Synthetic Biobots

Bioinformatics methods

Synthetic
Biobots

Author: Gerardo Cendejas-Mendoza

iGEM Design League

Season 2022

Introduction

In Synthetic Biobots we deal with a health and agriculture related problem, a harmful pest: the fall armyworm (*Spodoptera frugiperda*). It is easily found in the maize sold in mexican markets, which means that it is possible to find it even in products that have passed healthiness tests, and not even the most advanced pesticides or agricultural techniques can get rid of it completely. Since these products are often hazardous to health when used in excess or incorrectly, the solution is not very clear. In addition to human health, ecosystems are also affected by these treatments. With these concerns in mind, we raised the following research question: is it possible to create a good pesticide that is at the same time environmentally and health friendly?

The fall armyworm is widespread in various parts of the world. Its main host is maize plants, causing losses that exceed 50% of the production of low and medium potential, and about 30% for fields of higher productivity.[8] It also affects other crops such as sugarcane, beans and tomatoes, causing losses of up to 4.5 million Mexican pesos per crop. This pest is native to South America and has been reported in all Mexican states.[9]

There is a compound derived from black pepper (*Piper nigrum* L.) called piperine. Results have been obtained that indicate that at concentrations of 10 g/L and 20 g/L piperine has a pesticide capacity that inhibits the hatching and development of fall armyworm larvae, mainly of the most recently laid eggs, causing egg mortality of up to 88%.[19] Recent studies on the feasibility of plant extracts as pesticides against larvae of lepidoptera that infect common agricultural crops have suggested that in addition to piperine compounds such as β and α-asarone can have a pesticidal effect on *Spodoptera litura* larvae, a species related to the fall armyworm. It was found that the concentrations adequate to ensure the death of 50% of the larvae are 81.5, 6.24 and 2.22 micrograms per larva for peperin, β and α-asarone respectively. A synergistic effect was also shown in the biopesticide effect between extracts of *P. retrofractum* and *A. calamus*, whose main components are piperine for *P. retrofractum* (15.6%) and β and α-asarone, in a ratio of 4.38:1 for *A. calamus*.[22]

It also presents different activities that can make it a candidate for therapy to treat different diseases from a human perspective: slows growth tumor, decreases the concentration of glucose in the blood (diabetes mellitus), inhibits the production of reactive oxygen species, prevents alterations in the action potential of the heart, decreases the expression of genes of interleukins 4, 13 and necrosis factor alpha tumor (allergy), as well as many other effects that can help in the treatment of obesity, liver disorders, aging, cardiovascular diseases, etc.[20;11]

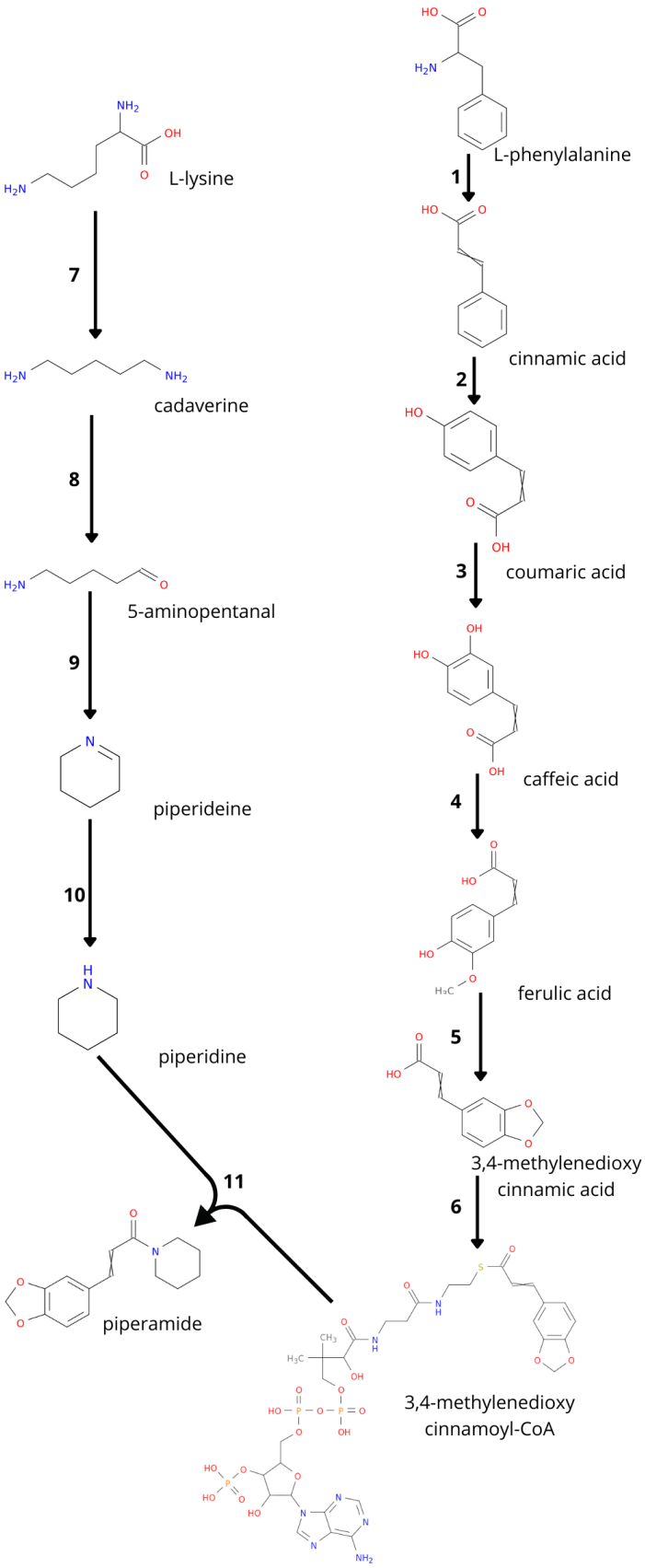
Piperine is synthesized from two precursors: piperoyl-CoA and piperidine, these are condensed by the enzyme piperine synthase to form piperine. Both precursors are synthesized from amino acids, piperoyl-CoA has phenylalanine as precursor and piperidine comes from lysine. The piperine biosynthetic pathway in the black pepper is not completely determined, it is known that phenylalanine reaches acid ferulic by the phenylpropanoid synthesis route, then it is extended in two carbons to feruperic acid which upon formation of a methylenedioxy bridge will form piperic acid, to this molecule an enzyme Pip-CoA ligase joins coenzyme A to form piperoyl-CoA. In the case of lysine, it is decarboxylated to form cadaverine, which by action of a monoamine oxidase is converted to 5-aminopentanal (KEGG: R06740), this molecule

cyclizes into piperideine (KEGG: R06741) which, when reduced, becomes piperidine. The enzyme that cyclizes 5-aminopentanal to piperideine is not still characterized.[16;17;18] In the Table 1 found in the Appendix section a list of all bioinformatic tools used for this project is shown.

Methods

In the black pepper fruit there are different compounds similar to piperine, which are collectively called piperamides and could have similar effects. It has been reported that a protein highly expressed in black pepper fruits similar to piperine synthase, called piperamide synthase, has a promiscuous substrate specificity and can accept different CoA donors to condense with piperidine and form a piperamide. One of the compounds that can be produces is a pieramide with similar structure similar to piperine but with two fewer carbons; this is result of the 3,4-Methylenedioxy cinnamoyl CoA as CoA donor instead of piperoyl CoA.[16] In order to see if this new piperamide could be a better option for a safe bio-pesticide active compuond we computed physicochemical descriptors and ADME parameters of both molecules using SwissADME tool.[7] One of the most important parameters calculated is skin permeability, this was computed using the QSPR model implemented by Abraham et al. 1995, the piperamide resulted with a lower skin permeability log k_p (-5.88 cm/s) than piperine (-5.58 cm/s), this makes the selected piperamide a better option since it will permeate less the skin of the farmers who use a piperamide-based bio-pesticide.[1;14] It also presents a better synthetic accessibility (2.59) compared to piperine (2.92).

Some of the enzymes involved in the biosynthetic pathway of piperine are known and have been characterized, the piperine synthase is one of these enzymes. Piperoyl CoA is formed by activation of piperic acid, this step is performed by a piperic acid CoA ligase, this enzyme was found by Schnabel et al. 2020. The methylendioxy bridge formed in feruperic acid in order to form piperic acid is performed by the enzyme CYP719A37. These enzymes were found by differential expression analysis of different *P. nigrum* tissues. Based on this information we propose a biosynthetic route for the synthesis of the selected piperamide, this can be seen in the figure shown in the next page.



Biosynthetic pathway of piperamide.

Anotation

Hu et al. 2019 provided the first ever reference genome of *P. nigrum*, with 761.2 Mb assembled in 26 pseudochromosomes, the complete assembly and genome sequence can be found in this link. The genes assembled in this project were not annotated for protein function, in order to know the function of the different genes of *P. nigrum* we performed two different annotation analyses. One analysis was performed using KofamKOALA, which is a Kegg ortholog assignment tool based on profile hmm and adaptive score threshold.[3] With this annotation a complete metabolome reconstruction was made using KEGG mapper reconstruct tool, the genes were anotated to 3803 different KEGG orthology IDs, 948 of these IDs were part of the Metabolic pathways present in the KEGG database (ID: 01100). The complete modules of the metabolic pathways are shown in the next list and in Figure 1. We found 455 different KEGG orthology IDs for biosynthesis of secondary metabolites pathways (ID: 01110), this can be seen in Figure 2.

Some of the more intersting pathways to this project are the phenylpropanoid biosynthesis pathway (ID: 00940) and the tropane, piperidine and pyridine alkaloid biosynthesis pathway (ID: 00960). In Figure 3 phenylpropanoid biosynthesis pathway can be seen, where some enzymes required in order to obtain ferulic acid from phenylalanine are shown in green, this are KEGG orthology IDs to which some genes from *P. nigrum* have been annotated. In the case of piperidine and pyridine alkaloid biosynthesis pathway, this information is shown in Figure 4, in this case the enzyme of interes is a primary-amine oxidase that performs the reaction that converts cadaverine into 5-Aminopentanal.

- Carbohydrate metabolism
  - Central carbohydrate metabolism
    - M00001 Glycolysis (Embden-Meyerhof pathway)
    - M00002 Glycolysis, core module involving three-carbon compounds
    - M00003 Gluconeogenesis
    - M00010 Citrate cycle, first carbon oxidation
    - M00004 Pentose phosphate pathway (Pentose phosphate cycle)
    - M00006 Pentose phosphate pathway, oxidative phase
    - M00007 Pentose phosphate pathway, non-oxidative phase
    - M00005 PRPP biosynthesis
  - Other carbohydrate metabolism
    - M00081 Pectin degradation
    - M00114 Ascorbate biosynthesis, plants
    - M00854 Glycogen biosynthesis
    - M00855 Glycogen degradation
    - M00549 Nucleotide sugar biosynthesis
    - M00132 Inositol phosphate metabolism
- Energy metabolism

  - Carbon fixation
    * M00165 Reductive pentose phosphate cycle (Calvin cycle)
    * M00166 Reductive pentose phosphate cycle
    * M00167 Reductive pentose phosphate cycle
    * M00168 CAM (Crassulacean acid metabolism), dark
    * M00169 CAM (Crassulacean acid metabolism), light
    * M00172 C4-dicarboxylic acid cycle, NADP - malic enzyme type
    * M00171 C4-dicarboxylic acid cycle, NAD - malic enzyme type
    * M00170 C4-dicarboxylic acid cycle, phosphoenolpyruvate carboxykinase type
  - Nitrogen metabolism
    * M00531 Assimilatory nitrate reduction
  - Sulfur metabolism
    * M00176 Assimilatory sulfate reduction

- Lipid metabolism
  - Fatty acid metabolism
    * M00082 Fatty acid biosynthesis, initiation
    * M00083 Fatty acid biosynthesis, elongation
    * M00415 Fatty acid elongation in endoplasmic reticulum
    * M00086 beta-Oxidation, acyl-CoA synthesis
    * M00087 beta-Oxidation
  - Lipid metabolism
    * M00089 Triacylglycerol biosynthesis
    * M00098 Acylglycerol degradation
    * M00092 Phosphatidylethanolamine (PE) biosynthesis
    * M00100 Sphingosine degradation
    * M00113 Jasmonic acid biosynthesis

- Nucleotide metabolism
  - Purine metabolism
    * M00049 Adenine ribonucleotide biosynthesis
    * M00050 Guanine ribonucleotide biosynthesis
    * M00053 Deoxyribonucleotide biosynthesis
  - Pyrimidine metabolism
    * M00052 Pyrimidine ribonucleotide biosynthesis

    - M00938 Pyrimidine deoxyribonucleotide biosynthesis
    - M00046 Pyrimidine degradation

- Amino acid metabolism
  - Serine and threonine metabolism
    - M00020 Serine biosynthesis
    - M00018 Threonine biosynthesis
    - M00555 Betaine biosynthesis
  - Cysteine and methionine metabolism
    - M00021 Cysteine biosynthesis
    - M00034 Methionine salvage pathway
    - M00368 Ethylene biosynthesis
  - Branched-chain amino acid metabolism
    - M00019 Valine/isoleucine biosynthesis
    - M00570 Isoleucine biosynthesis
    - M00432 Leucine biosynthesis
  - Arginine and proline metabolism
    - M00028 Ornithine biosynthesis
    - M00015 Proline biosynthesis
  - Aromatic amino acid metabolism
    - M00022 Shikimate pathway
    - M00023 Tryptophan biosynthesis
    - M00910 Phenylalanine biosynthesis
    - M00040 Tyrosine biosynthesis
    - M00044 Tyrosine degradation
  - Other amino acid metabolism
    - M00027 GABA (gamma-Aminobutyrate) shunt
    - M00118 Glutathione biosynthesis

- Glycan metabolism
  - Glycan biosynthesis
    - M00055 N-glycan precursor biosynthesis
    - M00072 N-glycosylation by oligosaccharyltransferase

- Metabolism of cofactors and vitamins

  - Cofactor and vitamin metabolism
    * M00899 Thiamine salvage pathway
    * M00125 Riboflavin biosynthesis, plants and bacteria
    * M00916 Pyridoxal-P biosynthesis
    * M00120 Coenzyme A biosynthesis
    * M00881 Lipoic acid biosynthesis, plants and bacteria
    * M00882 Lipoic acid biosynthesis, eukaryotes
    * M00126 Tetrahydrofolate biosynthesis
    * M00140 C1-unit interconversion, prokaryotes
    * M00112 Tocopherol/tocotorienol biosynthesis

- Biosynthesis of terpenoids and polyketides
  - Terpenoid backbone biosynthesis
    * M00095 C5 isoprenoid biosynthesis, mevalonate pathway
    * M00364 C10-C20 isoprenoid biosynthesis, bacteria
    * M00366 C10-C20 isoprenoid biosynthesis, plants
  - Plant terpenoid biosynthesis
    * M00097 beta-Carotene biosynthesis
    * M00372 Abscisic acid biosynthesis
    * M00371 Castasterone biosynthesis
    * M00927 Gibberellin A12 biosynthesis
    * M00928 Gibberellin A4/A1 biosynthesis

- Biosynthesis of other secondary metabolites
  - Biosynthesis of phytochemical compounds
    * M00039 Monolignol biosynthesis
    * M00137 Flavanone biosynthesis
    * M00138 Flavonoid biosynthesis

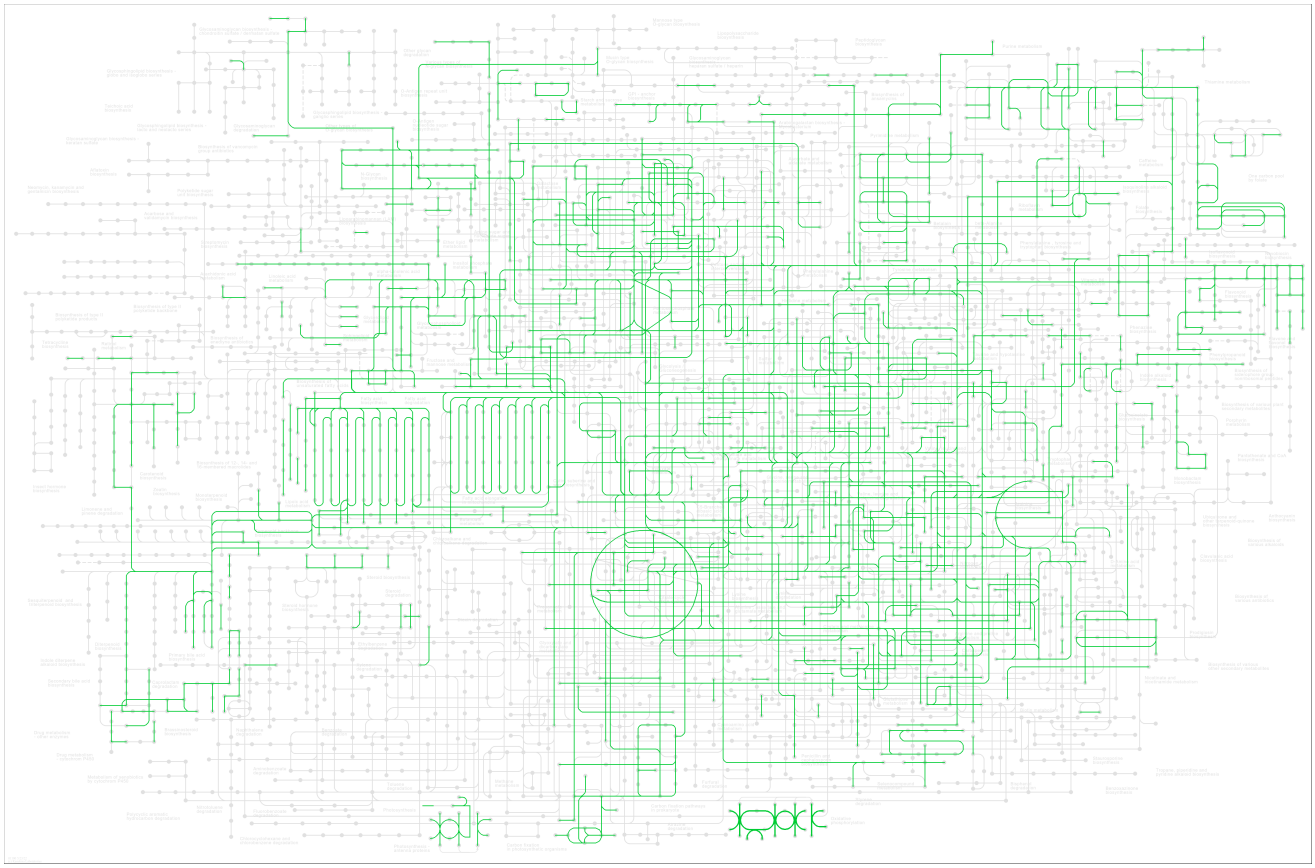


Figure 1: *Piper nigrum* metabolic pathways. KEGG ID 01100.

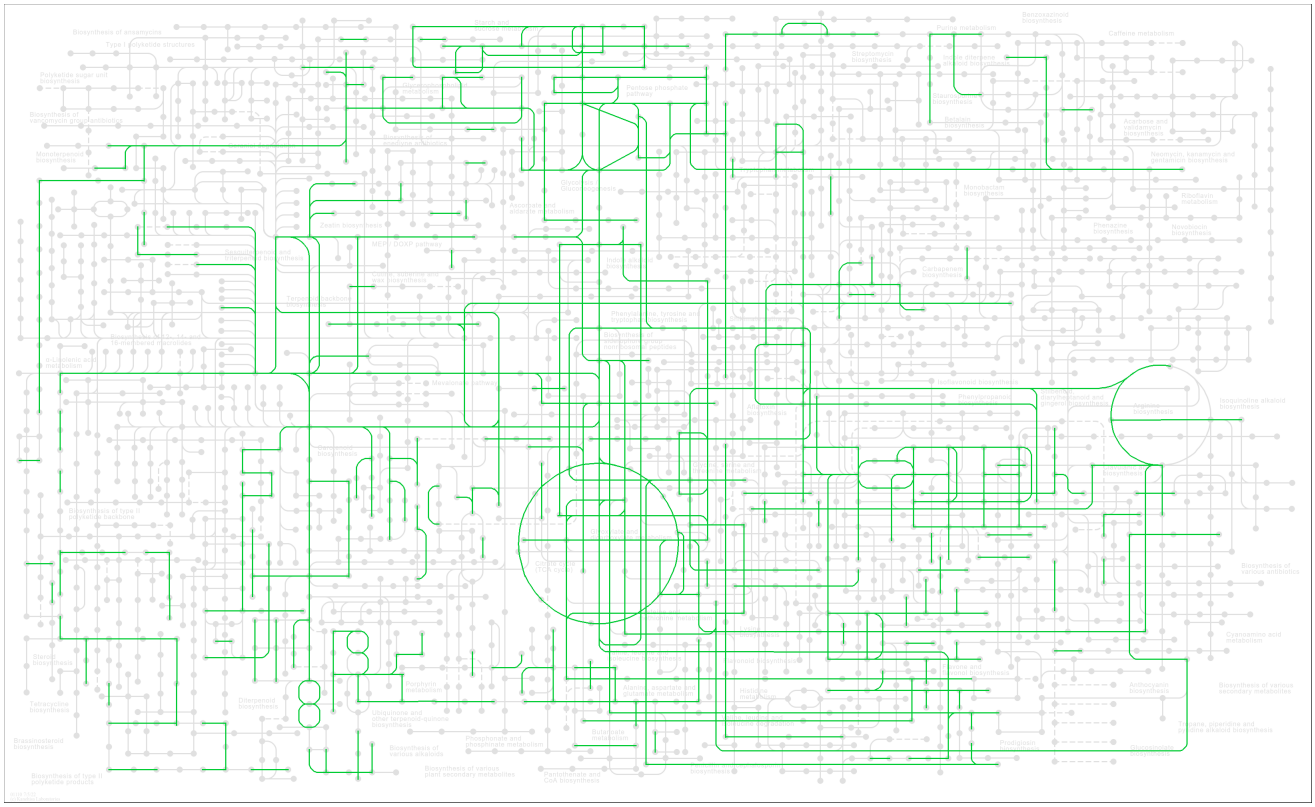


Figure 2: *Piper nigrum* biosynthesis of secondary metabolites pathways. KEGG ID 01110.

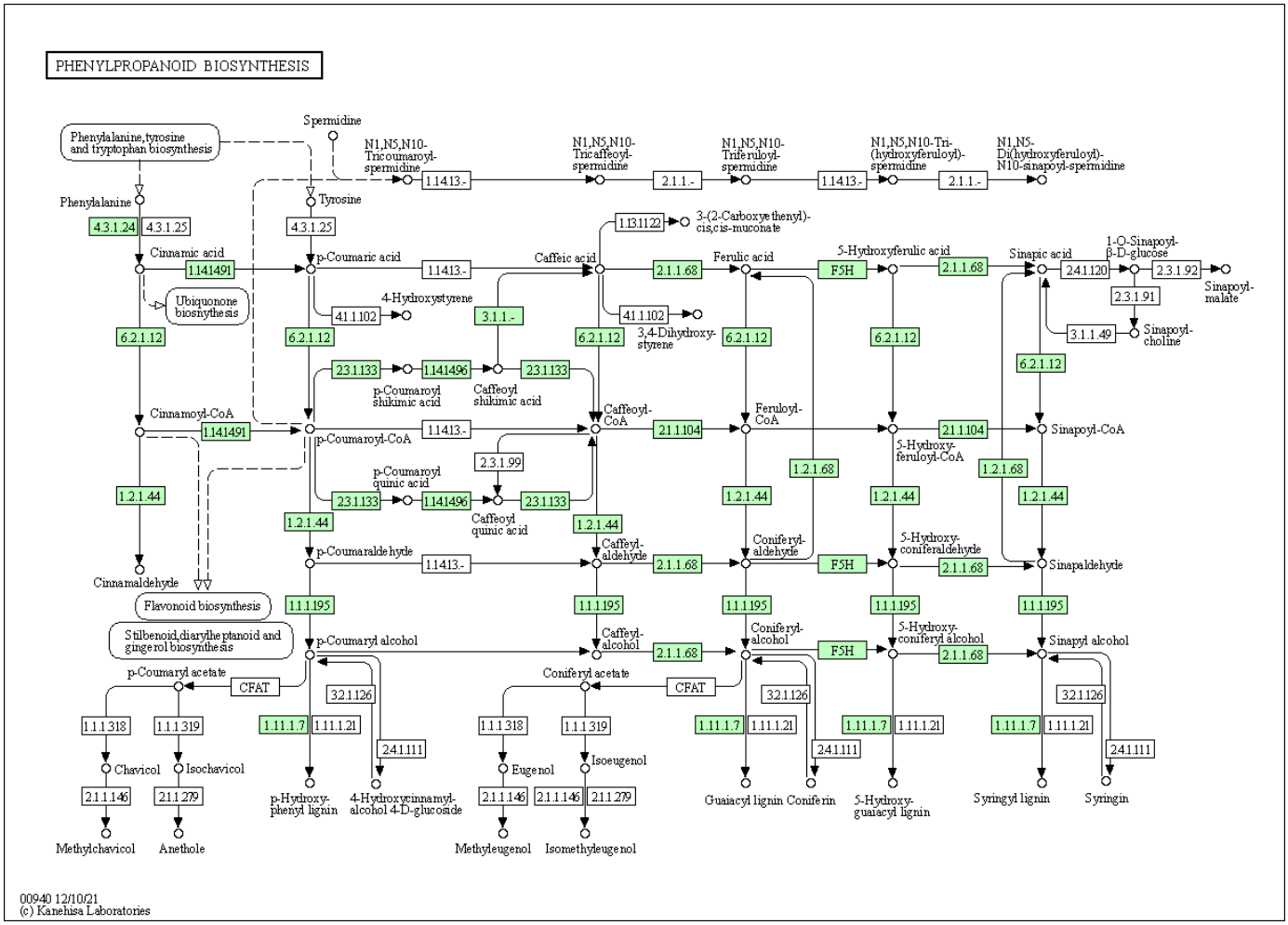


Figure 3: *Piper nigrum* phenylpropanoid biosynthesis pathway. KEGG ID 00940.

From this annotation some genes were found that corresponded to the KEGG orthology IDs responsible of performing some steps of the pathway. Pn8.2617 was annotated as a phenylalanine ammonia-lyase, the first enzyme of the proposed pathway, this enzyme was selected as the best candidate for enzyme 1 the fact that it obtained the highest score (1170.9) over the threshold (524.93) for KEGG Orthology ID K10775, with E-value $\approxeq 0$. The second enzyme of the pathway was also found in this way, Pn2.84 was selected as the best candidate for enzyme two of the proposed pathway based on the fact that it obtained the highest score (985.2) over the threshold (555.2) for KEGG Orthology ID K00487, with E-value $= 2.5 \times 10^{-297}$. Pn1.1317 was selected as the best candidate for enzyme number four based on the fact that it obtained the highest score (651.5) over the threshold (499.33) for KEGG Orthology ID K13066, with E-value $= 2.9 \times 10^{-196}$. The eighth enzyme of the reaction, Pn4.3222, was selected based on the fact that it obtained the highest score (889) over the threshold (550.3) for KEGG Orthology ID K00276, with E-value $= 2 \times 10^{-267}$.

The second annotation method used was the GeneFamilyClassifier tool, which is part of PlantTribes, a gene and gene family resource for comparative genomics in plants; this analysis was performed using the Galaxy project platform, a platform for accessible, reproducible and collaborative biomedical analyses. [21;6] 45,155 genes were annotated for AHRD, TAIR, Pfam domains, InterProScan Descriptions, GO Molecular Functions, Biological Processes and Cellular Components; this file can be reached here.

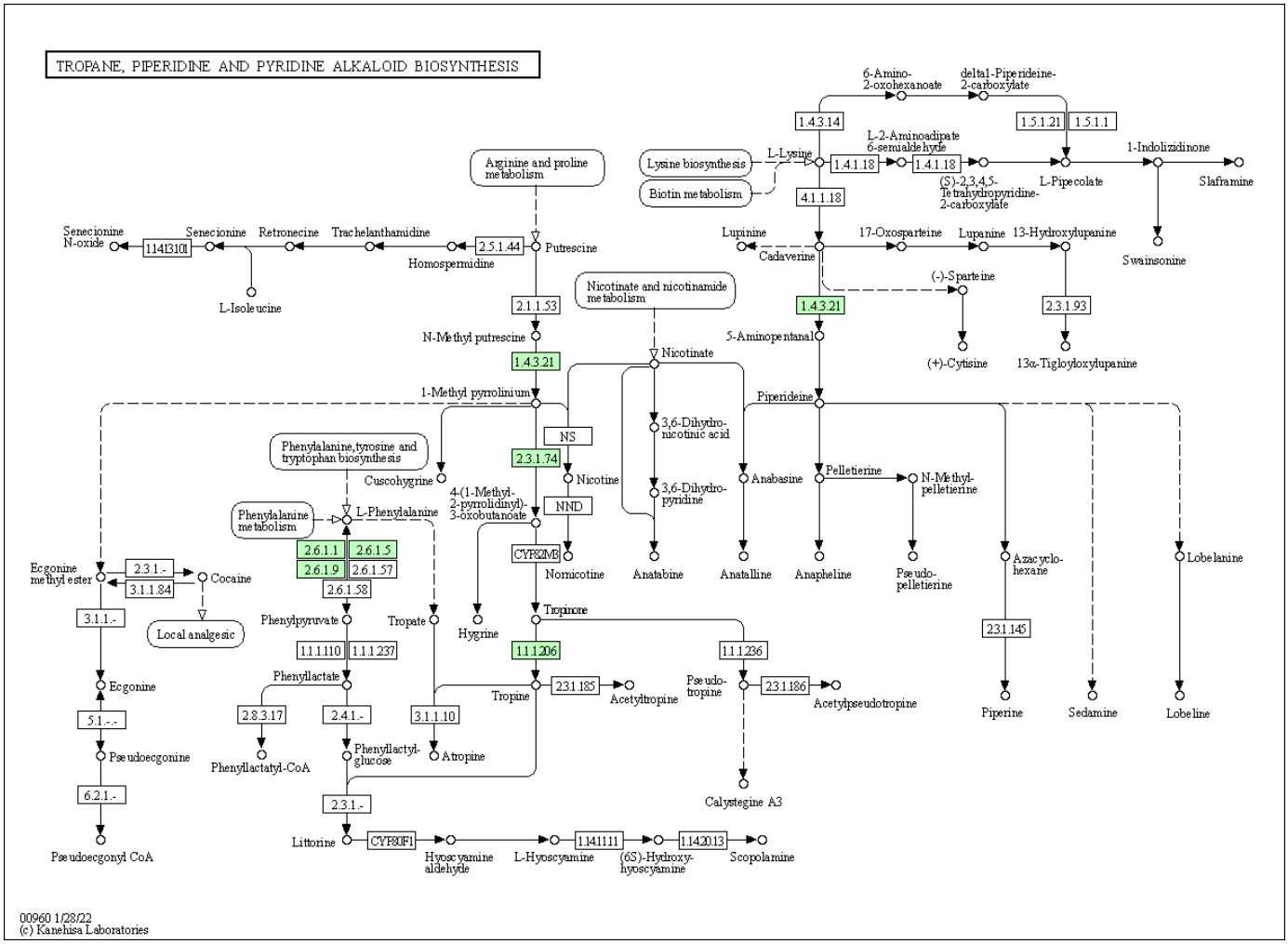


Figure 4: *Piper nigrum* tropane, piperidine and pyridine alkaloid biosynthesis pathway. KEGG ID 00960.

Differential expression analysis

In order to find the other enzymes present in the biosynthetic pathway a differential expression analysis was performed in the same way piperine synthase and other enzyme have been discovered.[16] RNAseq data from four different tissues of *P. nigrum* were used, this tissues were leaf, flowering spadix and black pepper fruits at two different stages: 20-30 days post anthesis and 40-60 days post anthesis; each tissue was represented by three biological replicates, having in total 12 samples. From now on the spadix of the panicle will be referred just as panicle. The mature fruit (40 days post anthesis) has the highest piperine and piperamide concentration from all the four tissues studied, considering this, it is expected that the genes over-expressed in mature fruit in comparison with the other tissues are related to piperine/piperidine biosynthesis.[16;18;17]

The RNAseq data were obtained from the SRA database with BioProject id: PRJEB38192, the sequencing technology used was Illumina HiSeq 2000 paired end sequencing.[13] The average number of reads per run was 35.46 Mbp, the quality of the reads was evaluated by using FastQC, a quality control tool for high throughput sequence data, and summarized with MultiQC.[2;10] The reads were chequed for adapter content and these were removed with Trimmomatic, the sequences were also quality trimmed using a sliding window of four nucleotides and a required average quality of 20, with a minimum length

of 125 bp.[4] After the raw data of the samples were cleaned, the average number of reads per run was 27.08 Mbp. The cleaned reads were mapped to the CDS sequences of the reference genome of *P. nigrum* by using Kallisto.[5]

With the count values obtained from the mapping, a differential expression analysis (DEA) was performed using the DESeq2 algorithm, a method for differential analysis of count data, using shrinkage estimation for dispersions and fold changes to improve stability and interpretability of estimates.[15] With this data a principal component analysis (PCA) was performed and can be seen in Figure 5. The principal component number one, which cointains 56% of the variance of the data, can differentiate between both the immature and mature fruit samples and the leaf and spadix tissues; the second component, with 36% of the variance, differentiates clearly between leaf and panicle tissues. Even though there is a distinction between both fruit stages, they can not be differentiated completely with this analysis, this indicates that there is a lot of similarity in the gene expression of both stages.

The differential expression analysis found 17071 differentially expressed genes across the selected tissues, this genes form a expression pattern for every tissue in the analysis. This genes were selected using a likelihood ratio test (LRT) to verify if they were significantly expressed in a different way across tissues. Genes with an absolute log2FoldChange value greater than one and an adjusted p-value lower than 0.05 in comparisons between mature fruit and every other tissue were considered as differentially expressed in mature fruit. These results can be seen in figures 6 and 7.

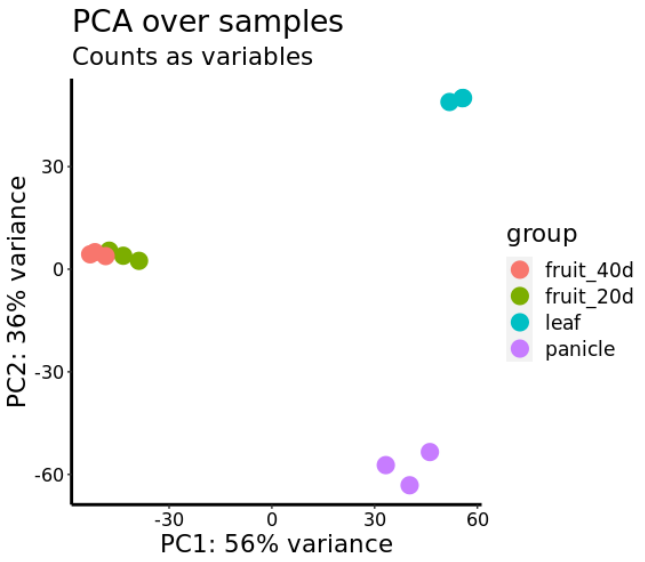


Figure 5: Principal component analysis of count values, RNAseq data of 12 samples of four different tissues of *P. nigrum*.

A search for the genes differentially expressed in mature fruit in comparison with every other tissue studied was performed, this results were verified by the LRT with adjusted p-value lower than 0.01 and an absolute log2FoldChange greater than 1. The results showed that 19,918 genes were differentially expressed in mature fruit in comparison with some tissue, but only 526 were in comparison to all tissues, from these genes, 10,972 were over-expressed in comparison with some tissue and only 223 genes were over-expressed in comparison to all tissues. From these genes, only those selected by hypothesis testing by LRT were conserved, 15,275 genes were differentially expressed in mature fruit in comparison with some tissue, but only 523 were in comparison to all tissues, this represents a reduction of three genes that were not considered as differentially expressed to all tissues after the hypothesis testing. From these genes, 9,022 were over-expressed in comparison with some tissue and 223 genes were over-expressed in comparison to all tissues, here we can see that there was no reduction in the number of genes interpreted as over-expressed to all other tissues, meaning a zero false discovery rate in this group (Figure 8).

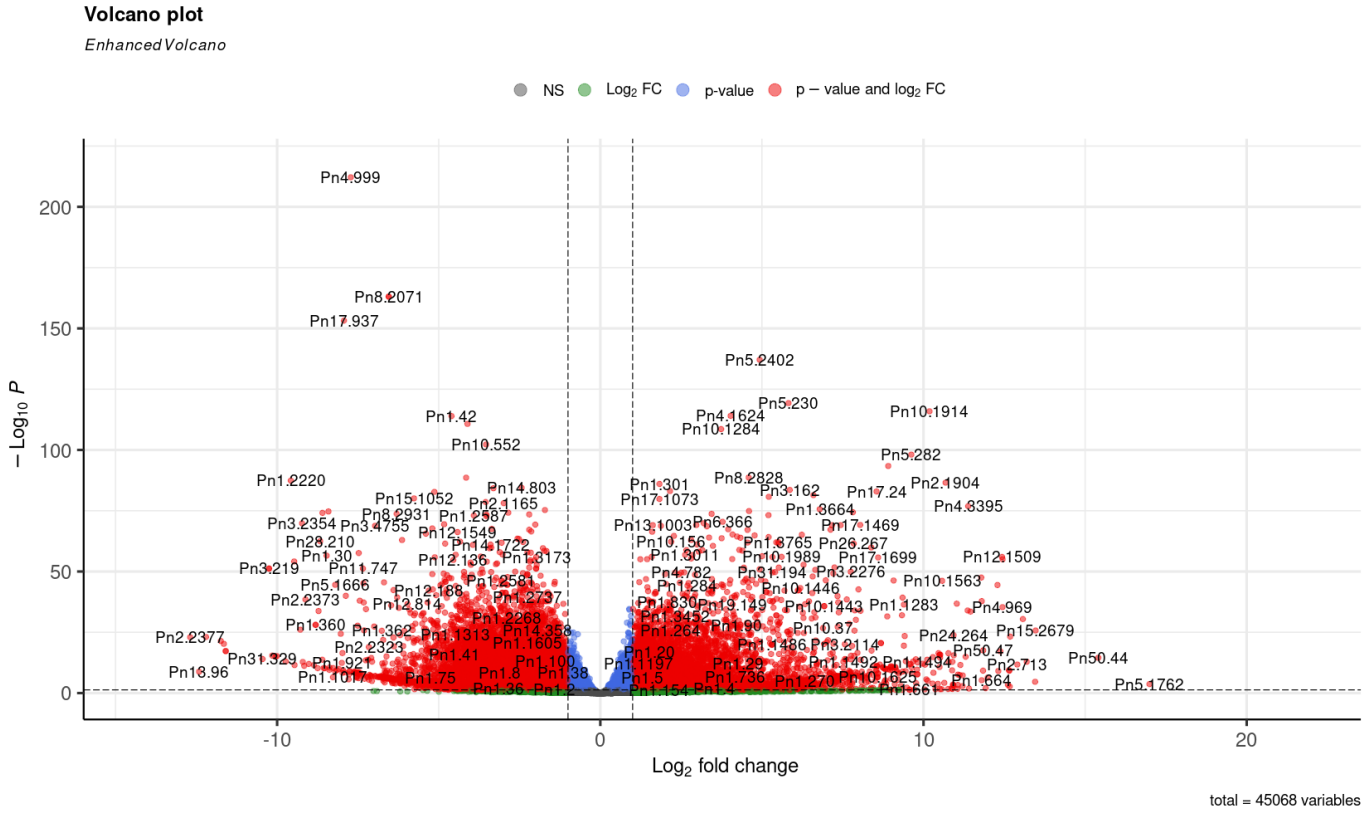


Figure 6: Volcano plot of differentially expressed genes in mature fruit.

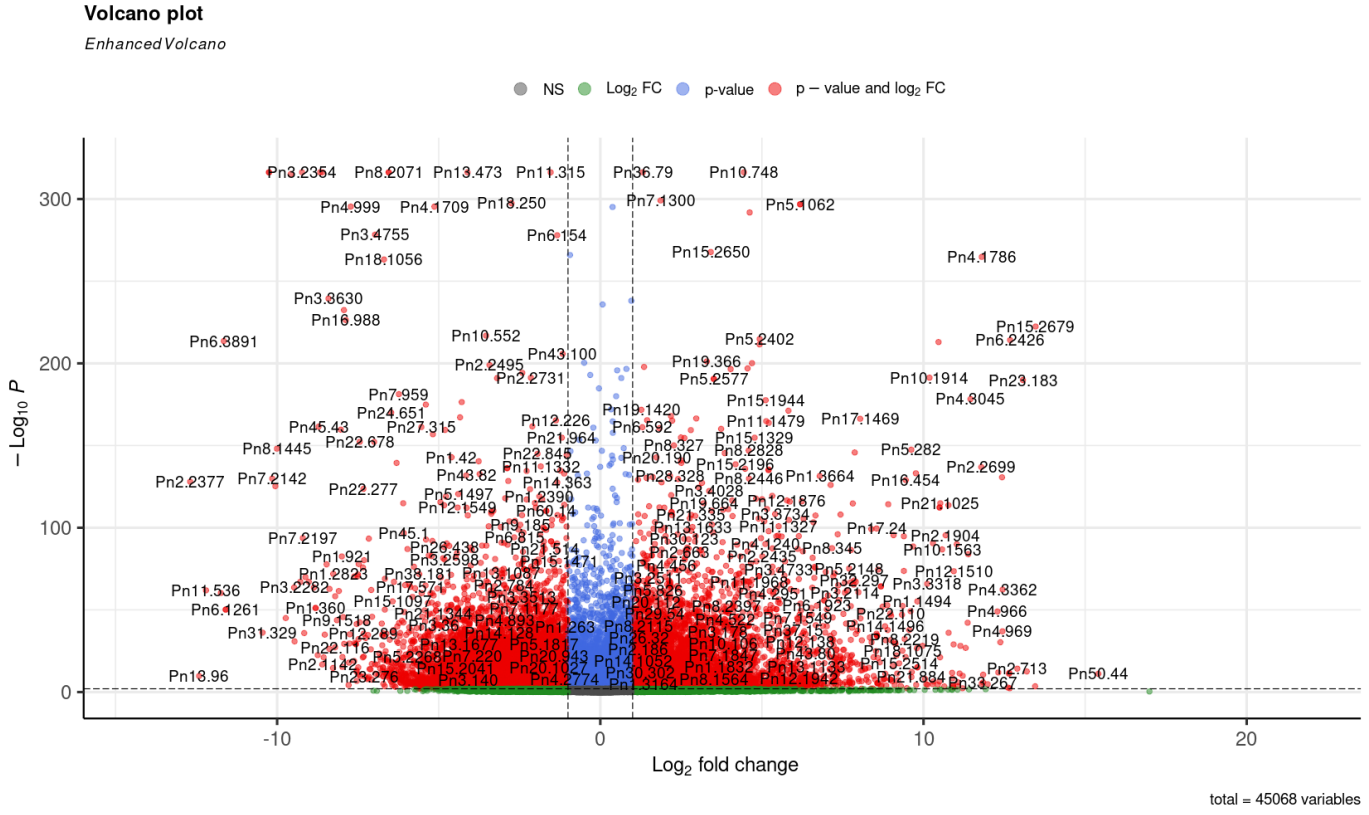


Figure 7: Volcano plot of differentially expressed genes in mature fruit, likelihood ratio test.

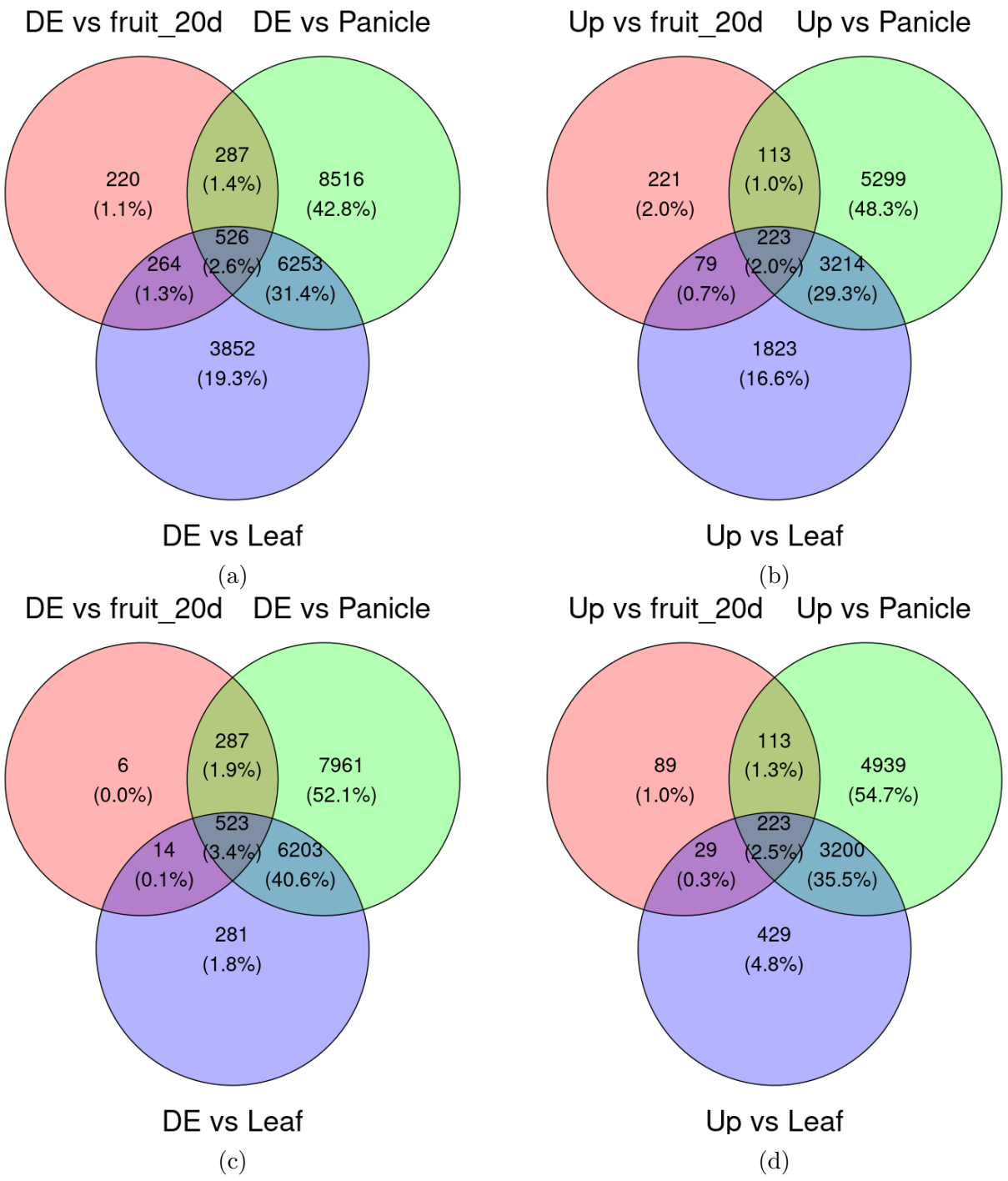


Figure 8: Venn diagram of differentially expressed genes in mature fruit. DE: Differentially expressed genes, Up: over-expressed genes. a-b pre hypothesis testing by LRT, c-d post hypothesis testing by LRT.

With the 523 genes found that are differentially expressed in mature fruit, a transcriptome profile was made (Figure 9). The heatmap shows the gene expression level for all the samples, the genes are clustered in eigth clusters found by kmeans, the number of adequate clusters was selected by the elbow method using the total within sum of squares of the cluster. The clusters 1, 2 and 7 are those of interest since there are the genes over-expressed in mature fruit in comparison to all tissues, probably involved in the biosynthesis of piperamides.

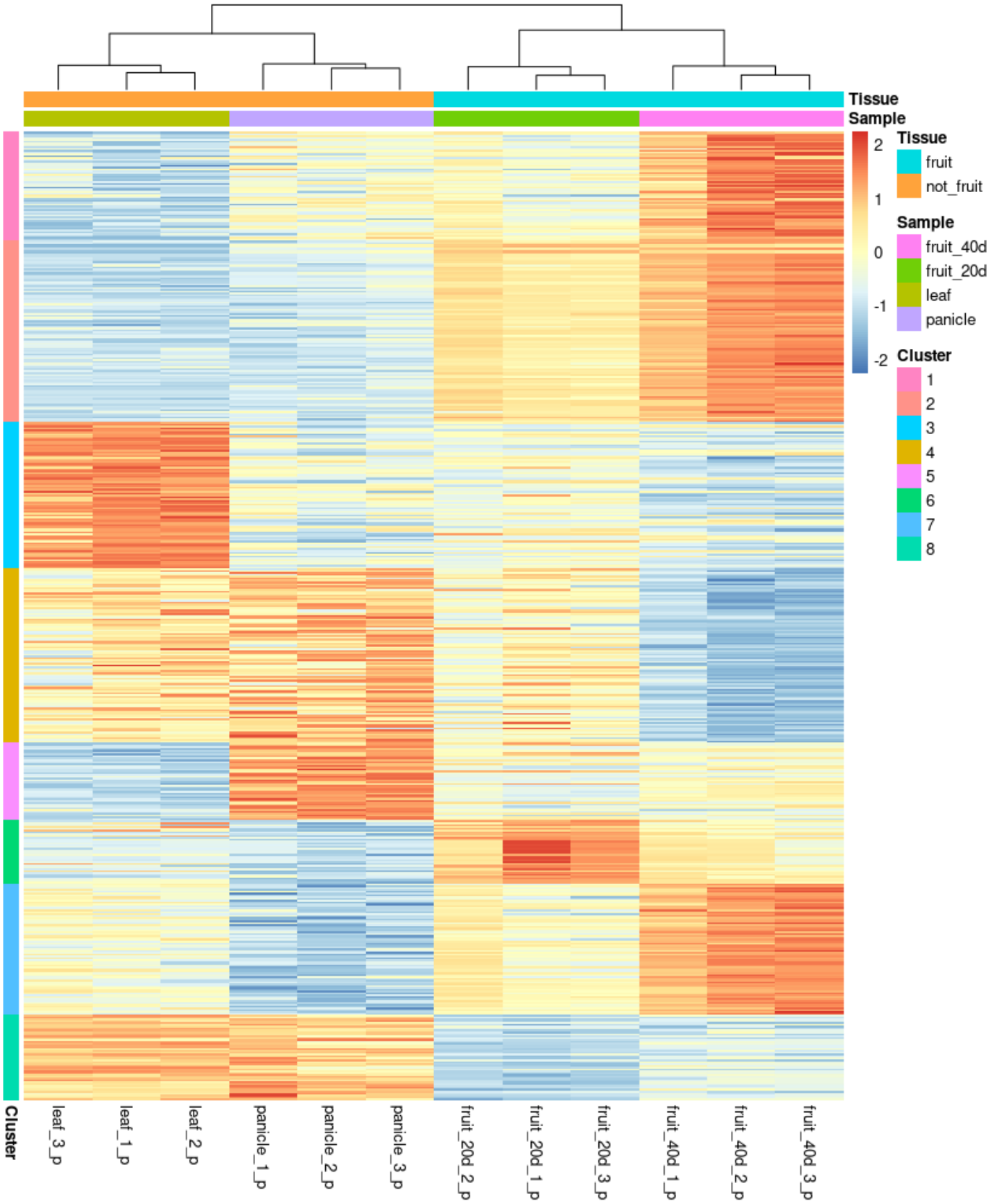


Figure 9: Transciptomic profile of differentially expressed genes in mature fruit.

The second gene more over-expressed in mature fruit is annotated as a terpenoid cyclases/protein prenyltransferases superfamily protein by the GeneFamilyClassifier tool. This gene has the highest log2FoldChange found (12.69) and an adjusted p-value of 9.74×10^{-129}. We propose that this enzyme could perform the step nine of the proposed biosynthetic pathway, in order to support this hypothesis a molecular docking methodology was performed, as with the other proposed enzymes. Even though the enzymes 5 and 6 are characterized for the synthesis of piperine, these enzyme cannot act on ligands with two less carbons like the proposed piperamide.[18;17] For enzyme five the gene Pn7.1626 was selected based on its similarity with the characterized protein (ID: QQS74306), with 76.908% identity and E-value ≈ 0. For enzyme number six the selected gene was Pn16.1198, the protein produced from this gene is annotated as a 4-coumarate:CoA ligase 2, which is the expected activity for the enzyme. It has a 75.632% identity with the know enzyme (ID: QGY72664), it is the sixth over-expressed gene in mature fruit with a log2FoldChange of 7.31 and a adjusted p-value of 5.89×10^{-81}. Its expression is higher in mature fruit than the known protein, which is the 189th over-expressed gene with a log2FoldChange of 1.84 and a adjusted p-value of 9.78×10^{-7}.

The piperine synthase (ID: QUS53100) cannot act on an acyl donor two carbons shorter than it normal substrate, piperoyl-CoA, but a similar enzyme called piperamide synthase (ID: QUS53101) can act on a wider variety of acyl donors, included the 3,4-methylenedioxy cinnamoyl CoA, which is the compound involved in our proposed biosynthetic pathway. The piperine synthase corresponds to the gene Pn12.1813, this gene is annotated as a HXXXD-type acyl-transferase family protein, which is the reaction type performed by the piperine synthase. This gene is over-expressed in the mature fruit in position 95, with a log2FoldChange of 4.77 and a adjusted p-value of 1.19×10^{-20}. The piperamide synthase corresponds to the gene Pn6.2477, this gene is also annotated as a HXXXD-type acyl-transferase family protein. This gene is over-expressed in the mature fruit in position 56, with a log2FoldChange of 5.44 and a adjusted p-value of 1.89×10^{-34}.

The three remaining enzymes were selected from the literature. Three of the genes selected were found over-expressed in the mature fruit, where the piperine and piperamides are present in the highest concentration in the plant. This expression pattern can be thought of as a temporal expression program for the biosynthetic pathway, in this way the different transcriptional units were designed, and the expression pattern of the construct works in a similar way to the temporal model found in the expression pattern of the pathway in *Piper nigrum*.

Appendix

Table 1: Bioinformatic tools used in the project.

Tool/program	Usage
AlphaFold2	Structure modelling algorithm using an artificial intelligence based methodology
Autodock Vina	Protein-ligand docking
AutoDockTools	Protein-ligand docking
Bioconductor	Open software development for computational biology and bioinformatics.
Biopython	Python Tools for Computational Molecular Biology
blast +	Suite of command-line tools to run BLAST (Basic Local Alignment Seacrh Tool)
CASTp3	Prediction of protein pockets/cavities
CIPRES Science Gateway	Public resource for inference of large phylogenetic trees.
Circos	Software package for visualizing data and information designed for visualizing genomic data
ColabFold	Collaboratory space that allows users to perform three-dimensional structure prediction using AlphaFold2
DeSeq2	Differential gene expression analysis based on the negative binomial distribution
DEXSeq	Inference of differential exon usage in RNA-Seq
Expasy	One of the main bioinformatics resources for proteomics.
FastQC	A quality control tool for high throughput sequence data.
FigTree	Phylogenetic trees visualization
Galaxy Platform	Accessible, reproducible and collaborative biomedical analyses.
Kallisto	Program for quantifying abundances of transcripts from RNA-Seq data
KofamKOALA	Kegg ortholog assignment based on profile hmm and adaptive score threshold.
Ligplot +	Program for automatically plotting protein-ligand interactions
MEGA 11	Molecular Evolutionary Genetics Analysis
MrBayes	Bayesian Inference of Phylogeny
MultiQC	Aggregate results from bioinformatics analyses across many samples into a single report
muscle	Multiple Sequence Comparison by Log- Expectation
NCBI conserved domain search interface	Conserved domain search of proteins

open babel	Format conversion of molecule structure files
P2rank	Machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure.
PlantTribes	Gene and gene family resource for comparative genomics in plants
Pymol	Molecule visualization
rdkit	Ligand energy minimization
Star	Ultrafast universal rna-seq aligner.
SWISS-MODEL	Three-dimensional protein structure prediction (homology based)
SwissADME	Web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules
Symbiology	Used to simulate enzymatic, promoter and various chemical reactions according to a mathematical model
Tracer	Software package for visualising and analysing the MCMC trace files generated through Bayesian phylogenetic inference.
Trimmomatic	A flexible read trimming tool for Illumina NGS data
UCSF ChimeraX	Molecule visualization an protein energy minimization

References

[1] Abraham, M. H., Chadha, H. S., and Mitchell, R. C. (1995). The factors that influence skin penetration of solutes. *Journal of pharmacy and pharmacology*, 47(1):8–16.

[2] Andrews, S. et al. (2010). Fastqc: a quality control tool for high throughput sequence data.

[3] Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). Kofamkoala: Kegg ortholog assignment based on profile hmm and adaptive score threshold. *Bioinformatics*, 36(7):2251–2252.

[4] Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

[5] Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic rna-seq quantification. *Nature biotechnology*, 34(5):525–527.

[6] Community, T. G. (2022). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50(W1):W345.

[7] Daina, A., Michielin, O., and Zoete, V. (2017). Swissadme: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7(1):1–13.

[8] De Groote, H., Kimenju, S. C., Munyua, B., Palmas, S., Kassie, M., and Bruce, A. (2020). Spread and impact of fall armyworm (spodoptera frugiperda je smith) in maize production areas of kenya. *Agriculture, ecosystems & environment*, 292:106804.

[9] EPPO (2022). Eppo global database. *Available online: https://gd.eppo.int/taxon/LAPHFR/distribution/MX*.

[10] Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048.

[11] Haq, I.-U., Imran, M., Nadeem, M., Tufail, T., Gondal, T. A., and Mubarak, M. S. (2021). Piperine: A review of its biological effects. *Phytotherapy Research*, 35(2):680–700.

[12] Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., Wu, H., Qin, X., Yan, L., Tan, L., et al. (2019). The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nature communications*, 10(1):1–11.

[13] Leinonen, R., Sugawara, H., Shumway, M., and Collaboration, I. N. S. D. (2010). The sequence read archive. *Nucleic acids research*, 39(suppl_1):D19–D21.

[14] Lian, G., Chen, L., and Han, L. (2008). An evaluation of mathematical models for predicting skin permeability. *Journal of pharmaceutical sciences*, 97(1):584–598.

[15] Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12):1–21.

[16] Schnabel, A., Athmer, B., Manke, K., Schumacher, F., Cotinguiba, F., and Vogt, T. (2021a). Identification and characterization of piperine synthase from black pepper, piper nigrum l. *Communications biology*, 4(1):1–10.

[17] Schnabel, A., Cotinguiba, F., Athmer, B., and Vogt, T. (2021b). Piper nigrum cyp719a37 catalyzes the decisive methylenedioxy bridge formation in piperine biosynthesis. *Plants*, 10(1):128.

[18] Schnabel, A., Cotinguiba, F., Athmer, B., Yang, C., Westermann, B., Schaks, A., Porzel, A., Brandt, W., Schumacher, F., and Vogt, T. (2020). A piperic acid coa ligase produces a putative precursor of piperine, the pungent principle from black pepper fruits. *The Plant Journal*, 102(3):569–581.

[19] Tavares, W., Cruz, I., Petacci, F., Freitas, S., Serrão, J., Zanuncio, J., et al. (2011). Insecticide activity of piperine: Toxicity to eggs of spodoptera frugiperda (lepidoptera: Noctuidae) and diatraea saccharalis (lepidoptera: Pyralidae) and phytotoxicity on several vegetables. *J. Med. Plants Res*, 5(21):5301–5306.

[20] Tiwari, A., Mahadik, K. R., and Gabhe, S. Y. (2020). Piperine: A comprehensive review of methods of isolation, purification, and biological properties. *Medicine in Drug Discovery*, 7:100027.

[21] Wall, P. K., Leebens-Mack, J., Müller, K. F., Field, D., Altman, N. S., and Depamphilis, C. W. (2007). Planttribes: a gene and gene family resource for comparative genomics in plants. *Nucleic acids research*, 36(suppl_1):D970–D976.

[22] Yooboon, T., Pengsook, A., Ratwatthananon, A., Pluempanupat, W., and Bullangpoti, V. (2019). A plant-based extract mixture for controlling spodoptera litura (lepidoptera: Noctuidae). *Chemical and Biological Technologies in Agriculture*, 6(1):1–10.

Contents

List of Figures