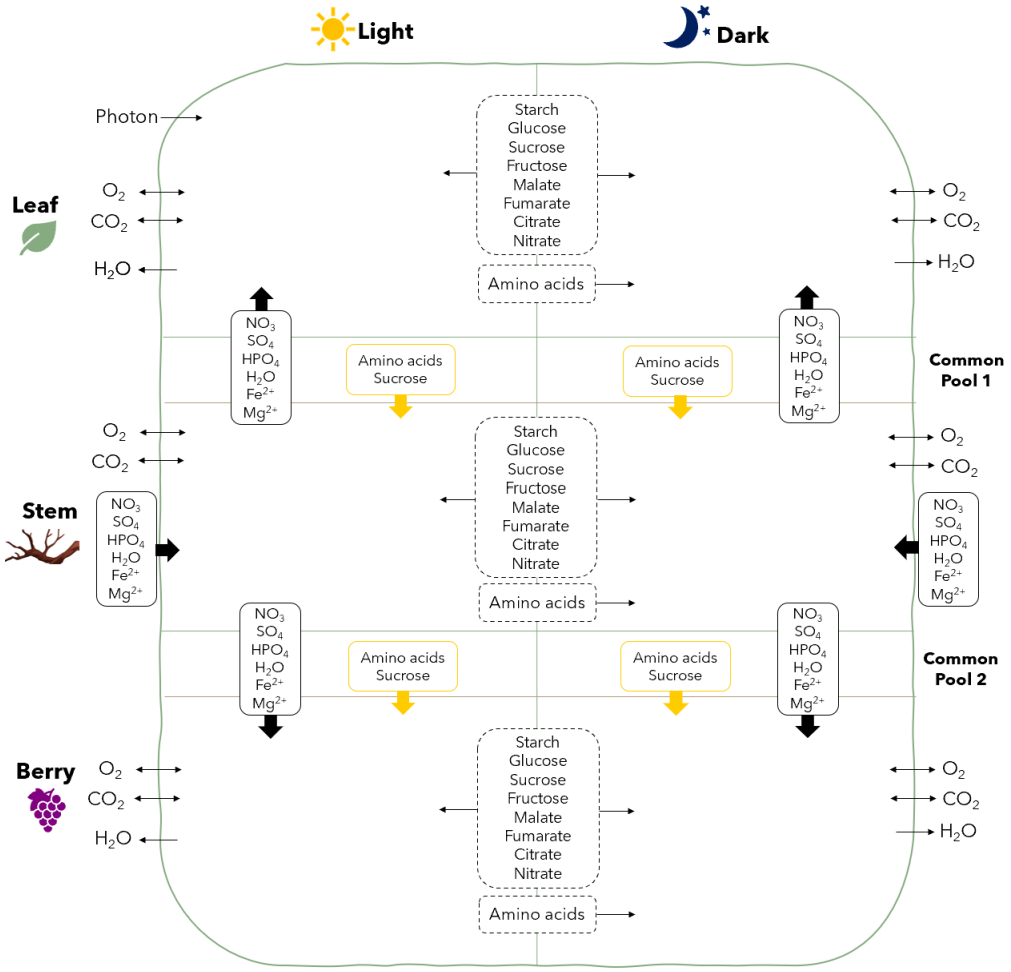


Figure 8. Schematic representation of the reconstructed diel multi-tissue models of *V. vinifera*, including the leaf, stem, and berry tissues and the common pools 1 and 2 in both light and dark phases. Photon uptake was allowed through the leaf in the light phase while mineral nutrients (nitrate, sulfate, phosphate, iron, magnesium) were allowed through the stem in both phases. Exchanges of carbon dioxide, oxygen, and water were allowed in all tissues and phases. Starch, glucose, sucrose, fructose, malate, fumarate, citrate, and nitrate were allowed to accumulate in the light and dark phases (dashed rectangle between phases). Amino acids can be stored in the light and used in the dark. Exchanges of amino acids, sucrose, and minerals were allowed between tissues through common pools.

pFBA was used to simulate the models, as described in the Materials and Methods section for photorespiration conditions. The phenotype predictions are available in the Supplementary File 7. A summary of the results is presented in Table 4 and the fluxes

for the storage metabolites between light and dark phases are shown in Tables Table ***5*** and S1, for green and mature berries, respectively.

Table 4. Summary of the phenotype predictions for the diel multi-tissue models of green and mature berries under photorespiration with biomass maximization as objective function and fixing the photon uptake to 300 mmol.gDW^{-1}.h^{-1}. This table shows the metabolites that are consumed and produced by the models. The fluxes of the metabolites are in mmol.gDW^{-1}.h^{-1} while biomass fluxes are in h^{-1}.

	photorespiration	
metabolite	**green**	**mature**
uptake		
Light__light	300.000	300.000
NITRATE__light	1.147	1.589
NITRATE__dark	0.764	1.059
OXYGEN-MOLECULE_dark	5.999	5.818
PROTON_light	26.330	27.210
PROTON_dark	16.790	16.020
SULFATE_light	0.098	0.112
SULFATE_dark	0.002	0.001
WATER_light	31.070	29.770
CARBON-DIOXIDE_light	35.830	34.920
production		
OXYGEN-MOLECULE_light	39.810	40.84
WATER_dark	1.455	2.788
CARBON-DIOXIDE_dark	0.000	0.000
HCO3_light	2.995	2.956
HCO3_dark	5.602	4.475
Pi_light	0.527	0.384
Pi_dark	0.527	0.384
total biomass	0.149	0.142

Significant differences were found between the light and dark phases, mainly in the leaf, as photosynthesis and photorespiration occur in this tissue. The light phase starts with photosynthesis light reactions and carbon dioxide fixation through the Calvin cycle in the leaf. The resulting carbohydrates are then used to produce all biomass precursors. Starch, sucrose, malate, and some amino acids are stored to be used in the leaf during the dark phase. At night, the active pathways include aerobic respiration, starch degradation, glycolysis, pentose phosphate, and citrate biosynthesis through the TCA

cycle. Sucrose was expected to be produced at night but instead, the model uses fructose 6-phosphate from starch degradation to start glycolysis. The sucrose requirements for biomass are fulfilled by accumulating very small quantities of sucrose between light and dark phases (flux value less than 0.02 mmol.gDW^{-1}.h^{-1}). This is an artifact as the model finds sucrose transport to the dark phase less costly than producing it. However, sucrose production at night can be assured by forcing flux in the respective reactions. In addition, starch is the main carbon compound stored in the light, and it is degraded in the dark phase to produce energy. This was expected as less energy is needed to mobilize plastidic starch reserves than vacuolar sucrose (Maurice Cheung et al., 2014).

Table 5. Fluxes for the metabolites stored between light and dark phases in the diel multi-tissue model with green berry. Positive fluxes indicate that the metabolites are stored in the light phase to be used in the dark while the metabolites with negative fluxes are stored in the dark to be used during the day. The fluxes are in mmol.gDW^{-1}.h^{-1}.

	reaction	flux
leaf	**CIT__vacu_leaf_light_dark_storage**	**-1.354**
	CYS__cyto_leaf_light_dark_storage	0.009
	ILE__cyto_leaf_light_dark_storage	0.025
	MAL__vacu_leaf_light_dark_storage	1.168
	MET__cyto_leaf_light_dark_storage	0.011
	NITRATE__vacu_leaf_light_dark_storage	**-0.020**
	PRO__cyto_leaf_light_dark_storage	0.360
	Starch__chlo_leaf_light_dark_storage	0.374
	SUCROSE__vacu_leaf_light_dark_storage	0.020
	THR__cyto_leaf_light_dark_storage	0.051
stem	**CIT__vacu_stem_light_dark_storage**	**-0.261**
	CYS__cyto_stem_light_dark_storage	0.003
	ILE__cyto_stem_light_dark_storage	0.009
	MAL__vacu_stem_light_dark_storage	0.060
	MET__cyto_stem_light_dark_storage	0.015
	NITRATE__vacu_stem_light_dark_storage	**-0.745**
	PRO__cyto_stem_light_dark_storage	0.276
	Starch__chlo_stem_light_dark_storage	0.009
berry	CYS__cyto_berry_light_dark_storage	0.009
	ILE__cyto_berry_light_dark_storage	0.025
	PRO__cyto_berry_light_dark_storage	0.162
	Starch__chlo_berry_light_dark_storage	0.013

Then, the citrate produced at night is stored in the vacuole to be used during the day, entering the TCA cycle. Nitrate is also transported from the dark to the light to support nitrogen assimilation, which was predicted to occur only during the day. These results were confirmed by experimental evidence and observed in other plant models (Maurice Cheung et al., 2014; Cunha et al., 2023b; Gauthier et al., 2010)

However, the entire TCA cycle was expected to occur in the dark phase. This does not happen in the leaf, as all citrate produced at night is stored in vacuoles to be used during the day. α-ketoglutarate is produced from the degradation of amino acids like glutamate and enters the cycle, which is complete until citrate production. The citrate accumulated, besides feeding the TCA cycle in the light phase, is used for the biosynthesis of Acetyl Co-A during the day, which is then used for lipid production. Therefore, the model finds it more efficient to store more citrate to be used during the day than to complete the TCA cycle in the leaf at night.

Ammonium is provided by the stem and transported to the leaf, where it is used for amino acid biosynthesis. In addition, phosphate, sulfate, pyruvate, formate, and glutamate are imported from the stem through common pool 1 in the light and dark phases to feed amino acid and citrate biosynthesis.

On the other hand, sugars and amino acids produced in the leaf are transported to the stem. The active pathways in the stem during the day have much lower fluxes than in the leaf. These include aerobic respiration, sucrose degradation, glycolysis, starch biosynthesis, pentose phosphate pathway, amino acid, and nucleotide biosynthesis, and degradation of beta-alanine and uracil.

The leaf and stem metabolisms in the dark phase are similar, and the same metabolites are stored between the light and dark phases. Also, in the stem, the TCA cycle is complete during the night, as expected, but a high percentage of the produced citrate is still stored (around 59%). The berry metabolism is very similar to the stem metabolism, but the reaction fluxes are even lower, except for the reactions related to folate biosynthesis. Formate and pyruvate are produced here and transported to the stem

through common pool 2 to be further transported from the stem to the leaf to be used for amino acid biosynthesis. Only starch and amino acids are exchanged in the berry between light and dark phases.

No significant differences were found in the phenotype predictions between green and mature berries. The total biomass rate is slightly higher in the green berry model (0.149 h^{-1}) than in the mature one (0.142 h^{-1}), and generally, the photosynthetic pathways and those related to cellular respiration have lower flux in the mature berry. The pathways related to secondary metabolite biosynthesis, mainly anthocyanins, have flux in the mature and not in the green berry, as expected, but these fluxes are very low; thus, no major differences were observed in the primary metabolism.

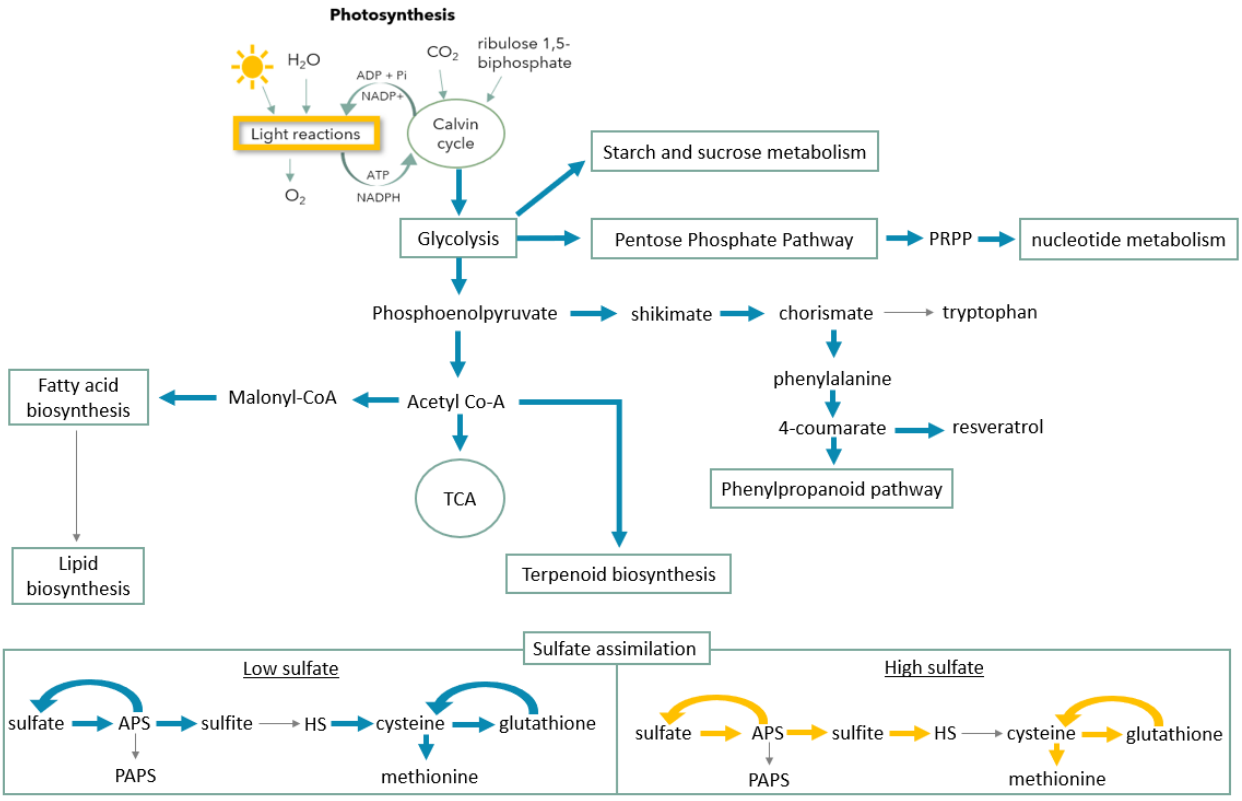
Sulfate Assimilation

Sulfur is an important nutrient taken up by plants from the soil in the form of sulfate, and it is the key element of the amino acids cysteine and methionine. Thus, a major part of sulfate is used for protein biosynthesis. Sulfur is also a component of glutathione, which is an important antioxidant agent, and S-adenosyl methionine and coenzyme A, which are cofactors for several enzymes. Elemental sulfur (S^0) is the oldest pesticide applied to grapevines and it is still widely used nowadays, being particularly effective against powdery mildew disease, one of the most common diseases affecting grapevines that is caused by the fungus *Erysiphe necator*. In addition, sulfur dioxide (SO_2) is often used as a conservative of table grapes or in winemaking to prevent oxidation and microbial contamination.

Plant exposure to high sulfur levels can lead to the accumulation of sulfur-derived compounds or affect the metabolism of phenolic compounds, which can change the flavor, aroma, and texture of grapes and wine (Cheng et al., 2016; Considine and Foyer, 2015). It was observed that residual sulfur on berries can lead to the formation of undesirable flavors, such as hydrogen sulfide (H_2S), during wine fermentation (Considine and Foyer, 2015).

514 *V. vinifera* diel multi-tissue models were simulated to assess the effect of different
515 sulfate concentrations on grapevine metabolism. Flux Variability Analysis (FVA) was
516 used to get the possible range of reaction fluxes while keeping at least 80% of the
517 maximum total biomass and fixing a photon uptake of 300 mmmol.gDW^{-1}.h^{-1}. Two
518 different flux values for sulfate uptake in the light phase were tested, 0.01 and 10
519 mmmol.gDW^{-1}.h^{-1}. The choice of these values was arbitrary, but the goal was to have
520 one value above and one below the unrestricted sulfate uptake flux (Table 4). Similar
521 results were obtained for the multi-tissue models with green and mature berries. Thus,
522 only the results for the green multi-tissue are described. The full results are available in
523 Supplementary File 8 and detailed in Supplementary Material.

524 With high sulfate (10 mmmol.gDW^{-1}.h^{-1}), the maximum flux for biomass production
525 decreased from 0.149 to 0.138 h^{-1}. Similarly, the production of all biomass components
526 also decreased as well as the flux for primary and secondary metabolism (Figure 9).



527

528 **Figure 9.** Simplified schema of the main metabolic pathways in the model affected by varying sulfate
529 levels. Pathways with increased maximum flux under low sulfate conditions are highlighted with a thick

blue arrow while pathways with increased flux under high sulfate conditions are highlighted with a thick yellow arrow. Pathways with decreased flux in both conditions are represented by a thin gray arrow.

As expected, the maximum fluxes of the reactions involved in sulfate assimilation and oxidation, and glutathione biosynthesis have increased. Surprisingly, in the model, the biosynthesis of cysteine and methionine decreased with high sulfate levels. During sulfate reduction, the reaction that produces H_2S has a higher maximum flux but the reaction that uses it to produce cysteine has a lower flux, which leads to a big increase in the flux of the H_2S exchange reaction (Figure 9). This could mean that when plants are exposed to high sulfur levels, they try to adapt to these conditions by adjusting their metabolism, leading to the accumulation of H_2S or other sulfur compounds that can alter the flavor and aroma of the grapes.

When plants are under a sulfate deficiency, the production of biomass and all its components also decreases. For a sulfate uptake of 0.01 $mmmol.gDW^{-1}.h^{-1}$, the maximum production of biomass greatly decreased from 0.149 to 0.018 h^{-1}. Hence, the plant has an excess of carbon skeletons, which are not being used for protein biosynthesis and are available for the biosynthesis of secondary metabolites, increasing the available flux for these pathways. Therefore, there was an increase in the flux of primary pathways, such as sucrose and starch biosynthesis and degradation, gluconeogenesis, and glycolysis, as well as in the pathways responsible for producing secondary metabolites, plant hormones, and amino acids that are precursors of secondary compounds, like phenylalanine (Figure 9). For instance, the maximum flux for resveratrol synthase reaction during the day increased from 0.41 to 1.75 $mmmol.gDW^{-1}.h^{-1}$. In addition, there was a great increase in the maximum flux for the storage of all amino acids, except for cysteine and methionine. Thus, sulfur levels in the soil can greatly influence grapevine metabolism and affect the flavor and aroma of grapes by sulfide or sulfur-compound accumulation or changes in the phenolic content in grapes.

The same approach was applied to assess the effect of different nitrate concentrations in the *V. vinifera* model and similar patterns were observed. The full results are available in Supplementary File 9 and described in Supplementary Material.

Machine Learning and Fluxomics

The potential of using ML to analyze fluxomics data generated by metabolic models was explored. As a significant number of samples is required to train good predictive ML models, the idea of applying ML to fluxomics data is to study the relationships between variables and their impact on the output class. With this in mind, 73 context-specific GSMMs were created by integrating RNA-Seq data from grapes in different developmental stages. Simulated fluxomics data was then obtained from each GSMM as described in Materials and Methods. In this case, the output class to be predicted by the models is the grape developmental phase, green or mature.

First, data was preprocessed and explored using unsupervised methods, starting with t-SNE for data visualization. Before applying the t-SNE, the reactions with the same value in all samples were removed, which greatly reduced the dataset from 8632 to 2322 features.

The results of the t-SNE are plotted in Figure 10. A clear separation between the two states is shown as all green samples are located at the top of the plot while most mature samples are at the bottom. However, six mature samples were grouped with the green ones. These comprise Cabernet Sauvignon samples from time points 5 and 6 and Pinot Noir samples from time point 5. As veraison is expected to occur between time points 3 and 4, these results indicate that not many differences exist at the fluxomics level between green and early mature samples. Even so, fluxomics seems to distinguish well the remaining green and mature samples.

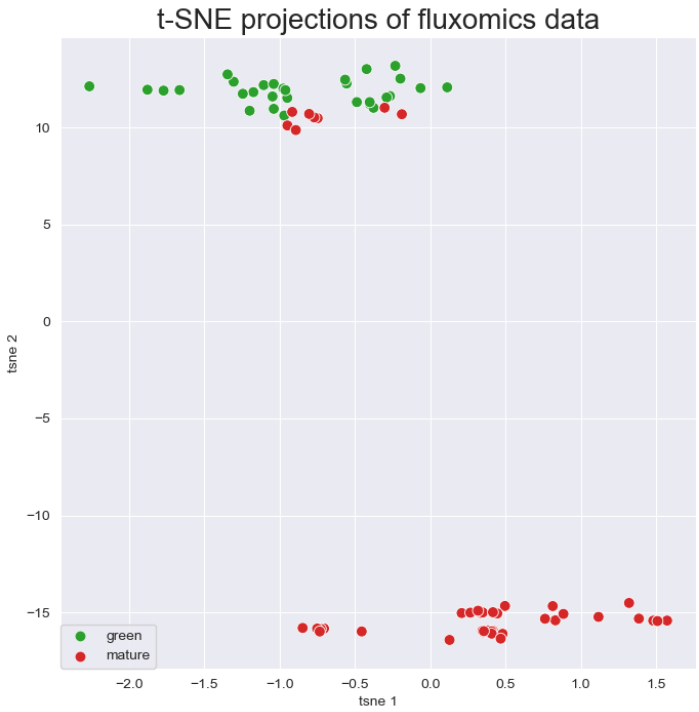


Figure 10. t-SNE visualization of the fluxomics data obtained from the 73 context-specific GSMMs. Samples are colored by the grape developmental stage.

For the supervised analysis, five different ML model architectures were applied including logistic regression (LR), K-nearest neighbors (KNN), decision trees (DT), support vector machine (SVM), and random forests (RF). These ML models were evaluated using repeated stratified cross-validation with 10 folds and 10 repeats. The average evaluation results are shown in Table 6.

Table 6. Evaluation results of the ML models for the metrics balanced accuracy, precision, recall, and F1 score.

	LR	KNN	DT	SVM	RF
BALANCED ACCURACY	0.96	0.95	0.96	0.93	0.97
PRECISION	0.97	0.96	0.98	0.96	0.98
RECALL	0.96	0.98	0.95	0.94	0.97
F1 SCORE	0.96	0.97	0.96	0.94	0.97

According to these results, the models are performing well in predicting the grape developmental stage with this fluxomics dataset. The model's performances across the different folds are robust, indicating that the models can learn meaningful patterns in the

data and handle data variations well. Overall, RF obtained higher values for all metrics, while SVM presented the worst performance. Despite the results being good and the models being able to generalize correctly on the test set of each fold, the dataset is very small (73 samples), which is a common problem when working with omics data. Larger datasets are needed to create better predictive models and draw more conclusions from the data. Nevertheless, these models represent a good start for understanding which reactions contribute most to the model's prediction. Hence, SHAP values (Lundberg et al.) were calculated for the two best models, RF and KNN, and the most contributing reactions are shown in Figure 11 for RF. The KNN results are described in Supplementary Material.

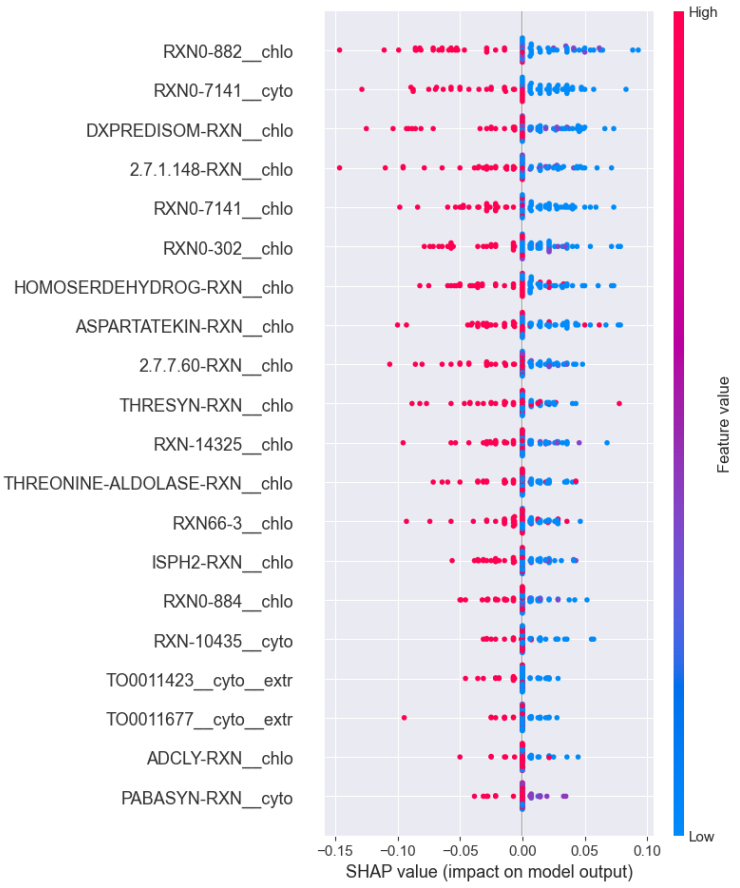


Figure 11. Beeswarm plot of SHAP values for the reactions that contribute most to RF's predictions. Features are ordered from higher to lower effects on the predictions. The dots represent a single observation, and the color indicates if the observation has a higher (pink) or a lower (blue) feature value compared to the other observations.

Overall, when the reactions presented high flux capacity (FCa) values, they had negative SHAP values, leading the model to predict the green state, while with lower FCa they exhibited higher positive SHAP values, leading the model to predict the mature state. Hence, all the reactions identified here presented an average FCa value higher in the green than in the mature grapes.

For the RF model, RXN0-882 in the chloroplast is the reaction that most contributes to the predictions. Similarly, high fluxes of this reaction have negative SHAP values (around -0.15), classifying the samples as green, while lower fluxes have positive SHAP values (close to 0.10), classifying the samples as mature. There are some exceptions to this trend, such as the chloroplastic THREONINE-ALDOLASE-RXN and THRESYN-RXN reactions that show positive SHAP values when presenting high flux, indicating that samples classified as mature by the models can also have high fluxes in these reactions.

Of the 20 reactions identified for each model, 10 have a high impact on both models, indicating that the results are reliable and robust and that these features are important for predicting the output. Most of these reactions are involved in the methylerythritol phosphate (MEP) pathway, which is responsible for the biosynthesis of the terpenoid precursors (Figure 4), threonine degradation into glycine, and the transport of glycerides. The remaining reactions identified only with the RF model are involved in the biosynthesis of nucleotides, 4-aminobenzoate, and threonine.

The accumulation of terpenoids in grapes typically starts before veraison, which can explain why the reactions associated with the biosynthesis of terpenoid precursors had higher FCa in the green state. However, terpenoid biosynthesis intensifies after veraison, which is not observed in the fluxes of these reactions. Fasoli et al., 2018 have also identified terpene metabolism as a negative biomarker for the onset of ripening. In addition, the abscisic acid (ABA) signaling is increased at veraison, and ABA is derived from carotenoids, whose biosynthesis starts with the MEP pathway. Thus, there is strong evidence that genes or reactions from the MEP pathway could be used as biomarkers for the onset of ripening.

In the green phase, as grapes are rapidly growing, the metabolism of amino acids, nucleotides, and lipids is expected to be more active than in the mature phase. 4-aminobenzoate is a precursor for the biosynthesis of various metabolites, such as tetrahydrofolates, which are involved in several processes like photorespiration, amino acid metabolism, and protein biosynthesis. These pathways are also expected to be more active in the green phase. This fact may explain why the reactions related to these pathways are important for the model's predictions. However, it is not clear why threonine metabolism is more important for the model than the metabolism of the other amino acids. Nevertheless, the models presented good predictions, associating high fluxes of these reactions to predict the green state and low fluxes to predict the mature state.

Materials and methods

Metabolic data source

The metabolic information of PlantCyc 14.0 (Zhang et al., 2010) and MetaCyc 26.1 (Caspi et al., 2016) databases was saved and organized in a repository named iplants, using Neo4j (Huang and Dong, 2013) and MongoDB (Jose and Abraham, 2017) database management systems. These NoSQL databases do not store data in relational tables, having instead a more flexible schema. Neo4j uses a graph structure, while MongoDB uses a document structure to represent data. The implementation and management of the databases were performed with Python 3.8, using the Neomodel package, an object graph mapper for Neo4j, and Mongoengine, an object document mapper for MongoDB. The Neo4j database saved the connections between metabolites, reactions, enzymes, genes, pathways, organisms, and models, while MongoDB stored metadata for all Neo4j entities. UniProt (The UniProt Consortium, 2017) data for enzymes were also collected when available, and added to MongoDB, including protein function, localization, sequence, and annotation status.

In addition, nine plant metabolic models of different species were integrated into the repository, comprising *A. thaliana, O. sativa, M. truncatula, S. viridis, G. max*, and *S.*

lycopersicum. This integration involved matching the iplants entry identifier with the identifier in the models and connecting the model object with the associated objects in the database.

Model Reconstruction

The reconstruction of the GSMM of *V. vinifera* was based on the PN40024.v4 genome and the PN40024.v4.1 annotation version, which includes 35922 genes and 41160 proteins (Velt et al., 2023). All the main steps and analyses were performed using Python 3.8 and *COBRApy* version 0.25 (Ebrahim et al., 2013) and are schematized in Figure 12.

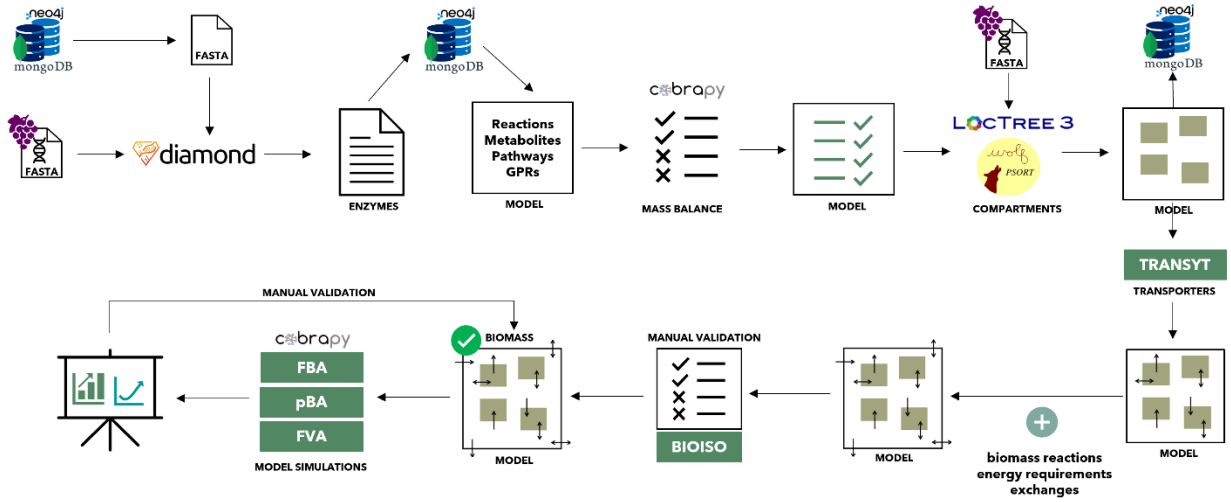


Figure 12. The main steps of the *V. vinifera* GSMM reconstruction. Genome annotation is performed using DIAMOND with the protein sequences from the database and *V. vinifera* proteins from the genome. The identified enzymes are used to get all metabolic information from the database as well as to define the GPR rules in the model. Next, the mass balance of reactions is checked and fixed whenever possible. Unbalanced non-essential reactions were removed from the model. Then, subcellular compartments were predicted using LocTree3 and Wolfpsort tools, and transporters between these compartments were predicted using TranSyt. Finally, biomass, exchange, and energy requirement reactions were defined, and the model was manually validated, using BioISO to verify biomass production. The final model was simulated using different methods, such as FBA and FVA, and the results were analyzed to validate the model.

The reconstruction started with genome annotation which was based on DIAMOND (Buchfink et al., 2014) similarity searches of *V. vinifera* proteins against the protein sequences in iplants, to assign enzymatic functions and associate reactions to *V. vinifera* proteins. Using only the iplants database to perform the annotation facilitated

the identification of the reactions to include in the model. DIAMOND searches were performed against the SwissProt database to determine the completeness of important enzyme annotations in iplants. In the course of these searches, it was found that more than 17,000 *V. vinifera* proteins had a match in SwissProt. However, most proteins were involved in other processes like transcription regulation, protein phosphorylation, and nuclear transport, which are not relevant to be included in the metabolic model.

Based on the iplants' annotation, the reactions linked to the identified enzymes and spontaneous reactions were assembled to create a draft metabolic network. WolfPsort (Horton et al., 2007) and LocTree3 (Goldberg et al., 2014) were used to predict the subcellular location of the proteins. However, due to contradictory outcomes from these tools, manual curation of the results was deemed necessary. For instance, certain enzymes catalyzing the reactions of the sphingolipid biosynthesis, like serine C-palmitoyltransferase and dihydroceramide fatty acyl 2-hydroxylase reactions, were predicted to be on the endoplasmic reticulum by Loctree3 and in the chloroplast or cytosol by WolfPSort. In this case, the annotations collected from UniProt were considered and the location of these enzymes was defined to be the endoplasmic reticulum. The thylakoidal and mitochondrial intermembrane compartments were added manually for the photosynthesis and oxidative phosphorylation reactions, respectively.

Transport reactions were automatically identified using TranSyT (Cunha et al., 2023a) and added to the model. Additional transporters were manually included in the model when required. All reactions were validated for mass and charge balance.

Biomass composition

The definition of tissue-specific biomass compositions is crucial to obtaining good models capable of simulating the specific metabolism of each tissue. Biomass is composed of macromolecules labeled "e-metabolites", which are required for cell growth, including RNA, DNA, proteins, carbohydrates, lipids, co-factors, and cell wall components. Ideally, experimental data should be used to define biomass composition for different tissues. However, as these data are not available for *V. vinifera*, the biomass content was estimated based on previously published plant GSMMs and

727 insights from the literature. Specifically, the biomass formulation for leaf, stem, and
728 green berry was based on the models of *A. thaliana* (Dal'Molin et al., 2010; de Oliveira
729 Dal'Molin et al., 2015), *S. lycopersicum* (Yuan et al., 2016), and *Q. suber* (Cunha et al.,
730 2023b), while the biomass of the mature berry was adjusted according to the literature
731 (Cheng et al., 2016) and the metabolomics data available for the same samples used to
732 obtain the RNA-Seq data (Fasoli et al., 2018). Details of the biomass composition are
733 available in Supplementary File 3.

734 The monomer contents for the production of DNA, RNA, and proteins were calculated
735 from the genome sequence using the biomass tool (Santos and Rocha, 2016), available
736 in *merlin* (Dias et al., 2015). The reactions for the production of the cell wall,
737 carbohydrates, fatty acids, lipids, and co-factors were also adapted from *A. thaliana* and
738 *Q. suber* models. The e-Cofactor metabolite includes several universal cofactors, such
739 as NAD(H), and vitamins, and in the case of leaf, it also includes pigments, such as
740 chlorophylls. The content of carbohydrates was adapted to reflect the grape
741 composition described in the literature. For instance, tartaric acid was added to the
742 model as it was described as the main organic acid found in grapes and it is not present
743 in any other plant model (Cheng et al., 2016). For mature berries, the content of sugars
744 and organic acids was adjusted to reflect the changes in berry composition during
745 maturation. In addition, some secondary metabolites were added to the carbohydrate
746 reaction of the mature berry based on metabolomics data (Fasoli et al., 2018). These
747 metabolites were found in mature grapes and mainly included anthocyanins, such as
748 malvidin 3-glucoside, peonidin 3-O-glucoside, and petunidin-3-O-glucoside.

749 *Manual validation*

750 Manual curation was an essential step during the reconstruction process. Literature and
751 biological databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG)
752 (Kanehisa et al., 2017), National Center for Biotechnology Information (NCBI) (Sayers
753 et al., 2022), BRaunschweig Enzyme Database (BRENDA) (Placzek et al., 2017) and
754 UniProt (The UniProt Consortium, 2017), were consulted to retrieve additional
755 information about specific reactions, enzymes, or pathways.

Validating the model required verifying its ability to produce biomass. *BioISO* (Cruz et al., 2024) was used to accomplish this goal, as it verifies the production of each biomass substrate to see which ones are missing. Then, the gaps in the model were analyzed and filled in when necessary for the production of essential metabolites. Also, it was assured that growth did not occur without photons and carbon sources, and no futile cycles were present. Then, dead-end metabolites and blocked reactions were identified, and each blocked reaction was analyzed and fixed by resorting to other databases. When no information was available, the blocked reactions were kept in the model.

Finally, the model's capability to accurately simulate key metabolic processes like photosynthesis, photorespiration, and respiration was confirmed. This was achieved through the application of diverse methods, including Flux Balance Analysis (FBA), parsimonious FBA (pFBA), and Flux Variability Analysis (FVA). FBA (Varma and Palsson, 1994), which uses linear programming to calculate an optimal flux distribution for a given objective function, has been extensively used to simulate GSMMs. However, usually, multiple optimal solutions exist in the solution space for a given objective. Hence, pFBA has emerged as a novel approach. It refines the traditional FBA by selecting a flux distribution from the FBA optimal space that minimizes the total sum of fluxes (Lewis et al., 2010). Likewise, FVA (Mahadevan and Schilling, 2003) is used to determine the span of flux variability of GSMMs in simulations by calculating the minimum and maximum flux of each reaction for a defined set of constraints.

Tissue-specific Models

Omics Data

Available RNA-Seq data from different tissues were used to create specific GSMMs for stem, leaf, and berry. Due to the absence of a single study providing RNA-Seq data for all three tissues, we resorted to using two distinct studies to gather the necessary information. Leaf and stem data were retrieved from the healthy samples of *V. vinifera*

Cabernet Sauvignon cultivar in the study of Massonnet et al., 2017b (Gene Expression Omnibus (GEO) accession: GSE97900). RNA-Seq data for the C. Sauvignon berries was obtained from the study of Fasoli et al., 2018 (GEO accession: GSE98923), which included berry samples in different developmental stages and associated metabolomics data. The time-point metadata for these samples was grouped into two developmental stages, green and mature, to create a metabolic model for each state. Samples until time point 4 were considered to be in the green state, while samples after time point 4 were considered to be mature. Although sample time 4 was collected 7 days after veraison, it was still considered to be in the green phase to reduce class imbalance.

These datasets were retrieved from GRape Expression ATlas (GREAT) (Velt and Rustenholz, 2023). GREAT is a gene expression atlas for grapevine that integrates all public RNA-Seq experiments, allowing the analysis and visualization of the data. The RNA-Seq data were already normalized in transcripts per million (TPM) and the reads were mapped to the new genome (PN40024.v4).

A dataset with all collected data and respective metadata was built, and a log2 transformation was applied to the gene expression values. Finally, the gene identifiers in the datasets were mapped to the protein identifiers present in the model to allow for omics integration.

Models

Tissue-specific models for stem, leaf, green berry, and mature berry were reconstructed using the FASTCORE algorithm (Vlassis et al., 2014) implemented in the *Troppo* package (Ferreira et al., 2020). This algorithm identifies the reactions that should be removed or kept in the model, based on the expression levels of the genes associated with each reaction through GPR rules, resulting in models with different reaction content.

The reconstructed generic model and the omics dataset were used as input for this algorithm, and the local T2 thresholding strategy (Richelle et al., 2019) was applied to preprocess the omics data before integrating it into the model. In this strategy, two global thresholds, upper and lower, are defined, and genes whose expression is below

the lower threshold are considered to be inactive, while genes whose expression is above the upper threshold are considered to be active. Genes with expression levels between these two thresholds may be active or not and further analysis is performed by comparing the expression values to a local threshold. This is a gene-specific threshold that accounts for the expression levels of each gene over all samples, while the global thresholds have the same value for all genes. This strategy was employed as it was described to obtain better results (Richelle et al., 2019).

In this work, we selected the percentiles 25 and 75 for the global lower and upper thresholds, respectively, and the percentile 50 (median) for the local threshold. The pseudo-reaction representing the drain of macromolecules required to create a new unit of biomass was included in the set of protected reactions of the algorithm so that all tissue-specific models would be able to produce biomass.

Phenotype predictions

Phenotype predictions of the tissue-specific models were performed by pFBA, using two different strategies as applied in other plant models (Dal'Molin et al., 2010; Cunha et al., 2023b) and based on *A. thaliana* experimental measures (Niemann et al., 1995). The first consisted of fixing the biomass growth rate to 0.11h^{-1} and defining the minimization of the photon/sucrose uptake as the objective function. The second strategy defined the biomass growth rate as the objective function and fixed the photon uptake to 100 mmmol.gDW^{-1}.h^{-1} for photosynthesis and photorespiration and the sucrose uptake to 1 mmmol.gDW^{-1}.h^{-1} for respiration. Photorespiration was simulated by constraining the carboxylation (RIBULOSE-BISPHOSPHATE-CARBOXYLASE-RXN) and oxygenation (RXN-961) reactions by Rubisco with a flux ratio of 3:1 (Dal'Molin et al., 2010; Cunha et al., 2023b).

Differential Flux Analysis

Differential flux analysis between the created tissue models was performed using the approach of (Nanda and Ghosh, 2021). In this approach, sample fluxes for the tissue models were generated using Artificial Co-ordinate Hit and Run (ACHR) sampler

(Bordel et al., 2010) from *CobraPy*, with a thinning factor of 100 and a sample size of 10000 for each model. Pairwise Kolmogorov-Smirnov tests were used to compare the flux distribution of the distinct tissue models. The flux change (FC) of each reaction between the two models was also calculated as shown in Equation (1), where $\bar{S}_{model1}$ and $\bar{S}_{model2}$ represent the mean of the flux distributions for a reaction in *model1* and *model2*, respectively. Reactions with an absolute value less than 0.82 (equivalent to a 10-fold change in flux) were considered insignificant.

$$FC = \frac{\bar{S}_{model1} - \bar{S}_{model2}}{|\bar{S}_{model1} + \bar{S}_{model2}|} \quad (1)$$

For reactions that are absent in a model, their flux is assumed to be zero in that model, and bootstrapping is used to estimate the 95% confidence interval of their fluxes. If zero is outside the interval, the reactions are considered to have differential flux in the two models. In addition, the p-values of altered reactions were adjusted by a Benjamini-Hochberg correction, with a significance level of 0.05. The differential pathways between models were obtained using hypergeometric enrichment tests that select the pathways that are over-represented due to the higher number of altered reactions and not by chance.

T-distributed Stochastic Neighbor Embedding (t-SNE) (Van Der Maaten and Hinton, 2008) was used to visually compare the sampled flux data of the different tissue models. This tool is a dimensionality reduction algorithm that allows for nonlinear data separation. Before applying t-SNE, the flux data was filtered, keeping only the reactions with altered flux between models, and scaled to z-scores.

Diel multi-tissue models

Models

Multi-tissue models were created by joining the tissue-specific models and connecting them by two common pools, one between stem and leaf, and the other between stem and berry, based on a previous approach (de Oliveira Dal'Molin et al., 2015). Transport reactions between tissues and common pools were manually added when required.

Given that a model for the root was not developed, we assumed that the uptake of minerals occurs in the stem. Exchanges of water, oxygen, and carbon dioxide were allowed in all tissues, and light absorption was allowed only in the leaf model. Two multi-tissue models were created, one with the berry in the green phase and the other with the mature berry.

Additionally, diel models were created to account for light and dark phases. All reactions and metabolites were duplicated for each phase, and new reactions were added to allow the exchange of some metabolites between the two phases. These are called storage metabolites and include the 20 amino acids, nitrate, citrate, malate, glucose, sucrose, fructose, and starch, which can be produced in one phase and used in the other, as previously described (Maurice Cheung et al., 2014).

Phenotype predictions

Phenotype predictions using multi-tissue diel models were also performed with pFBA using the second strategy mentioned above for photorespiration conditions, but with a photon uptake of 300 mmmol.gDW^{-1}.h^{-1} as flux values were very low with 100 mmmol.gDW^{-1}.h^{-1}. As in other plant diel models (Maurice Cheung et al., 2014; Shaw and Cheung, 2018; Cunha et al., 2023b), the nitrate uptake was constrained to a ratio of 3:2 in the light and dark cycle.

Machine learning and Fluxomics

Data

All samples from the RNA-Seq dataset from Fasoli et al., 2018 were used to create simulated fluxomics data for grapes in the green and mature state, by using the iMS7199 model and the aforementioned phenotype prediction approaches to reach flux distributions. Firstly, the mean expression value of all replicates was calculated to represent each biological sample, resulting in a dataset with 73 samples, 55% from Cabernet Sauvignon and 45% from Pinot Noir, and the log2 expressions of the 6018 genes in the model. As performed before, these samples were discretized into two

903 developmental stages, green and mature, which represent the output class to be later
904 predicted by the ML models. Then, the final RNA-Seq dataset was integrated with the
905 generic *V. vinifera* model to create GSMMs representing each sample in the dataset
906 using the FASTCORE algorithm as described before. In total, 73 contex
907 models were created, one per sample. The resulting sample-specific GSMMs were
908 simulated using FVA and the flux capacity (FCa) of each reaction was calculated by
909 subtracting the maximum and minimum flux obtained for each reaction while keeping
910 80% of the maximum biomass value (Equation (2)). Before running FVA, all reactions
911 were made irreversible to facilitate the interpretation of results. The reactions absent in
912 the models were considered to have a capacity of 0.

(2)

$$FCa(r) = Flux_{max}(r) - Flux_{min}(r)$$

913 *Models*

914 The analysis of fluxomics with ML was performed in Python 3.11 with Scikit-learn 1.2.2.
915 For the unsupervised analysis, the dataset was filtered to remove the reactions with the
916 same FCa across all samples using *VarianceThreshold* and scaled by *StandardScaler*.
917 t-SNE was applied to visualize the distribution of the data. For the supervised analysis,
918 the original dataset was divided into train and test sets by cross-validation with 10 folds
919 and repeated 10 times, using the *RepeatedStratifiedKfold* function. In each iteration,
920 feature selection was performed using *VarianceThreshold* and, as the number of
921 features was still high, the *SelectKBest* function was used to select the 500 most
922 relevant features based on ANOVA F-values. In addition, the resulting dataset was also
923 scaled by *StandardScaler*. Then, an ML model fitted the train data and predicted the
924 output classes for the test set. Five different ML models were tested including logistic
925 regression, K-nearest neighbors, decision trees, support vector machine, and random
926 forests. These were evaluated by different metrics, such as recall (Equation (3)),
927 precision (Equation (4)), balanced accuracy (Equation (5)), and F1 score (Equation (6)),
928 which were averaged across all train-test splits.

$$recall = \frac{TP}{TP + FN} \quad (3)$$

(4)

(5)

$$precision = \frac{TP}{TP + FP}$$

$$balanced\ accuracy = 0.5 * (recall + \frac{TN}{TN + FP})$$

$$F1\ score = \frac{2 * TP}{TP + 0.5\ (FP + FN)}$$

The importance of each feature in the prediction of the output was analyzed by calculating the SHAP values for each classifier. SHAP values (SHAPley Additive exPlanations) are defined based on the contribution of each feature to the prediction of each sample and are used to increase the interpretability of ML models. Larger absolute SHAP values have a larger effect on the prediction (Lundberg et al.). The SHAP values for each fold of the repeated cross-validation were calculated, and the average SHAP values for each sample were calculated to give a more stable representation of the feature contributions.

Conclusion

In this work, we reconstructed the first GSSM of *V. vinifera*. This model is based on the latest *V. vinifera* genome and database knowledge, including primary and secondary metabolic pathways, mainly related to flavonoids and hormone biosynthesis. The model can simulate grapevine metabolism under photosynthesis, photorespiration, and respiration. RNA-Seq data was integrated with this generic model to build tissue-specific models for the leaf, stem, green berry, and mature berry of *V. vinifera* Cabernet Sauvignon cultivar. Multi-tissue models were built by connecting the tissue-specific models, and the diel cycle was introduced in the models by replicating the multi-tissue model for both light and dark phases. Two diel multi-tissue GSMMs were built, one using the green berry tissue and the other using the mature berry tissue. The models were used to simulate the metabolic responses of grapevine to different levels of sulfate and nitrate. The results indicated that with low nitrate or low sulfate, less biomass is produced, and more flux is expected in the respiratory pathways, fatty acid production, and secondary pathways. Conversely, with high levels of nitrate or sulfate, the maximum

flux of secondary reactions has decreased, as well as most primary pathways of sugar metabolism. Hence, controlling the soil levels of nitrate and sulfate in specific stages of development can help control the phenolic and sugar content in the grapes, which will affect their quality.

The reconstructed metabolic models developed here can be a valuable tool for analyzing and predicting the metabolic behavior of grapevine under different environmental conditions and assessing its metabolic potential and fruit quality, which can be important for wine production.

Fluxomics data were generated from GSMMs of green and mature grapes and analyzed using ML techniques. The resulting models obtained very good results in predicting the grape developmental stage, with accuracy, precision, recall, and F1 scores higher than 90%. The reactions that contributed the most to the model's predictions are associated with different pathways, including MEP, threonine, nucleotide metabolism, and ascorbate degradation, and presented higher fluxes in the green state. Although these pathways are not the main differences between green and mature grapes found in the literature, the results suggest that their fluxes are significantly different between the two states.

A deeper understanding of plant metabolic pathways is essential to develop more robust GSMMs. Additionally, the creation of larger omics datasets is crucial for developing more realistic predictive ML models, enabling more advanced analyses such as the identification of biomarkers for disease or environmental stress resistance. This approach not only represents a novel and pioneering effort in integrating omics, GSMMs, and ML in plant metabolism studies but also showcases the significant potential of applying this strategy for more insightful analyses, as additional data becomes available.

Supplemental Data

Supplemental Material. Technical and additional details about the methods and results.

Supplemental File 1. List of the attributes for all objects in MongoDB.

Supplemental File 2. Full comparisons between the metabolic content of iMS7199 and other plant models.

Supplemental File 3. Details on the definition of biomass compositions.

Supplemental File 4. Pathway content by tissue.

Supplemental File 5. pFBA phenotype predictions of each tissue GSMM.

Supplemental File 6. Full differential flux analysis results.

Supplemental File 7. pFBA phenotype predictions of the diel GSMMs.

Supplemental File 8. Effects of varying sulfate levels on the phenotype predictions of the diel GSMMs.

Supplemental File 9. Effects of varying nitrate levels on the phenotype predictions of the diel GSMMs.

Supplemental File 10. Zip folder with all GSSMs reconstructed in this work in the Systems Biology Markup Language (SBML) format: iMS7199, leaf, stem, berry green, berry mature, diel with green berry, and diel with mature berry models.

Competing interests

No competing interest is declared.

Author contributions statement

M.S. was involved in the conceptualization and writing of the original draft. M.R. and O.D. were responsible for the conceptualization, reviewing, and editing of the manuscript.

Acknowledgments

The authors thank the Portuguese Foundation for Science and Technology (FCT) which supported this work under the scope of the strategic funding of UID/BIO/04469/2020 unit and the PhD scholarship (SFRH/BD/144643/2019) to Marta Sampaio. Oscar Dias also acknowledges FCT for the Assistant Research contract obtained under CEEC Individual 2018 (DOI 10.54499/CEECIND/03425/2018/CP1581/CT0020).

References

Antonakoudis, A., Barbosa, R., Kotidis, P., and Kontoravdi, C. (2020). The era of big data: Genome-scale modelling meets machine learning. Comput. Struct. Biotechnol. J. 18: 3287–3300.

Bogart, E. and Myers, C.R. (2016). Multiscale metabolic modeling of C4 plants: Connecting nonlinear genome-scale models to leaf-scale metabolism in developing maize leaves. PLoS One 11: 1–27.

Bordel, S., Agren, R., and Nielsen, J. (2010). Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes. PLOS Comput. Biol. 6: e1000859.

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 2014 121 12: 59–60.

Caspi, R. et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 44: D471–D480.

Chatterjee, A., Huma, B., Shaw, R., and Kundu, S. (2017). Reconstruction of Oryza sativa indica genome scale metabolic model and its responses to varying RuBisCO activity, light intensity, and enzymatic cost conditions. Front. Plant Sci. 8: 1–18.

Cheng, Y.-L. et al. (2016). Grape and Wine Metabolites: Biotechnological Approaches to Improve Wine Quality. In Intech, p. 13.

Cheung, C.Y.M., Williams, T.C.R., Poolman, M.G., Fell, D.A., Ratcliffe, R.G., and Sweetlove, L.J. (2013). A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. Plant J. 75: 1050–1061.

Collakova, E., Yen, J.Y., and Senger, R.S. (2012). Are we ready for genome-scale modeling in plants? Plant Sci. 191–192: 53–70.

Considine, M.J. and Foyer, C.H. (2015). Metabolic responses to sulfur dioxide in grapevine (Vitis vinifera L.): Photosynthetic tissues and berries. Front. Plant Sci. 6: 127145.

Cruz, F., Capela, J., Ferreira, E.C., Rocha, M., and Dias, O. (2024). BioISO: an objective-oriented application for assisting the curation of genome-scale metabolic models. IEEE/ACM Trans. Comput. Biol. Bioinforma. PP: 1–13.

Cunha, E., Lagoa, D., Faria, J.P., Liu, F., Henry, C.S., and Dias, O. (2023a). TranSyT, an innovative framework for identifying transport systems. Bioinformatics 39.

Cunha, E., Silva, M., Chaves, I., Demirci, H., Lagoa, D.R., Lima, D., Rocha, M., Rocha, I., and Dias, O. (2023b). The first multi-tissue genome-scale metabolic model of a woody plant highlights suberin

biosynthesis pathways in Quercus suber. PLOS Comput. Biol. 19: e1011499.

Dal'Molin, C.G. de O., Quek, L.E., Palfreyman, R.W., Brumbley, S.M., and Nielsen, L.K. (2010). AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol. 152: 579–589.

Dias, O., Rocha, M., Ferreira, E.C., and Rocha, I. (2015). Reconstructing genome-scale metabolic models with merlin. Nucleic Acids Res. 43: 3899–3910.

Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. 7: 1–6.

Fasoli, M., Richter, C.L., Zenoni, S., Bertini, E., Vitulo, N., Dal Santo, S., Dokoozlian, N., Pezzotti, M., and Tornielli, G.B. (2018). Timing and order of the molecular events marking the onset of berry ripening in grapevine. Plant Physiol. 178: 1187–1206.

Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., and Palsson, B.Ø. (2008). Reconstruction of Biochemical Networks in Microbial Organisms. Nat. Rev. Microbiol.

Ferreira, J., Vieira, V., Gomes, J., Correia, S., and Rocha, M. (2020). Troppo - A Python Framework for the Reconstruction of Context-Specific Metabolic Models. Adv. Intell. Syst. Comput. 1005: 146–153.

Gauthier, P.P.G., Bligny, R., Gout, E., Mahé, A., Nogués, S., Hodges, M., and Tcherkez, G.G.B. (2010). In folio isotopic tracing demonstrates that nitrogen assimilation into glutamate is mostly independent from current CO2 assimilation in illuminated leaves of Brassica napus. New Phytol. 185: 988–999.

Goldberg, T. et al. (2014). LocTree3 prediction of localization. Nucleic Acids Res. 42: W350.

Gomes de Oliveira Dal'Molin, C. and Nielsen, L.K. (2018). Plant genome-scale reconstruction: from single cell to multi-tissue modelling and omics analyses. Curr. Opin. Biotechnol. 49: 42–48.

Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., and Lee, S.Y. (2019). Current status and applications of genome-scale metabolic models. Genome Biol. 20: 1–18.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35.

Huang, H. and Dong, Z. (2013). Research on architecture and query performance based on distributed graph database Neo4j. 2013 3rd Int. Conf. Consum. Electron. Commun. Networks, CECNet 2013 - Proc.: 533–536.

International Organisation of Vine and Wine (2023). STATE OF THE WORLD VINE AND WINE SECTOR IN 2022.

Jose, B. and Abraham, S. (2017). Exploring the merits of nosql: A study based on mongodb. 2017 Int. Conf. Networks Adv. Comput. Technol. NetACT 2017: 266–271.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45: D353–D361.

Kaul, S. et al. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nat. 2000 4086814 408: 796–815.

Kieber, J.J. and Schaller, G.E. (2014). Cytokinins. Arabidopsis Book 12: e0168.

Kim, Y., Kim, G.B., and Lee, S.Y. (2021). Machine learning applications in genome-scale metabolic modeling. Curr. Opin. Syst. Biol. 25: 42–49.

Lakshmanan, M., Lim, S.H., Mohanty, B., Kim, J.K., Ha, S.H., and Lee, D.Y. (2015). Unraveling the light-specific metabolic and regulatory signatures of rice through combined in silico modeling and multiomics analysis. Plant Physiol. 169: 3002–3020.

1086 Lewis, N.E. et al. (2010). Omic data from evolved E. coli are consistent with computed optimal growth
1087 from genome-scale models. Mol. Syst. Biol. 6.

1088 Lundberg, S.M., Allen, P.G., and Lee, S.-I. A Unified Approach to Interpreting Model Predictions.

1089 Van Der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. 9: 2579–
1090 2605.

1091 Mahadevan, R. and Schilling, C.H. (2003). The effects of alternate optimal solutions in constraint-based
1092 genome-scale metabolic models. Metab. Eng. 5: 264–276.

1093 Massonnet, M., Fasoli, M., Tornielli, G.B., Altieri, M., Sandri, M., Zuccolotto, P., Paci, P., Gardiman, M.,
1094 Zenoni, S., and Pezzotti, M. (2017a). Ripening transcriptomic program in red and white grapevine
1095 varieties correlates with berry skin anthocyanin accumulation. Plant Physiol. 174: 2376–2396.

1096 Massonnet, M., Figueroa-Balderas, R., Galarneau, E.R.A., Miki, S., Lawrence, D.P., Sun, Q., Wallis,
1097 C.M., Baumgartner, K., and Cantu, D. (2017b). Neofusicoccum parvum colonization of the
1098 grapevine woody stem triggers asynchronous host responses at the site of infection and in the
1099 leaves. Front. Plant Sci. 8.

1100 Maurice Cheung, C.Y., Poolman, M.G., Fell, D.A., George Ratcliffe, R., and Sweetlove, L.J. (2014). A diel
1101 flux balance model captures interactions between light and dark metabolism during day-night cycles
1102 in C3 and crassulacean acid metabolism leaves. Plant Physiol. 165: 917–929.

1103 Moreira, T.B., Shaw, R., Luo, X., Ganguly, O., Kim, H.S., Coelho, L.G.F., Cheung, C.Y.M., and Williams,
1104 T.C.R. (2019). A genome-scale metabolic model of soybean (Glycine max) highlights metabolic
1105 fluxes in seedlings. Plant Physiol. 180: 1912–1929.

1106 Nanda, P. and Ghosh, A. (2021). Genome scale-differential flux analysis reveals deregulation of lung cell
1107 metabolism on SARS-CoV-2 infection. PLoS Comput. Biol. 17: 1–22.

1108 Niemann, G.J., Pureveen, J.B.M., Eijkel, G.B., Poorter, H., and Boon, J.J. (1995). Differential chemical
1109 allocation and plant adaptation: A Py-MS Study of 24 species differing in relative growth rate. Plant
1110 Soil 175: 275–289.

1111 de Oliveira Dal'Molin, C.G., Quek, L.E., Saa, P.A., and Nielsen, L.K. (2015). A multi-tissue genome-scale
1112 metabolic modeling framework for the analysis of whole plant systems. Front. Plant Sci. 6: 1–12.

1113 Pfau, T., Christian, N., Masakapalli, S.K., Sweetlove, L.J., Poolman, M.G., and Ebenhöh, O. (2018). The
1114 intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. Sci.
1115 Rep. 8: 1–11.

1116 Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J., and Schomburg, D. (2017).
1117 BRENDA in 2017: new perspectives and new tools in BRENDA. Nucleic Acids Res. 45: D380–
1118 D388.

1119 Poolman, M.G., Kundu, S., Shaw, R., and Fell, D.A. (2013). Responses to light intensity in a genome-
1120 scale model of rice metabolism. Plant Physiol. 162: 1060–1072.

1121 Poolman, M.G., Miguet, L., Sweetlove, L.J., and Fell, D.A. (2009). A genome-scale metabolic model of
1122 Arabidopsis and some of its properties. Plant Physiol. 151: 1570–1581.

1123 Rana, P., Berry, C., Ghosh, P., and Fong, S.S. (2020). Recent advances on constraint-based models by
1124 integrating machine learning. Curr. Opin. Biotechnol. 64: 85–91.

1125 Richelle, A., Joshi, C., and Lewis, N.E. (2019). Assessing key decisions for transcriptomic data integration
1126 in biochemical networks. PLoS Comput. Biol. 15: 1–18.

1127 Saad, N.M., Sekar, M., Gan, S.H., Lum, P.T., Vaijanathappa, J., and Ravi, S. (2020). Resveratrol: Latest
1128 scientific evidences of its chemical, biological activities and therapeutic potentials. Pharmacogn. J.
1129 12: 1779–1791.

1130 Saha, R., Suthers, P.F., and Maranas, C.D. (2011). Zea mays irs1563: A comprehensive genome-scale
1131 metabolic reconstruction of maize metabolism. PLoS One 6.

1132 Sampaio, M., Rocha, M., and Dias, O. (2022). Exploring synergies between plant metabolic modelling
1133 and machine learning. Comput. Struct. Biotechnol. J. 20: 1885–1900.

1134 Santos, S. and Rocha, I. (2016). Estimation of biomass composition from genomic and transcriptomic
1135 information. J. Integr. Bioinform. 13: 285.

1136 Sayers, E.W. et al. (2022). Database resources of the national center for biotechnology information.
1137 Nucleic Acids Res. 50: D20–D26.

1138 Shaw, R. and Cheung, C.Y.M. (2018). A dynamic multi-tissue flux balance model captures carbon and
1139 nitrogen metabolism and optimal resource partitioning during arabidopsis growth. Front. Plant Sci. 9:
1140 1–15.

1141 Shaw, R. and Cheung, C.Y.M. (2020). Multi-tissue to whole plant metabolic modelling. Cell. Mol. Life Sci.
1142 77: 489–495.

1143 Shaw, R. and Maurice Cheung, C.Y. (2019). A mass and charge balanced metabolic model of Setaria
1144 viridis revealed mechanisms of proton balancing in C4 plants. BMC Bioinformatics 20: 1–11.

1145 Simons, M., Saha, R., Amiour, N., Kumar, A., Guillard, L., Clément, G., Miquel, M., Li, Z., Mouille, G., Lea,
1146 P.J., Hirel, B., and Maranas, C.D. (2014). Assessing the metabolic impact of nitrogen availability
1147 using a compartmentalized maize leaf genome-scale model. Plant Physiol. 166: 1659–1674.

1148 Singh, J., Kumar, M., Sharma, A., Pandey, G., Chae, K., and Lee, S. (2016). Phenolic Compounds of
1149 Grapes and Wines: Key Compounds and Implications in Sensory Perception.

1150 Sweetlove, L.J., Beard, K.F.M., Nunes-Nesi, A., Fernie, A.R., and Ratcliffe, R.G. (2010). Not just a circle:
1151 flux modes in the plant TCA cycle. Trends Plant Sci. 15: 462–470.

1152 Sweetlove, L.J. and George Ratcliffe, R. (2011). Flux-balance modeling of plant metabolism. Front. Plant
1153 Sci. 2: 1–10.

1154 The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45:
1155 D158.

1156 Varma, A. and Palsson, B.O. (1994). Stoichiometric flux balance models quantitatively predict growth and
1157 metabolic by-product secretion in wild-type Escherichia coli W3110. Appl. Environ. Microbiol. 60:
1158 3724–31.

1159 Velt, A. et al. (2023). An improved reference of the grapevine genome reasserts the origin of the
1160 PN40024 highly homozygous genotype. G3 Genes|Genomes|Genetics 13: 67.

1161 Velt, A. and Rustenholz, C. (2023). GREAT_PN40024.v4.2_beta.

1162 Vlassis, N., Pacheco, M.P., and Sauter, T. (2014). Fast Reconstruction of Compact Context-Specific
1163 Metabolic Network Models. PLoS Comput. Biol. 10.

1164 Wasternack, C. and Song, S. (2017). Jasmonates: biosynthesis, metabolism, and signaling by proteins
1165 activating and repressing transcription. J. Exp. Bot. 68: 1303–1321.

1166 Williams, T.C.R., Miguet, L., Masakapalli, S.K., Kruger, N.J., Sweetlove, L.J., and Ratcliffe, R.G. (2008).
1167 Metabolic Network Fluxes in Heterotrophic Arabidopsis Cells: Stability of the Flux Distribution under
1168 Different Oxygenation Conditions. Plant Physiol. 148: 704–718.

1169 Yuan, H., Cheung, C.Y.M., Poolman, M.G., Hilbers, P.A.J., and van Riel, N.A.W. (2016). A genome-scale
1170 metabolic network reconstruction of tomato (Solanum lycopersicum L.) and its application to
1171 photorespiratory metabolism. Plant J. 85: 289–304.

1172 Zampieri, G., Coggins, M., Valle, G., and Angione, C. (2017). A poly-omics machine-learning method to

predict metabolite production in CHO cells. In 2nd International Electronic Conference on Metabolomics, p. 4993.

Zhang, P. et al. (2010). Creation of a genome-wide metabolic pathway database for populus trichocarpa using a new approach for reconstruction and curation of metabolic pathways for plants. Plant Physiol. 153: 1479–1491.

Parsed Citations

Antonakoudis, A., Barbosa, R., Kotidis, P., and Kontoravdi, C. (2020). The era of big data: Genome-scale modelling meets machine learning. Comput. Struct. Biotechnol. J. 18: 3287–3300.
Google Scholar: Author Only Title Only Author and Title

Bogart, E. and Myers, C.R. (2016). Multiscale metabolic modeling of C4 plants: Connecting nonlinear genome-scale models to leaf-scale metabolism in developing maize leaves. PLoS One 11: 1–27.
Google Scholar: Author Only Title Only Author and Title

Bordel, S., Agren, R., and Nielsen, J. (2010). Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes. PLOS Comput. Biol. 6: e1000859.
Google Scholar: Author Only Title Only Author and Title

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 2014 121 12: 59–60.
Google Scholar: Author Only Title Only Author and Title

Caspi, R. et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 44: D471–D480.
Google Scholar: Author Only Title Only Author and Title

Chatterjee, A., Huma, B., Shaw, R., and Kundu, S. (2017). Reconstruction of Oryza sativa indica genome scale metabolic model and its responses to varying RuBisCO activity, light intensity, and enzymatic cost conditions. Front. Plant Sci. 8: 1–18.
Google Scholar: Author Only Title Only Author and Title

Cheng, Y.-L. et al. (2016). Grape and Wine Metabolites: Biotechnological Approaches to Improve Wine Quality. In Intech, p. 13.
Google Scholar: Author Only Title Only Author and Title

Cheung, C.Y.M., Williams, T.C.R., Poolman, M.G., Fell, D.A., Ratcliffe, R.G., and Sweetlove, L.J. (2013). A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. Plant J. 75: 1050–1061.
Google Scholar: Author Only Title Only Author and Title

Collakova, E., Yen, J.Y., and Senger, R.S. (2012). Are we ready for genome-scale modeling in plants? Plant Sci. 191–192: 53–70.
Google Scholar: Author Only Title Only Author and Title

Considine, M.J. and Foyer, C.H. (2015). Metabolic responses to sulfur dioxide in grapevine (Vitis vinifera L.): Photosynthetic tissues and berries. Front. Plant Sci. 6: 127145.
Google Scholar: Author Only Title Only Author and Title

Cruz, F., Capela, J., Ferreira, E.C., Rocha, M., and Dias, O. (2024). BioISO: an objective-oriented application for assisting the curation of genome-scale metabolic models. IEEE/ACM Trans. Comput. Biol. Bioinforma. PP: 1–13.
Google Scholar: Author Only Title Only Author and Title

Cunha, E., Lagoa, D., Faria, J.P., Liu, F., Henry, C.S., and Dias, O. (2023a). TranSyT, an innovative framework for identifying transport systems. Bioinformatics 39.
Google Scholar: Author Only Title Only Author and Title

Cunha, E., Silva, M., Chaves, I., Demirci, H., Lagoa, D.R., Lima, D., Rocha, M., Rocha, I., and Dias, O. (2023b). The first multi-tissue genome-scale metabolic model of a woody plant highlights suberin biosynthesis pathways in Quercus suber. PLOS Comput. Biol. 19: e1011499.
Google Scholar: Author Only Title Only Author and Title

Dal'Molin, C.G. de O., Quek, L.E., Palfreyman, R.W., Brumbley, S.M., and Nielsen, L.K. (2010). AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol. 152: 579–589.
Google Scholar: Author Only Title Only Author and Title

Dias, O., Rocha, M., Ferreira, E.C., and Rocha, I. (2015). Reconstructing genome-scale metabolic models with merlin. Nucleic Acids Res. 43: 3899–3910.
Google Scholar: Author Only Title Only Author and Title

Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. 7: 1–6.
Google Scholar: Author Only Title Only Author and Title

Fasoli, M., Richter, C.L., Zenoni, S., Bertini, E., Vitulo, N., Dal Santo, S., Dokoozlian, N., Pezzotti, M., and Tornielli, G.B. (2018). Timing and order of the molecular events marking the onset of berry ripening in grapevine. Plant Physiol. 178: 1187–1206.
Google Scholar: Author Only Title Only Author and Title

Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., and Palsson, B.Ø. (2008). Reconstruction of Biochemical Networks in Microbial Organisms. Nat. Rev. Microbiol.
Google Scholar: Author Only Title Only Author and Title

Ferreira, J., Vieira, V., Gomes, J., Correia, S., and Rocha, M. (2020). Troppo - A Python Framework for the Reconstruction of Context-Specific Metabolic Models. Adv. Intell. Syst. Comput. 1005: 146–153.
Google Scholar: Author Only Title Only Author and Title

Gauthier, P.P.G., Bligny, R., Gout, E., Mahé, A., Nogués, S., Hodges, M., and Tcherkez, G.G.B. (2010). In folio isotopic tracing demonstrates that nitrogen assimilation into glutamate is mostly independent from current CO2 assimilation in illuminated leaves of Brassica napus. New Phytol. 185: 988–999.
Google Scholar: Author Only Title Only Author and Title

Goldberg, T. et al. (2014). LocTree3 prediction of localization. Nucleic Acids Res. 42: W350.
Google Scholar: Author Only Title Only Author and Title

Gomes de Oliveira Dal'Molin, C. and Nielsen, L.K. (2018). Plant genome-scale reconstruction: from single cell to multi-tissue modelling and omics analyses. Curr. Opin. Biotechnol. 49: 42–48.
Google Scholar: Author Only Title Only Author and Title

Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., and Lee, S.Y. (2019). Current status and applications of genome-scale metabolic models. Genome Biol. 20: 1–18.
Google Scholar: Author Only Title Only Author and Title

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35.
Google Scholar: Author Only Title Only Author and Title

Huang, H. and Dong, Z. (2013). Research on architecture and query performance based on distributed graph database Neo4j. 2013 3rd Int. Conf. Consum. Electron. Commun. Networks, CECNet 2013 - Proc.: 533–536.
Google Scholar: Author Only Title Only Author and Title

International Organisation of Vine and Wine (2023). STATE OF THE WORLD VINE AND WINE SECTOR IN 2022.

Jose, B. and Abraham, S. (2017). Exploring the merits of nosql: A study based on mongodb. 2017 Int. Conf. Networks Adv. Comput. Technol. NetACT 2017: 266–271.
Google Scholar: Author Only Title Only Author and Title

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45: D353–D361.
Google Scholar: Author Only Title Only Author and Title

Kaul, S. et al. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nat. 2000 4086814 408: 796–815.
Google Scholar: Author Only Title Only Author and Title

Kieber, J.J. and Schaller, G.E. (2014). Cytokinins. Arabidopsis Book 12: e0168.
Google Scholar: Author Only Title Only Author and Title

Kim, Y., Kim, G.B., and Lee, S.Y. (2021). Machine learning applications in genome-scale metabolic modeling. Curr. Opin. Syst. Biol. 25: 42–49.
Google Scholar: Author Only Title Only Author and Title

Lakshmanan, M., Lim, S.H., Mohanty, B., Kim, J.K., Ha, S.H., and Lee, D.Y. (2015). Unraveling the light-specific metabolic and regulatory signatures of rice through combined in silico modeling and multiomics analysis. Plant Physiol. 169: 3002–3020.
Google Scholar: Author Only Title Only Author and Title

Lewis, N.E. et al. (2010). Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6.
Google Scholar: Author Only Title Only Author and Title

Lundberg, S.M., Allen, P.G., and Lee, S.-I. A Unified Approach to Interpreting Model Predictions.

Van Der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. 9: 2579–2605.
Google Scholar: Author Only Title Only Author and Title

Mahadevan, R. and Schilling, C.H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5: 264–276.
Google Scholar: Author Only Title Only Author and Title

Massonnet, M., Fasoli, M., Tornielli, G.B., Altieri, M., Sandri, M., Zuccolotto, P., Paci, P., Gardiman, M., Zenoni, S., and Pezzotti, M.

(2017a). Ripening transcriptomic program in red and white grapevine varieties correlates with berry skin anthocyanin accumulation. Plant Physiol. 174: 2376–2396.
Google Scholar: Author Only Title Only Author and Title

Massonnet, M., Figueroa-Balderas, R., Galarneau, E.R.A., Miki, S., Lawrence, D.P., Sun, Q., Wallis, C.M., Baumgartner, K., and Cantu, D. (2017b). Neofusicoccum parvum colonization of the grapevine woody stem triggers asynchronous host responses at the site of infection and in the leaves. Front. Plant Sci. 8.
Google Scholar: Author Only Title Only Author and Title

Maurice Cheung, C.Y., Poolman, M.G., Fell, D.A., George Ratcliffe, R., and Sweetlove, L.J. (2014). A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C3 and crassulacean acid metabolism leaves. Plant Physiol. 165: 917–929.
Google Scholar: Author Only Title Only Author and Title

Moreira, T.B., Shaw, R., Luo, X., Ganguly, O., Kim, H.S., Coelho, L.G.F., Cheung, C.Y.M., and Williams, T.C.R. (2019). A genome-scale metabolic model of soybean (Glycine max) highlights metabolic fluxes in seedlings. Plant Physiol. 180: 1912–1929.
Google Scholar: Author Only Title Only Author and Title

Nanda, P. and Ghosh, A. (2021). Genome scale-differential flux analysis reveals deregulation of lung cell metabolism on SARS-CoV-2 infection. PLoS Comput. Biol. 17: 1–22.
Google Scholar: Author Only Title Only Author and Title

Niemann, G.J., Pureveen, J.B.M., Eijkel, G.B., Poorter, H., and Boon, J.J. (1995). Differential chemical allocation and plant adaptation: A Py-MS Study of 24 species differing in relative growth rate. Plant Soil 175: 275–289.
Google Scholar: Author Only Title Only Author and Title

de Oliveira Dal'Molin, C.G., Quek, L.E., Saa, P.A., and Nielsen, L.K. (2015). A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. Front. Plant Sci. 6: 1–12.
Google Scholar: Author Only Title Only Author and Title

Pfau, T., Christian, N., Masakapalli, S.K., Sweetlove, L.J., Poolman, M.G., and Ebenhöh, O. (2018). The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. Sci. Rep. 8: 1–11.
Google Scholar: Author Only Title Only Author and Title

Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J., and Schomburg, D. (2017). BRENDA in 2017: new perspectives and new tools in BRENDA. Nucleic Acids Res. 45: D380–D388.
Google Scholar: Author Only Title Only Author and Title

Poolman, M.G., Kundu, S., Shaw, R., and Fell, D.A. (2013). Responses to light intensity in a genome-scale model of rice metabolism. Plant Physiol. 162: 1060–1072.
Google Scholar: Author Only Title Only Author and Title

Poolman, M.G., Miguet, L., Sweetlove, L.J., and Fell, D.A. (2009). A genome-scale metabolic model of Arabidopsis and some of its properties. Plant Physiol. 151: 1570–1581.
Google Scholar: Author Only Title Only Author and Title

Rana, P., Berry, C., Ghosh, P., and Fong, S.S. (2020). Recent advances on constraint-based models by integrating machine learning. Curr. Opin. Biotechnol. 64: 85–91.
Google Scholar: Author Only Title Only Author and Title

Richelle, A., Joshi, C., and Lewis, N.E. (2019). Assessing key decisions for transcriptomic data integration in biochemical networks. PLoS Comput. Biol. 15: 1–18.
Google Scholar: Author Only Title Only Author and Title

Saad, N.M., Sekar, M., Gan, S.H., Lum, P.T., Vaijanathappa, J., and Ravi, S. (2020). Resveratrol: Latest scientific evidences of its chemical, biological activities and therapeutic potentials. Pharmacogn. J. 12: 1779–1791.
Google Scholar: Author Only Title Only Author and Title

Saha, R., Suthers, P.F., and Maranas, C.D. (2011). Zea mays irs1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS One 6.
Google Scholar: Author Only Title Only Author and Title

Sampaio, M., Rocha, M., and Dias, O. (2022). Exploring synergies between plant metabolic modelling and machine learning. Comput. Struct. Biotechnol. J. 20: 1885–1900.
Google Scholar: Author Only Title Only Author and Title

Santos, S. and Rocha, I. (2016). Estimation of biomass composition from genomic and transcriptomic information. J. Integr. Bioinform. 13: 285.
Google Scholar: Author Only Title Only Author and Title

Sayers, E.W. et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. 50: D20–D26.
Google Scholar: Author Only Title Only Author and Title

Shaw, R. and Cheung, C.Y.M. (2018). A dynamic multi-tissue flux balance model captures carbon and nitrogen metabolism and optimal resource partitioning during arabidopsis growth. Front. Plant Sci. 9: 1–15.
Google Scholar: Author Only Title Only Author and Title

Shaw, R. and Cheung, C.Y.M. (2020). Multi-tissue to whole plant metabolic modelling. Cell. Mol. Life Sci. 77: 489–495.
Google Scholar: Author Only Title Only Author and Title

Shaw, R. and Maurice Cheung, C.Y. (2019). A mass and charge balanced metabolic model of Setaria viridis revealed mechanisms of proton balancing in C4 plants. BMC Bioinformatics 20: 1–11.
Google Scholar: Author Only Title Only Author and Title

Simons, M., Saha, R., Amiour, N., Kumar, A., Guillard, L., Clément, G., Miquel, M., Li, Z., Mouille, G., Lea, P.J., Hirel, B., and Maranas, C.D. (2014). Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. Plant Physiol. 166: 1659–1674.
Google Scholar: Author Only Title Only Author and Title

Singh, J., Kumar, M., Sharma, A., Pandey, G., Chae, K., and Lee, S. (2016). Phenolic Compounds of Grapes and Wines: Key Compounds and Implications in Sensory Perception.

Sweetlove, L.J., Beard, K.F.M., Nunes-Nesi, A., Fernie, A.R., and Ratcliffe, R.G. (2010). Not just a circle: flux modes in the plant TCA cycle. Trends Plant Sci. 15: 462–470.
Google Scholar: Author Only Title Only Author and Title

Sweetlove, L.J. and George Ratcliffe, R. (2011). Flux-balance modeling of plant metabolism. Front. Plant Sci. 2: 1–10.
Google Scholar: Author Only Title Only Author and Title

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45: D158.
Google Scholar: Author Only Title Only Author and Title

Varma, A. and Palsson, B.O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl. Environ. Microbiol. 60: 3724–31.
Google Scholar: Author Only Title Only Author and Title

Velt, A. et al. (2023). An improved reference of the grapevine genome reasserts the origin of the PN40024 highly homozygous genotype. G3 Genes|Genomes|Genetics 13: 67.
Google Scholar: Author Only Title Only Author and Title

Velt, A. and Rustenholz, C. (2023). GREAT_PN40024.v4.2_beta.

Vlassis, N., Pacheco, M.P., and Sauter, T. (2014). Fast Reconstruction of Compact Context-Specific Metabolic Network Models. PLoS Comput. Biol. 10.
Google Scholar: Author Only Title Only Author and Title

Wasternack, C. and Song, S. (2017). Jasmonates: biosynthesis, metabolism, and signaling by proteins activating and repressing transcription. J. Exp. Bot. 68: 1303–1321.
Google Scholar: Author Only Title Only Author and Title

Williams, T.C.R., Miguet, L., Masakapalli, S.K., Kruger, N.J., Sweetlove, L.J., and Ratcliffe, R.G. (2008). Metabolic Network Fluxes in Heterotrophic Arabidopsis Cells: Stability of the Flux Distribution under Different Oxygenation Conditions. Plant Physiol. 148: 704–718.
Google Scholar: Author Only Title Only Author and Title

Yuan, H., Cheung, C.Y.M., Poolman, M.G., Hilbers, P.A.J., and van Riel, N.A.W. (2016). A genome-scale metabolic network reconstruction of tomato (Solanum lycopersicum L.) and its application to photorespiratory metabolism. Plant J. 85: 289–304.
Google Scholar: Author Only Title Only Author and Title

Zampieri, G., Coggins, M., Valle, G., and Angione, C. (2017). A poly-omics machine-learning method to predict metabolite production in CHO cells. In 2nd International Electronic Conference on Metabolomics, p. 4993.
Google Scholar: Author Only Title Only Author and Title

Zhang, P. et al. (2010). Creation of a genome-wide metabolic pathway database for populus trichocarpa using a new approach for reconstruction and curation of metabolic pathways for plants. Plant Physiol. 153: 1479–1491.
Google Scholar: Author Only Title Only Author and Title