

Analysis of ANS structures and molecular evolution in Rosaceae

Chang Liu^{1*}

¹ College of Horticulture, Sichuan Agricultural University, Chengdu, Sichuan, 611130, China

Abstract. Anthocyanidin Synthase (ANS) is a key rate-limiting enzyme involves in the biosynthesis of anthocyanin. Notably, anthocyanin contents of Rosaceae have a significant impact on its economic value as major economic crops. Thus, the role of ANS in the regulation of anthocyanin contents in the Rosaceae family is worthy of investigation. Since the characterizations and expression pattern of ANS in the Rosaceae family are largely unknown so far, a systematic analysis is required to extend current understanding. Using the bioinformatics methods, we performed comprehensive bioinformatic analysis for the ANS amino acid sequences of 11 major Rosaceae species, which shared a high degree of similarity and identity, and also clustered closely in the phylogenetic analysis. Through homology modeling, 11 Rosaceae species ANS were divided into two groups: one group with similar structures to AtANS and the other with similar structures to AtACO2. Further protein-ligand docking showed that most ANS might be involved in the anthocyanin synthesis. This study found that ANS from 11 Rosaceae species shared highly similar characteristics, protein structures, and binding characteristics, helping increase the understanding of ANS evolution in Rosaceae fruits and providing supports for studies in the regulation of anthocyanin biosynthesis of Rosaceae plants.

1. Introduction

The Rosaceae family (Rosaceae) has a large variety and number of plants, containing about 124 genera and more than 3,300 species, and is widely distributed throughout the world. Common rose plants mainly consist of apple (*Malus domestica*), pear (*Pyrus communis*), sweet cherry (*Prunus avium*), raspberry (*Rubus idaeus*), Chinese pear (*Pyrus Pyrifolia*), peach (*Prunus Persica*), and strawberry (*Fragaria × ananassa*) with high economic values.

As important economic crops, skin colour is one of the most important appearance qualities of Rosaceae fruits. Anthocyanin is the main pigments that determines the colour of fruit skins, and the type, amount, distribution of its accumulation is closely related to the fruit colour [1] [2]. Additionally, anthocyanin can attract pollinators, prevent ultraviolet light from burning plants, and resist pests and pathogens, which is essential to improve the yield of crops. Recently, there has been an increasing requirement in fruits' health care functions among consumers. Anthocyanin, an important antioxidant phenolic substance, has the activities of preventing cardiovascular disease, anti-inflammatory, and anticancer [3]. Thus, anthocyanin contents in Rosaceae plants not only affect the skin colour which is an important index for evaluation of fruit quality and commodity, but also directly affect its nutritional value and market competitiveness. Therefore, research on the biosynthesis of anthocyanin in Rosaceae has attracted much attention.

Anthocyanin, a natural water-soluble pigment composed of an anthocyanidin backbone with sugar and acyl conjugates [4], has typical flavonoid characteristics and is a phenolic compound. As with other polyphenolic substances, anthocyanin occurs naturally as glycosides of polyhydroxy and polymethoxy derivatives of 2-phenylbenzopyrylium salts. There are about 20 species of anthocyanidins that have been discovered in the world, of which six are the most common, termed as pelargonidin, cyanidin, delphinidin, peonidin, petunidin, malvidin [5]. Anthocyanin biosynthesis belongs to the branching pathway of flavonoid compounds, which is synthesized through shikimic pathway [6]. The biosynthetic pathway of anthocyanin is initiated from 4-coumaroyl-CoA and malonyl-CoA, which converted to naringenin through Chalcone Synthase (CHS) and Chalcone Isomerase (CHI). Naringenin is generated to dihydrokaempferol by Flavanone 3-hydroxylase (F3H), which is further hydroxylated by flavonoid 3'-hydroxylase (F3'H) or flavonoid 3',5'-hydroxylase (F3'5'H) to two other dihydroflavonols. Catalyzed by Dihydroflavonol 4-Reductase (DFR), these dihydroflavonols are converted to colorless leucoanthocyanidins. Subsequently, colored anthocyanidins are yield by Anthocyanidin Synthase (ANS). Finally, sugar molecules can be attached to anthocyanidins by Flavonoid 3-O-Glucosyltransferase (UGT) to generate anthocyanin.

Anthocyanidin Synthase (ANS) belongs to the 2-oxoglutarate Fe(II)-dependent dioxygenase (2OGD) family, which contains a C-terminal 2-

* Corresponding author's e-mail: 202001709@stu.sicau.edu.cn

oxoglutarate/Fe(II)-dependent oxygenase domain (PF03171) and an N-terminal non-heme dioxygenase domain (PF14226). Additionally, ANS is a key rate-limiting enzyme at the end of anthocyanin biosynthetic pathway, which catalyzes anthocyanin formation from colorless substrates through Fe(II) and 2-oxoglutarate. Recently, the relationship between the expression level of *ANS* and anthocyanin accumulation have been studied in many plant species of Rosaceae. In strawberry (*Fragaria x ananassa*), FaANS might promote the biosynthesis of flavonols[7]. In pear (*Pyrus communis* L.), anthocyanin accumulation is positively correlated with the expression of five anthocyanin biosynthetic genes (*PcPAL*, *PcF3H*, *PcDFR*, *PcANS*, and *PcUGT*) during pear ripening [8]. Moreover, *ANS* is down-regulated in less colored specie ‘Zaobaimi’ but upregulated in deeper colored ‘Yunhong-1’ pears [9]. The reduced anthocyanin contents of yellow raspberry (*Rubus idaeus* L.) fruits may be due to the absence of most important catalytic residues in the ANS of yellow raspberry compared to other known ANS of Rosaceae [10]. In sweet cherries (*Prunus avium* L.), the expression of *ANS* in UV-C irradiated samples was significantly enhanced, the total phenolic, flavonoid, and anthocyanin contents were considerably elevated consistently [11] [12]. However, most studies on *ANS* gene focus on *Arabidopsis thaliana*, the regulatory role of anthocyanin biosynthesis in Rosaceae fruit is still insufficient.

In this study, the ANS amino acid sequences of 11 species of Rosaceae [pear (*Pyrus communis*), sweet cherry (*Prunus avium*), raspberry (*Rubus idaeus*), Chinese pear (*Pyrus pyrifolia*), peach (*Prunus persica*), strawberry (*Fragaria x ananassa*), sago (*Malus sylvestris*), loquat (*Eriobotrya japonica*), European plum (*Prunus domestica*), and almond (*Prunus dulcis*)] were obtained. With highly similar characteristics, protein structures, and conserved regions, the understanding of ANS evolution in Rosaceae fruits has increased, and through homology modeling and further protein-ligand docking analysis, the role of *ANS* in the regulation of anthocyanin biosynthesis of Rosaceous plants also become clear, providing supports for studies of both *ANS* and Rosaceae plants.

2. Method and Result

2.1 Identification of ANS sequences of major rosaceous crop species

To obtain the unpublished ANS information of 20 major rosaceous crop species [Maloideae (apple, pear, quince, loquat and medlar), Amygdaloideae (plum, cherry, almond, apricot, peach and damson), Rosoideae (raspberry, strawberry and rose)], we used BLAST

(https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) approach to screen the ANS sequences based on the amino acid sequence of apple ANS (*Malus domestica*, MdANS). By excluding incomplete sequences of several species (cherry plum, Japanese plum, apricot, Chinese pear, sour cherry, rose), ANS sequences from totally 10 species were selected for further analysis (Figure 1A). The lengths of the predicted ANS protein sequences were highly similar among most of the species, ranging from 314 to 381 amino acids, with the exception of RiANS (413 amino acids), which was about 15% longer than that of MdANS (Figure 1A).

For protein sequences of these predicted rosaceous ANS, the percentage of similarity and identity to MdANS varied from 100% to 81%, and 99.72% to 31.76%, respectively (Figure 1A). Moreover, the ANS amino acid sequences of *Pyrus communis*, *Prunus avium*, *Rubus idaeus*, *Pyrus pyrifolia*, *Prunus persica*, *Fragaria x ananassa*, *Prunus dulcis*, and *Prunus domestica* predicted by ESript3 (<https://esript.ibcp.fr/ESript/ESript/index.php>) were highly identity to that of MdANS, indicating that the predicted rosaceous ANS might be involved in anthocyanin biosynthetic pathway and assume a role of converting colorless leucoanthocyanidins into colored anthocyanidins. Though the amino acid sequences of predicted MsANS and EjANS differed from that of MdANS, several amino acid sites (e.g. substrate-binding sites: 233Ala, 306Ala; metal binding sites: 234His, 288Lys) still remained the same. The functions of these amino acid sites as well as whether these sites function as key enzymic sites still need further studies. Additionally, all predicted ANS protein sequences contain conserved regions: DIOX_N domain (PF14226, 51-163 aa) and 2OG-Fe(II)_Oxy domain (PF03171, 216-311 aa), suggesting that they all belong to the 2OGD family, consistent with the reported functional ANS (e.g. MdANS and AtANS). Collectively, these data indicate that the predicted ANS of 10 major rosaceous crop species share a high genetic homology.

2.2 Phylogenetic tree analysis

To solve the evolutionary relationship of ANS in these rosaceous crop species, we used MEGA7 to conduct a phylogenetic analysis. All ANS protein sequences from 11 major rosaceous species were related with each other. Furthermore, the ANS clustered according to their taxonomic relationships in the Rosaceae. The species from Maloideae (Apple, pear, loquat, crab, nashi) clustered together closely, while the species from Rosoideae Focke (strawberry, raspberry) clustered more distantly (Figure 1).

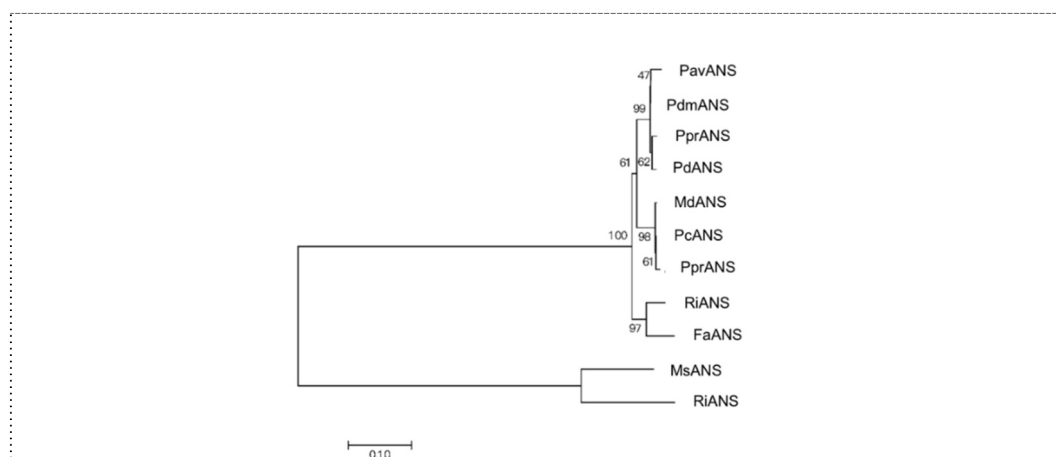


Figure 1. Phylogenetic relationships among rosaceous ANS proteins. Phylogeny of ANS from 11 major rosaceous species. Sequences were aligned using Clustal W and the phylogenetic and molecular evolutionary analysis was conducted using MEGA 7.

2.3 Homologous modeling of ANS and substrate docking

To understand whether partial differences in ANS sequences lead to differences in the 3D structure of proteins, we first performed ANS homology modeling of multiple species (*Pyrus communis*, *Prunus avium*, *Rubus idaeus*, *Pyrus pyrifolia*, *Prunus persica*, *Fragaria × ananassa*, *Malus sylvestris*, *Eriobotrya japonica*, *Prunus*

domestica, *Prunus dulcis*) using SWISS-MODEL with AtANS as a template. Figure 2 showed that predicted ANS structures of 8 major Rosaceous crop species: *Malus domestica* (B), *Pyrus communis* (C), *Prunus avium* (D), *Rubus idaeus* (E), *Pyrus pyrifolia* (F), *Prunus persica* (G), *Fragaria × ananassa* (H), *Prunus dulcis* (I) are highly similar to that of AtANS (A), whereas the predicted ANS structures of *Malus sylvestris* (K) and *Rhaphiolepis bibas* (L) are highly similar to that of 1-aminocyclopropane-1-carboxylate oxidase 2 (AtACO2) (J), a 2OGD involved in the biosynthesis of ethylene, but not anthocyanin.

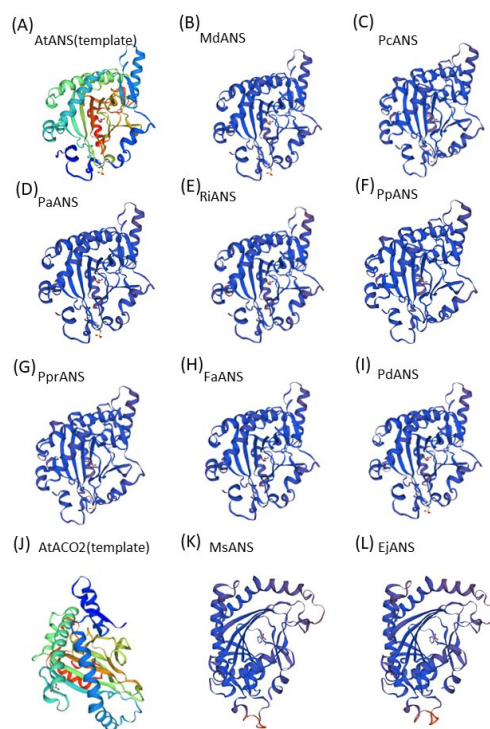
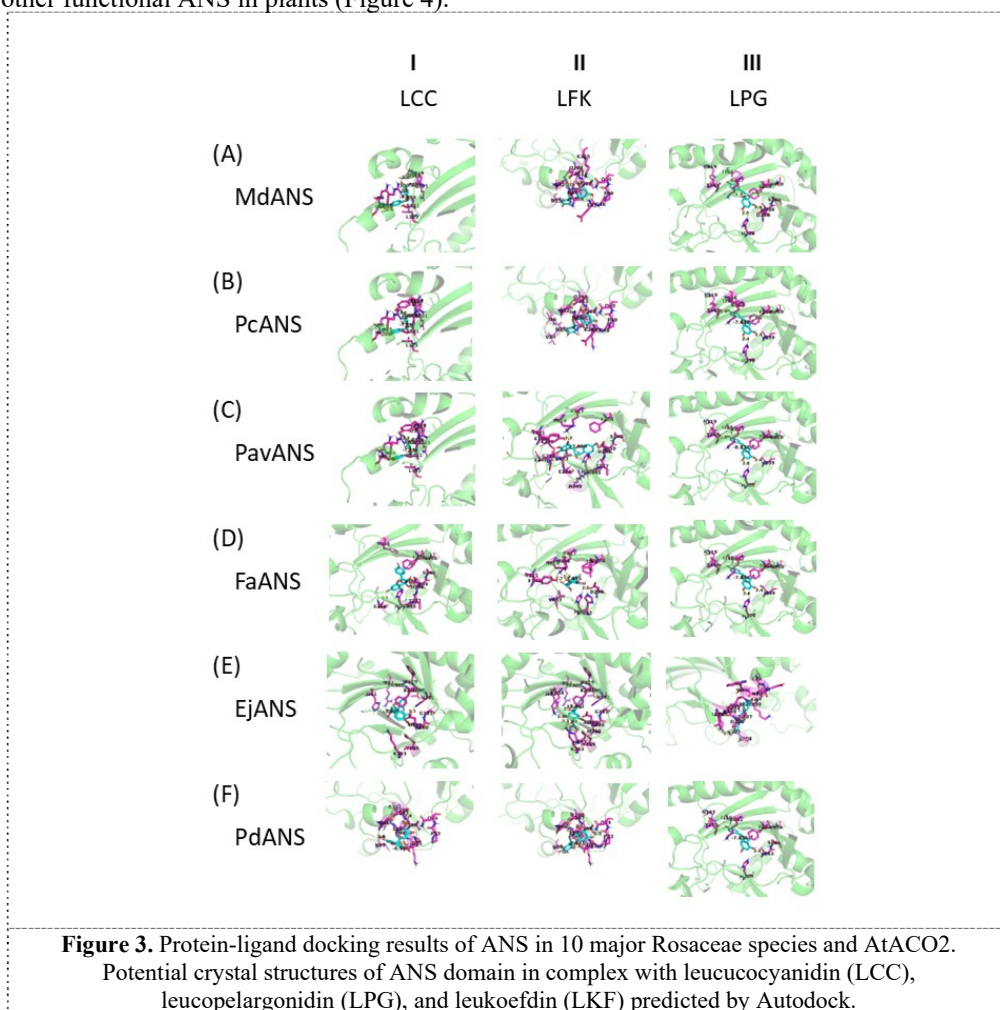
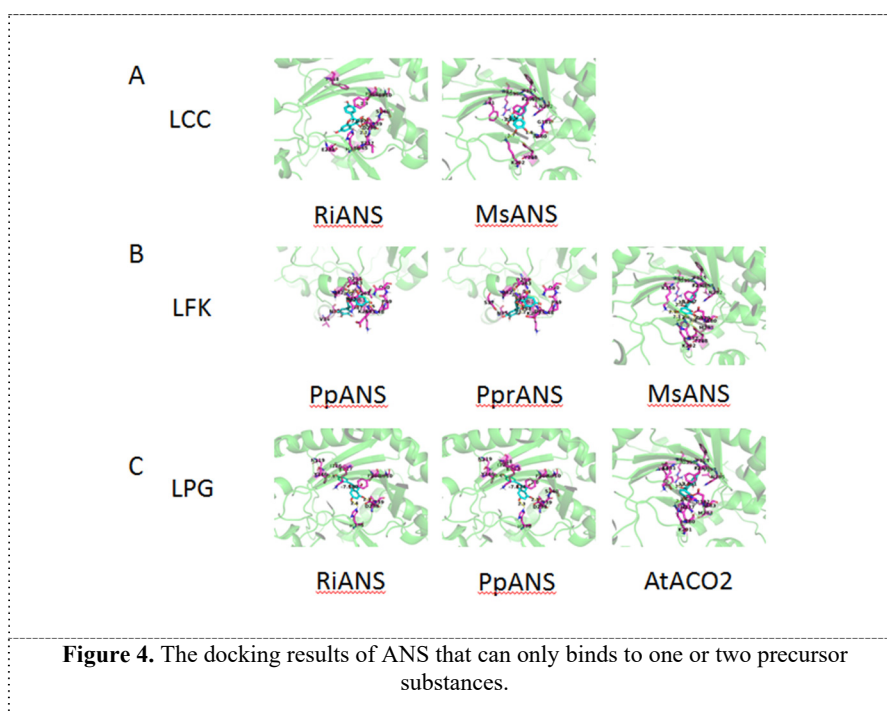


Figure 2. Predicted ANS 3D structures of 10 major Rosaceae species. Homologous modeling of rosaceous ANS proteins performed by SWISS-MODEL with AtANS (A) and AtACO2 (H) as templates.

To further detect whether the predicted ANS proteins can bind to substrates, protein-ligand docking analysis by Autodock software was used. The result exhibited that ANS of *Malus domestica*, *Pyrus communis*, *Prunus avium*, *Fragaria x ananassa*, *Prunus dulcis* could bind to leucocyanidin (LCC), leucopelargonidin (LPG), and leukoefdin (LKF), just like that of AtANS (Figure 3). While ANS of *Rubus idaeus*, *Pyrus pyrifolia*, *Malus sylvestris* could not bind to LCC, LPG, and LKF, respectively (Figure 4). Also, ANS of *Prunus persica* could only bind to LKF, which might be due to the existence of other functional ANS in plants (Figure 4).

MdANS, PcANS, PaANS showed highly similar binding sites when the ANS combined with LCC (Figure 3AI, BI, and CI). The binding free energy is -7.907, -7.882, -8.312, respectively. This binding result might be due to their highly similar ANS structures. Meanwhile, RiANS and FaANS showed similar binding sites to LCC, consistent with the phylogenetic analysis (Figure 4). The PprANS did not bind to LCC, which still needs more scientific research focus.





When binding to LKF, the results can separate the 10 species into three groups. Group A (MsANS, PcANS, PpANS, PprANS), group B (PaANS, FaANS) and group C (MsANS, RiANS). There were significant differences in the binding results among the groups, and the docking results between ANS and LKF were similar in each group. It suggests that LKF docking with ANS may be more complex and elaborate, and the result is more closely related to the evolutionary relationship between various species of Rosaceae.

When combined with LPG, these ten species of ANS showed similar binding sites. These results indicate that in Rosaceae, LPG binds to ANS in a specific protein site which plays an important role in anthocyanin synthesis.

In protein 3D structure prediction, we found that the structures of MsANS and RiANS are more similar to the structure of AtACO2. To further analyze the relationship between AtACO2 and these two ANS, the docking of AtACO2 and three anthocyanins has been conducted. Interestingly, when binding to LFK, MsANS, and RiANS showed similar binding sites to AtACO2-LPG result.

3. Discussion

The study of 2OGD has emerged as one of the focuses in plant gene function research. ANS catalyzes the formation of anthocyanins from colorless substrates as a key rate-limiting enzyme in anthocyanin biosynthesis. ANS of ten major crops from the rosaceae family were chosen [Pear (*Pyrus communis*), Sweet cherry (*Prunus avium*), raspberry (*Rubus idaeus*), Chinese pear (*Pyrus Pyrifolia*), Peach (*Prunus persica*), Strawberry (*Fragaria × Ananassa*), Sago (*Malus Sylvestris*), loquat (*Eriobotrya japonica*), European plum (*Prunus domestica*), almond (*Prunus*

dulcis)] to analyze the gene structures, conserved structures, phylogenetic tree, and 3D structures. Our findings revealed that ANS from ten major Rosaceae crops were highly conserved, encoding 2 typical conserved domains: DIOX_N [N-terminal non-heme dioxygenase domain (PF14226)] and 2OG-Fe(II)-oxy [C-terminal 2-oxyglutarate/Fe(II)-dependent oxygenase domain (PF03171)], and share high homology. It is speculated that the biological functions of PcANS, PaANS, RiANS, PprANS, PpANS, FaANS, MsANS, EjANS, PdmANS, PdANS are similar to that in MdANS. Phylogenetic analysis revealed that plants from the same subfamily were clustered together, implying that they were genetically related. The phylogenetic evolutionary tree has many levels and branches step by step, indicating that ANS protein evolution is both conserved and specific.

Anthocyanin synthesis is influenced by the 2OGD ANS. Plants' anthocyanin content can be increased or decreased by controlling the expression levels of the ANS gene. As a result, plant varieties with different anthocyanin content can be cultivated to meet market demand, increasing the economic value of crops. The findings of this study analyzed the evolutionary relationship of the ANS genes in the Rosaceae family, laying a theoretical foundation for breeding anthocyanin-rich rosaceae crops.

4. Method

4.1 Collection of the ANS amino acid sequences and the related information of 11 major rosaceous crop species

11 major Rosaceous crop species [pear (*Pyrus communis*), sweet cherry (*Prunus avium*), raspberry (*Rubus idaeus*), Chinese pear (*Pyrus pyrifolia*), peach (*Prunus persica*), strawberry (*Fragaria × ananassa*), sago (*Malus*

sylvestris), loquat (*Eriobotrya japonica*), European plum (*Prunus domestica*), and almond (*Prunus dulcis*)] were chosen and UNIPROT (<https://www.uniprot.org>) database was used to find the ANS amino acid sequences in FASTA format. Unregistered ANS amino acid sequences of crab apple (*Malus sylvestris*), loquat (*Eriobotrya japonica*), European plum (*Prunus domestica*), almond (*Prunus dulcis*), strawberry (*Fragaria × ananassa*) were found in UNIPROT, then the sequences of these five species were obtained through the comparison with MdANS by BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).

The genebank numbers, protein lengths, and DIOX_N and 2-OG-Fe(II) lengths of 11 Rosaceous crop species were recruited from NCBI (<https://ncbi.nlm.nih.gov>). The similarity and identity of ANS of 10 Rosaceous crop species to MdANS were analyzed by BLASTP.

4.2 Alignment of amino acid sequences in multiple species of Rosaceae family

ClustalW (<https://www.genome.jp/tools-bin/clustalw>) was used to execute multiple alignment of the amino acid sequences of the major rosaceous crop species (*Pyrus communis*, *Prunus avium*, *Rubus idaeus*, *Pyrus pyrifolia*, *Prunus persica*, *Fragaria × ananassa*, *Prunus dulcis*, *Prunus domestica*, *Malus sylvestris*, *Eriobotrya japonica*) in slow/accurate mode.

Esript3.0 (<https://esript.ibcp.fr/ESript/ESript/index.php>) was used to get visualized multiple alignment diagram.

4.3 Construction of the Phylogenetic tree

ANS proteins of 10 major rosaceous crop species containing 2OG-Fe(II) oxygenase superfamily (PF03171) and non-haem dioxygenase in morphine synthesis N-terminal (PF14226) were selected by alignment with MdANS using BLASTP. For each gene, the longest protein sequence was selected for phylogenetic analysis at phylogeny. Phylogenetic trees were constructed by MEGA 7 in Neighbor-joining method. Parameter for the NJ tree were set as poisson model and complete detection with the bootstrap values as 500.

4.4 Homology modeling of ANSs and protein-ligand docking

The structures of AtANS and AtACO2 have been resolved and are in the PDB database under the SMTL ID: 1gp4.1.A and 5gj9.1, respectively. The structures of AtANS and AtACO2 were used as templates to access the SWISS-MODEL website (<http://www.expasy.ch/swissmod/swiss-model.html>), the protein sequences of predicted ANS of *Pyrus communis*, *Prunus avium*, *Rubus idaeus*, *Pyrus pyrifolia*, *Prunus persica*, *Fragaria × ananassa*, *Malus sylvestris*, *Eriobotrya japonica*, *Prunus domestica*, and *Prunus dulcis* were entered in the "automatic modelling" mode.

Then the protein model files returned from the calculation were downloaded.

The predicted structure files from SWISS-MODEL and substrates: Leucocyanidin (LCC), Leucopelargonidin (LPG), leukoefdin(LKF) from PubChem (<https://pubchem.ncbi.nlm.nih.gov>) were downloaded and docked using Autodock software.

Autodock 4.0 was used to dock the three substances LCC, LPG, LKF and ten ANS from Rosaceae family: *Pyrus communis*, *Prunus avium*, *Rubus idaeus*, *Pyrus pyrifolia*, *Prunus persica*, *Fragaria × ananassa*, *Malus sylvestris*, *Eriobotrya japonica*, *Prunus dulcis*. The size of the docking box was 60×60×60. The docking parameters were assigned following the default values.

References

1. Sass-Kiss, A., Kiss, J., Milotay, P., Kerek, M.M. & Toth-Markus, M. (2005) Differences in anthocyanin and carotenoid content of fruits and vegetables. Food Research International 38: 1023-1029.
2. Lightbourn, G.J. *et al.* (2008) Effects of anthocyanin and carotenoid combinations on foliage and immature fruit color of *Capsicum annuum* L. Journal of Heredity 99: 105-111.
3. He, J.A. & Giusti, M.M. (2010) Anthocyanins: Natural Colorants with Health-Promoting Properties. Annual Review of Food Science and Technology, Vol 1: 163-187.
4. Stommel, J.R., Lightbourn, G.J., Winkel, B.S. & Griesbach, R.J. (2009) Transcription Factor Families Regulate the Anthocyanin Biosynthetic Pathway in *Capsicum annuum*. Journal of the American Society for Horticultural Science 134: 244-251.
5. Tanaka, Y., Sasaki, N. & Ohmiya, A. (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant Journal 54: 733-749.
6. Holton, T.A. & Cornish, E.C. (1995) Genetics and Biochemistry of Anthocyanin Biosynthesis. The Plant cell 7: 1071-1083.
7. Almeida, J.R. *et al.* (2007) Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria xananassa*). Arch Biochem Biophys 465: 61-71.
8. Li, L. *et al.* (2012) Differential expression of anthocyanin biosynthetic genes and transcription factor PcMYB10 in pears (*Pyrus communis* L.). PLoS One 7: e46070.
9. Zhang, X. *et al.* (2011) Differential Gene Expression Analysis of Yunnan Red Pear, *Pyrus Pyrifolia*, During Fruit Skin Coloration. Plant Molecular Biology Reporter 29: 305-314.
10. Rafique, M.Z. *et al.* (2016) Nonsense Mutation Inside Anthocyanidin Synthase Gene Controls Pigmentation in Yellow Raspberry (*Rubus idaeus* L.). Front Plant Sci 7: 1892.

11. Zhang, Q. *et al.* (2021) Postharvest UV-C irradiation increased the flavonoids and anthocyanins accumulation, phenylpropanoid pathway gene expression, and antioxidant activity in sweet cherries (*Prunus avium* L.). *Postharvest Biology and Technology* 175: 111490.
12. Starkevič, P. *et al.* (2015) Expression and Anthocyanin Biosynthesis-Modulating Potential of Sweet Cherry (*Prunus avium* L.) MYB10 and bHLH Genes. *PLoS One* **10**: e0126991.