



HAL
open science

Origin and evolution of geraniol biosynthesis in rose petals

Corentin Conart

► **To cite this version:**

Corentin Conart. Origin and evolution of geraniol biosynthesis in rose petals. *Vegetal Biology*. Université Jean Monnet - Saint-Etienne, 2022. English. NNT : 2022STET0033 . tel-04276357

HAL Id: tel-04276357

<https://theses.hal.science/tel-04276357>

Submitted on 9 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N°d'ordre NNT : 2022STET0033

THÈSE de DOCTORAT DE L'UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE

Membre de l'Université de LYON

École Doctorale N° 488
Sciences Ingénierie Santé

Spécialité / discipline de doctorat :
Biologie et physiologie Végétale

Soutenue publiquement le 08/11/2022, par :
Corentin Conart

Origin and evolution of geraniol biosynthesis in rose petals

Devant le jury composé de :

HUGUENEY Philippe, Directeur de recherche, Université de Strasbourg (INRAE Colmar),
Président

GIGLIOLI-GUIVARC'H Nathalie, Professeure, Université de Tours (BBV), **Rapporteuse**
GAQUEREL Emmanuel, Professeur, CNRS-Université de Strasbourg (IBMP), **Rapporteur**
REBOLLO Rita, Chercheuse, INSA Lyon (BF2I), **Examinatrice**
HUGUENEY Philippe, Directeur de recherche, Université de Strasbourg (INRAE Colmar),
Examineur

CAISSARD Jean-Claude, Maître de conférence (HDR), Université de Lyon-Saint-Étienne
(LBVpam), **Directeur de thèse**
BOACHON Benoît, Ingénieur de recherche, Université de Lyon-Saint-Étienne (LBVpam),
Co-directeur de thèse
BAUDINO Sylvie, Professeure, Université de Lyon-Saint-Étienne (LBVpam), **Directrice du
laboratoire**

Origin and evolution of geraniol biosynthesis in rose petals

Thèse de doctorat de l'Université de Lyon-Saint-Étienne

*préparée au Laboratoire de Biotechnologies Végétales
appliquées aux plantes aromatiques et médicinales*

École doctorale n°488, Sciences Ingénierie Santé (SIS)

Discipline : Biologie et Physiologie Végétale

Thèse présentée et soutenue à Saint-Étienne, le 8 novembre 2022, par

Corentin Conart

Composition du jury :

Nathalie GIGLIOLI-GUIVARC'H,
Professeure, Université de Tours (BBV)

Rapporteure

Emmanuel GAQUEREL,
Professeur, CNRS-Université de Strasbourg (IBMP)

Rapporteur

Rita REBOLLO,
Chercheuse, INSA Lyon (BF2I)

Examinatrice

Philippe HUGUENEY,
Directeur de recherche, Université de Strasbourg (INRAE Colmar)

Examinateur

Jean-Claude CAISSARD,
Maître de conférence (HDR), Université de Lyon-Saint-Étienne
(LBVpam)

Directeur de thèse

Benoît BOACHON,
Ingénieur de recherche, CNRS (LBVpam)

Co-encadrant de
thèse

Sylvie BAUDINO,
Professeure, Université de Lyon-Saint-Étienne (LBVpam)

Directrice du
laboratoire

Table of contents

| | |
|---|-----|
| Table of contents | 3 |
| Table of figures | 4 |
| Table of tables | 6 |
| Table of supplemental tables | 7 |
| Table of supplemental figures | 8 |
| Abbreviations | 9 |
| Acknowledgments | 11 |
| Introduction | 13 |
| I- Rose fragrance biosynthetic pathways..... | 21 |
| I-1. Biosynthesis of phenolic methyl ethers..... | 22 |
| I-2. Biosynthesis of phenylpropanoids..... | 22 |
| I-3. Biosynthesis of terpenes and terpenoids..... | 24 |
| II- Diversity of enzymes producing terpenes in plants..... | 26 |
| II-1. The TPS family..... | 26 |
| II-2. Other enzymes producing terpenes and terpenoids..... | 29 |
| II-3. The NUDX1 family..... | 29 |
| III- Biosynthesis of terpene precursors in plants..... | 31 |
| III-1. The MVA and MEP pathways..... | 31 |
| III-2. Key regulators for IPP and DMAPP supply..... | 32 |
| III-3. Active sites and functioning of IDSs..... | 34 |
| III-4. Dimeric structures of LSU and SSU..... | 36 |
| III-5. Homodimeric structures of IDSs..... | 38 |
| Chapter 1 | 39 |
| General outline Chapter 1..... | 39 |
| Article 1 Functional diversification in the <i>Nudix hydrolase</i> gene family drives sesquiterpene biosynthesis in <i>Rosa x wichurana</i> | 40 |
| Chapter 2 | 55 |
| General outline Chapter 2..... | 55 |
| Article 2 Duplication and specialization of <i>NUDX1</i> in <i>Rosaceae</i> led to geraniol production in rose petals | 56 |
| Chapter 3 A conserved bifunctional geranyl/farnesyl diphosphate synthase in <i>Rosaceae</i> provides cytosolic GPP precursor for <i>NUDX1-1a</i>-dependent geraniol biosynthesis in rose flowers | 77 |
| General outline Chapter 3..... | 77 |
| I- Introduction..... | 78 |
| II- Results..... | 79 |
| III- Discussion..... | 92 |
| IV- Material and methods..... | 94 |
| General discussion | 101 |
| I – Diversification of <i>NUDX1</i> in <i>Rosaceae</i> and in <i>Rosa</i> | 102 |
| II – Specialization of <i>NUDX1</i> in terpene production..... | 105 |
| III – An evolutive scenario of geraniol production..... | 109 |
| Appendix 1 | 113 |
| Appendix 2 | 129 |
| Appendix 3 | 153 |
| Bibliography | 165 |

Table of figures

| | | |
|---------------------|--|----|
| Figure 1 | Network representing the relationships among copies of GAPDH obtained from <i>Rosa</i> species..... | 13 |
| Figure 2 | Classification of cultivated roses..... | 14 |
| Figure 3 | Isoprenoid biosynthetic pathway in plant cell..... | 16 |
| Figure 4 | Geraniol dependent Nudix hydrolase I pathway in rose..... | 18 |
| Figure 5 | Phenolic methyl ethers biosynthesis pathway in rose..... | 20 |
| Figure 6 | Phenylpropanoids biosynthesis pathway in rose..... | 21 |
| Figure 7 | Reaction mechanism of terpenes synthase..... | 26 |
| Figure 8 | Mode of action of prenyl diphosphate synthase..... | 32 |
| Figure 9 | biosynthesis of prenyl diphosphate in plant..... | 34 |
| Figure 1.1 | (a) Alignment of amino acid sequences of NUDX1 (b) Unrooted neighbor joining tree displaying the relationships of rose NUDX1 proteins with AtNUDX1 from <i>A. thaliana</i> | 42 |
| Figure 1.2 | (a) Gene expression levels of <i>NUDX1-1</i> and <i>NUDX1-2</i> in 16 F ₁ individuals from a cross between <i>R. chinensis</i> cv. 'Old Blush' (OB) and <i>R. x wichurana</i> (Rw) and their parents. (b) Heatmap correlation of <i>NUDX1-1</i> and <i>NUDX1-2</i> expressions with some of the volatile scent compounds..... | 43 |
| Figure 1.3 | Genetic map of the male linkage group 7, based on SNP and <i>RwNUDX1-2</i> markers (b) A LOD score curve obtained from interval mapping analysis of <i>E,E</i> -farnesol contents in flowers on the male linkage group B7..... | 44 |
| Figure 1.4 | Patterns of 4odellin glycoside accumulation following transient expression of <i>NUDX1</i> genes in <i>N. benthamiana</i> | 46 |
| Figure 1.5 | (a) Superimposition of the RhNUDX1/GPP complex (pink) and apo RhNUDX1 (blue). (b) Structure of the RhNUDX1/GPP complex. | 46 |
| Figure 1.6 | Comparison of GPP ligand position (a) in the crystal structure of RhNUDX1 and (b) in the model obtained by molecular 4odelling. | 46 |
| Figure 2.1 | Overview of the evolution of the Rosaceae family and of the <i>Rosa</i> genus..... | 58 |
| Figure 2.2 | ML tree of genomic sequences of NUDX1 homologs in the Rosaceae..... | 59 |
| Figure 2.3 | Gene map of RcNUDX1 in Old Blush..... | 60 |
| Figure 2.4 | Synteny map of the Rosaceae genomes..... | 61 |
| Figure 2.5 | ML tree of genomic sequences of the Nudx1-1 clade. | 63 |
| Figure 2.6 | Organization of the shared Tes around the NUDX1-1a and NUDX1-1b sequences in three accessions: Old Blush, Moschata, and Laevigata..... | 65 |
| Figure 2.7 | Correlation between the expression of NUDX1-1 homologs and the number of gene sequences in rose species. | 66 |
| Figure 2.8 | Alignment interpretation of box38 of chromosomes 2 and 4 of Old Blush genome. | 67 |

Table of figures

| | | |
|----------------------|---|----|
| Figure 2.9 | Confocal laser scanning microscopy of transient expression of GFP constructs in agroinfiltrated petals of Old Blush. | 68 |
| Figure 2.10 | Scenario of evolution of NUDX1 in botanical roses..... | 69 |
| Figure 3.1 | Geraniol is synthesized through the MVA pathway in cytosol of rose flowers..... | 80 |
| Figure 3.2 | The OB genome contains six IDS candidates of which two FPPS-like synthases localized in the cytosol..... | 82 |
| Figure 3.3 | RcG/FPPS1 is a bifunctional enzyme producing GPPS and FPPS in vitro and <i>in planta</i> | 84 |
| Figure 3.4 | <i>RcG/FPPS1</i> rhythmical expression pattern precedes GPP and FPP accumulation in rose flowers and geraniol emission..... | 86 |
| Figure 3.5 | The bifunctional G/FPPS activity is conserved in <i>Rosaceae</i> | 88 |
| Figure 3.6 | Two amino acid are involved in the GPPS activity of roses FPPS1..... | 90 |

Table of tables

| | | |
|--------------------|--|----|
| Table 1 | Classification of the genus <i>Rosa</i> | 12 |
| Table 2 | Characterized GES in plants..... | 18 |
| Table 3 | Characterized genes implicated in rose compound biosynthesis in different rose species or cultivars..... | 19 |
| Table 4 | Family of TPS and their function in different organism groups..... | 25 |
| Table 1.1 | Major volatile compounds extracted from petals of OB and <i>R. x wichurana</i> (Rw) at fully open flower stage (stage four) and analyzed by GC-MS..... | 43 |
| Table 1.2 | Kinetic parameters of NUDX1 proteins: RcNUDX1-1a, RcNUDX1-1b, RhNUDX1, RwNUDX1-2c, and AtNUDX1, with several potential substrates..... | 45 |
| Table 2.1 | Comparison of Geraniol Concentration and Expression of NUDX1-1 Homologs in Wild and Heritage Roses..... | 62 |
| Table 3.1 | Kinetic parameters of OB IDS candidates and RcG/FPPS1 mutants..... | 83 |

Table of tables

| | | |
|--------------------|--|----|
| Table 1 | Classification of the genus <i>Rosa</i> | 12 |
| Table 2 | Characterized GES in plants..... | 18 |
| Table 3 | Characterized genes implicated in rose compound biosynthesis in different rose species or cultivars..... | 19 |
| Table 4 | Family of TPS and their function in different organism groups..... | 25 |
| Table 1.1 | Major volatile compounds extracted from petals of OB and <i>R. x wichurana</i> (Rw) at fully open flower stage (stage four) and analyzed by GC-MS..... | 43 |
| Table 1.2 | Kinetic parameters of NUDX1 proteins: RcNUDX1-1a, RcNUDX1-1b, RhNUDX1, RwNUDX1-2c, and AtNUDX1, with several potential substrates..... | 45 |
| Table 2.1 | Comparison of Geraniol Concentration and Expression of NUDX1-1 Homologs in Wild and Heritage Roses..... | 62 |
| Table 3.1 | Kinetic parameters of OB IDS candidates and RcG/FPPS1 mutants..... | 83 |

Table of supplemental figures

| | | |
|--------------------|---|-----|
| Figure S1.1 | Alignment of amino acid sequences of four RcNUDX1-1a copies..... | 122 |
| Figure S1.2 | Distribution in the OW progeny of volatile compound amounts..... | 123 |
| Figure S1.3 | Localization of the QTLs for E,E-farnesol and other sesquiterpenoid compounds..... | 124 |
| Figure S1.4 | Microsynteny analysis at the NUDX1-2c locus..... | 125 |
| Figure S1.5 | Amino acid sequence alignment and structure indication of NUDX1 proteins..... | 126 |
| Figure S1.6 | Analysis of the 20 ns trajectories for NUDX1 proteins..... | 127 |
| Figure S1.7 | Models of NUDX1–substrate interactions..... | 128 |
| Figure S2.1 | Expression of <i>RcNUDX1</i> genes in petals of full-opened flowers of Old Blush..... | 130 |
| Figure S2.2 | Extended ML tree of genomic sequences of the Nudx1-1 clade. | 131 |
| Figure S2.3 | Heatmap correlation between expression of <i>NUDX1-1</i> homologs and concentrations of volatile compounds in 34 accessions of botanical roses..... | 132 |
| Figure S2.4 | Study of intergenic homologies in Old Blush chromosome 2. | 133 |
| Figure S2.5 | Estimation of the number of copies of <i>NUDX1-1</i> homologs in 12 accessions of wild diploid rose, and Old Blush. | 134 |
| Figure S2.6 | Multiple sequence alignments (MAFFT) of <i>Copia R24588</i> hits in Old Blush homozygous genome..... | 135 |
| Figure S2.7 | Organization of the box38 repetitions upstream the coding region of NUDX1-1a homologs..... | 136 |
| Figure S3.1 | Subcellular fractionation of rose petal crude extracts..... | 154 |
| Figure S3.2 | The trans-short-chain IDSs member in OB and their characteristics..... | 155 |
| Figure S3.3 | RcG/FPPS1 is a homodimer and RcGPPS.SSU forms a heterodimer with RcGGPPS.LSU1..... | 156 |
| Figure S3.4 | Enzymatic characterization of the five rose IDS candidates..... | 157 |
| Figure S3.5 | Transcriptomic analysis of and metabolite profiling of OB genes during flower development and rhythmicity..... | 158 |
| Figure S3.6 | FvFPPS1.2 having His residue at position 100 has low G/FPPS activity..... | 160 |
| Figure S3.7 | List of m/z ions used to calculate VOCs labelling..... | 161 |
| Figure S3.8 | List of primers used in this study..... | 162 |
| Figure S3.9 | Accession numbers of genes used in this study..... | 163 |

Abbreviations

| | |
|---|---|
| AAT: alcohol acetyltransferase | IDS: isoprenyl-diphosphate synthase |
| ADH: alcohol dehydrogenase | IP: isopentenyl monophosphate |
| BLAST: Basic Local Alignment Search Tool | IPK: isopentenyl monophosphate kinase |
| CCD: Carotenoid cleavage dioxygenase | IPP: isopentenyl diphosphate |
| ChPP: chrysanthemyl diphosphate | Km: Michaelis & Menten constant |
| DMAP: dimethylallyl monophosphate | LINS: linalool synthase |
| DMAPP: dimethylallyl diphosphate | LPP: lavandulyl diphosphate |
| DMT: 3,5-dimethoxytoluene | LSU: large sub unit |
| DOX: 1-deoxy-D-xylulose5-phosphate | MD: molecular dynamic |
| DXR: 1-deoxy-D-xylulose 5-phosphate reductoisomerase | MEP: 2-C- methyl-D-erythritol 4-phosphate |
| DXS: 1-deoxy-D-xylulose 5-phosphate synthase | MULE: Mutator-like transposable elements |
| EGS: eugenol synthase | MITE: Miniature Inverted-repeat Transposable |
| FARM: first aspartate rich motif | MVA: mevalonate |
| FP: farnesyl monophosphate | NPP: neryl diphosphate |
| FPP: farnesyl pyrophosphate | NUDX: Nudix hydrolase |
| FPPS: farnesyl diphosphate synthase | OB: <i>Rosa chinensis</i> cv. 'Old Blush' |
| GAPDH: Glyceraldehyde-3-phosphate dehydrogenase | OMT: O-methyltransferase |
| GDS: germacrene D synthase | OOMT: orcinol O-methyltransferase |
| GES: geraniol synthase | PAAS: phenylacetaldehyde synthase |
| GGPPS: geranylgeranyl diphosphate synthase | POMT: phloroglucinol O-methyltransferase |
| GGPP: geranylgeranyl diphosphate | QTL: quantitative trait loci |
| GP: geranyl monophosphate | SARM: second aspartate rich motif |
| GPP: geranyl diphosphate | SSU: small sub unit |
| GPPS: geranyl-geranyl diphosphate synthase | TIR: Tandem inverted repeat |
| HMGR: 3-hydroxy-3-methylglutaryl CoA reductase | TMB: 1,3,5-trimethoxybenzene |
| IDI: isopentenyl diphosphate isomerase | TPS: terpene synthase |
| | VIGS: virus induced gene silencing |
| | VOCs: volatile organic compounds |

Acknowledgments

First, I would like to express my deep gratitude to my thesis committee who agreed to evaluate my work and take their time to read it.

A thesis can't be done without a supervisor, so I would like to give a special thanks to Jean-Claude Caissard, my PhD supervisor, who was always there to support me during my thesis. I will always remember your happiness each time I presented you the good results.

Then, I would like to thank Benoît Boachon, my PhD co-supervisor, for his trust and his time spent to teach me so many techniques which I can't enumerate/count.

Sylvie Baudino, the director of the laboratory, I thank you to welcome me in the LBVpam team and provide us this good environment for scientific research.

I can't write acknowledgements without thinking of all colleagues, with a special thanks to Aya and Ludivine for the care they give to my "baby" Old Blush. I also have to thank Aurelie, without you, experiments wouldn't work.

Finally, for all the French and international collaborators, it was always a pleasure to present you my work and discuss about science, and I feel lucky for the opportunities to share some scientific moment with all of you.

Quatre ans, ça passe vite, Saint-Etienne et ses rencontres seront de merveilleux souvenirs. Charlotte, Léa : Merci pour tout ! Je fais court car cela serait trop long. En remontant dans les étages, les chimistes Fabien et Samira j'ai une pensée pour vous en finissant cette thèse. L'ancienne collègue, mais surtout amie, Laure, avec qui j'ai pu échanger sur tout, merci, je te souhaite plein de belles choses dans ta vie de scientifique (ton tapuscrit m'a beaucoup aidé). Comment ne pas finir en parlant de personnes rencontrées à la Métare sans parler de Saretta qui m'a apporté énormément, encouragé, soutenu, motivé et bien plus ! Ces quatre années de thèse auraient été beaucoup plus moroses sans avoir eu la chance de vous rencontrer et découvrir. Pour tous les moments passés ensemble et souvenirs créés, je tenais à vous remercier.

J'ai une très grosse pensée à l'association scientifique des étudiants jeunes chercheurs de Saint-Etienne (ASEC) pour toutes les rencontres entre doctorants et post-doctorants quelle m'a permis de faire. Et notamment le bureau de l'ASEC de l'année 2019 avec qui c'était un plaisir d'être le « secrétaire », Carine, Marion, Bastien, Cyriac, Mathilde et Elodie. Les rencontres ASEC n'auront donc jamais été à sec !

Je tiens à remercier ma famille qui m'a toujours encouragé durant mon parcours universitaire et m'a soutenu dans le choix de faire une thèse. J'inclus également les « copains » Kevin et Vincent avec qui les 590 km qui séparent la Marne de Saint-Etienne n'ont été qu'une raison de plus pour partager l'apéritif à chacun de mes retours.

Introduction |

Roses belong to the *Rosaceae* family. *Rosaceae* is part of the *Rosid* clade, one of the biggest monophyletic groups of flowering plants. *Rosaceae* include flowers with five petals and numerous stamens. It is divided into three subfamilies: *Rosoideae*, *Amygdaloideae* and *Dryaoideae* including about 2000, 1000 and 30 species respectively. They include high valuable species for human as *Fragaria*, *Rosa* (subfamily *Rosoideae*), *Rubus*, *Malus*, *Prunus*, and *Pyrus* (subfamily *Amygdaloideae*) (Xiang *et al.* 2016). Thanks to their importance for human food, many of these species are sequenced with available genome on-line (<https://www.rosaceae.org/>; Jung *et al.* 2019).

The genus *Rosa* consists of around 150-200 wild species, and contains four subgenus named *Hulthemia*, *Platyrhodon*, *Hesperhodos*, and *Rosa*. Subgenus *Rosa* contains twelve sections (Table 1; Rehder, 1940; Wissemann, 2003), which are not totally validate by phylogenetic studies (Figure 1; Fougère-Denazan *et al.* 2015). Indeed, it is not monophyletic, but it is divided into more than two subclades in which the biggest are *Synstylae* and *Cinnamomeae*. Beside the reticulate evolution of the genus *Rosa*, it is assumed that some clades are monophyletic and considered as more ancient roses as *Laevigatae* (*R. laevigata*), *Bracteatae* (*R. bracteata*), and *Banksianae* (*R. banksiae*) (Fougère-Denazan *et al.* 2015; Zhu *et al.* 2015). Ploidy level in rose is also complicated, ranging from diploid to decaploid (Wissemann, 2003). This may explain a part of the difficulty to produce a phylogeny of *Rosa* genus due to interspecific hybridizations. For example, in the same section *Caninae*, *R. canina*, a wild European rose species, is pentaploid, but other species are tetraploid and hexaploid (Lim *et al.* 2005; Ritz *et al.* 2005). Furthermore, genomic organisation of *Rosaceae* is quite complicated. Ancestral *Rosaceae* are supposed to have nine chromosomes. Remodelling of genome in *Rosaceae* has evolved differently by fusions and fissions. In some of *Rosaceae*, there has been some whole genome duplication for example in *Malus* and *Pyrus*, but most of species of the *Rosoideae* subfamily such as *Fragaria* and *Rosa* have seven chromosomes (Vilanova *et al.* 2008; Jung *et al.* 2012; Xiang *et al.* 2016). As a consequence, classification and phylogeny of *Rosaceae* and *Rosa* are still much debated.

Table 1 | Classification of the genus *Rosa* (Rehder, 1940 revised by Wissemann, 2003)

| Subgenus | Section | |
|------------------|---------------------------------|--|
| <i>Hulthemia</i> | | |
| <i>Rosa</i> | <i>Pimpinellifoliae</i> | |
| | <i>Rosa</i> = <i>Gallicanae</i> | |
| | <i>Caninae</i> | |
| | <i>Carolinae</i> | |
| | <i>Cinnamomae</i> | |
| | <i>Synstylae</i> | |
| | <i>Indicae</i> | |
| | <i>Banksianae</i> | |
| | <i>Laevigatae</i> | |
| | <i>Bracteatae</i> | |
| | <i>Platyrhodon</i> | |
| | <i>Hesperodos</i> | |

Humans used roses since antiquity for their fragrance and pleasant perfume. Nowadays, most of humans know roses for their flowers and associated symbols. Humans have spread roses all around the world and have created thousand varieties with many different characters. Rose breeders used this diversity to make new varieties. They have crossed roses from the *Chinenses* section originated from East Asia, roses from the *Gallicanae* section from central Europe, roses from *Caninae* section from Europe and Asia, and many other species. Cultivated roses can be classified as species roses (also named wild roses), old garden roses (cultivated since antiquity), and modern roses (roses crossed and selected by humans; Wissemann, 2003; Figure 2). Numerous crosses made by breeders have allowed the generation of thousand hybrids (more than 30,000), in which *R. x hybrida* cv. 'La France' created in 1867 is the most famous as it is considered as the first “true” modern rose. This hybrid combines the growth-vigour of European roses, a typical rose fragrance, and the recurrent blooming of roses from the *Chinenses* section. However, such classifications of horticultural roses seem artificial melting a “temporal vision of the plant’s origin and the static view of its place in systems of classification” (Oghina-Pavie, 2015). Nevertheless, in a genetic point of view, roses from Europe and Middle-East were progressively crossed with roses from Asia (Liorzou *et al.* 2016) sometimes for different fragrances, sometimes for everblooming traits, or sometimes for colours for example.

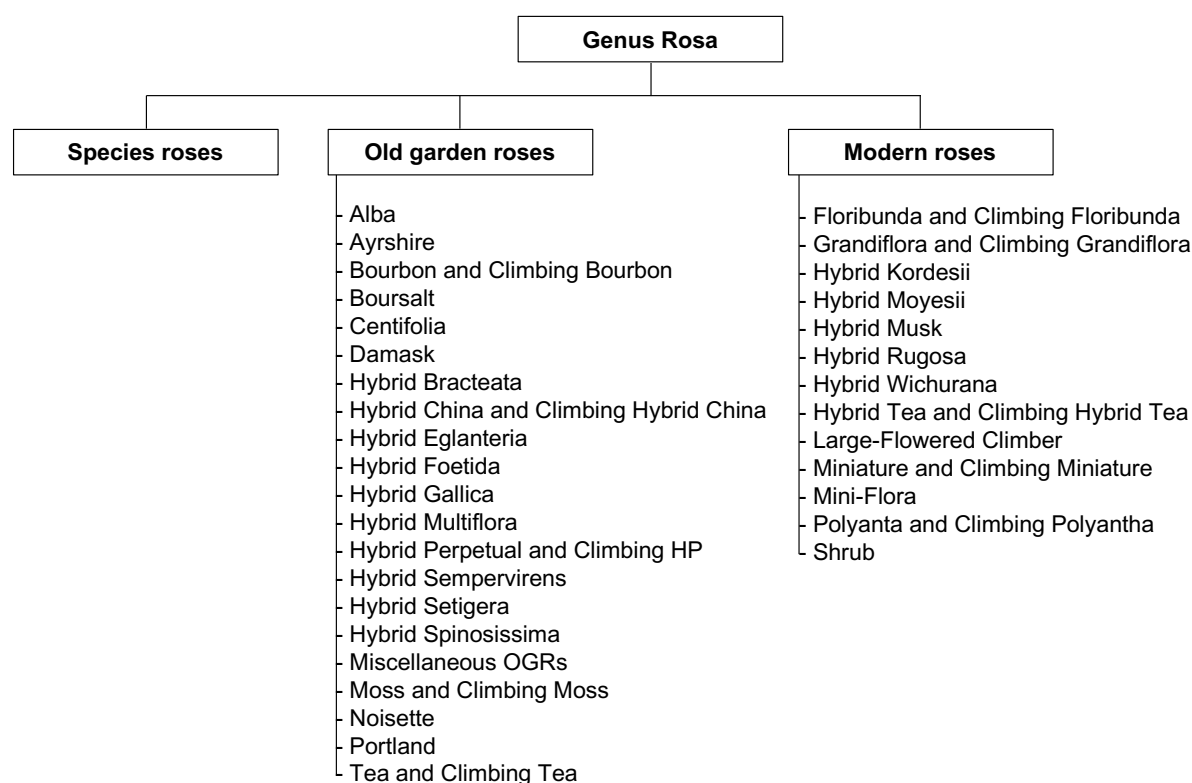


Figure 2 | Classification of cultivated roses (American Rose Society, 2000)

For the selection of new rose varieties, many characters are used: flower type, colour, shape, disease resistance, fragrance, vase life, number of flowers per stem, *etc.* Hybridization and natural or directed

polyploidization allow the selection of numerous traits in roses. Some crosses between wild species led to horticultural groups, themselves included in middle age roses and modern roses: Noisette, Bourbon, Portland, Tea, and Tea Hybrids roses for example (Figure 2). Between 7-8 to 15-20 wild species (depending on the sources) have been used intensively to produce today's cultivars (Kumari *et al.* 2021, Liorzou *et al.* 2016). The research interest for roses is clearly visible due to their diversity in morphological traits. These interests can explain the research effort on producing high quality genome references. Three genomes of *R. chinensis* cv. 'Old Blush' (shortly named Old Blush in our work) have been published, two of them coming from haploid calli (Raymond *et al.* 2018; Hibrand Saint-Oyant *et al.* 2018).

Scent extracts of very scented roses have been used in perfumes to make some important valuable products, even in the Antiquity (Krüssmann 1981; Widrechner 1981). Varieties of *R. x damascena* and *R. x centifolia* are used for hundred years for these extractions, from enfleurage in olive oil in Antiquity, to essential oil distillation in the Middle Ages, and solvent or supercritical CO₂ extraction nowadays. Rose scent is also a very important trait for selection of garden roses, and recently for the cut-flower market, but it is a complex trait because of the mix of hundreds of VOCs (volatile organic compounds). For example, the broodstocks used today contain phenolic methyl ethers such as TMB and DMT (1,3,5-trimethoxybenzene and 3,5-dimethoxytoluene respectively). European roses do not produce phenolic methyl ethers. These compounds come from the ancient crosses with Chinese roses, which are known to have a "tea scent" due to DMT. Tea and Tea Hybrid roses are the result of these crossings, and explain the presence of this compound in modern roses (Scalliet *et al.* 2002; 2006; 2008). As a consequence, the genetic background of the broodstocks is largely made of genes from roses of the *Chinenses* section (Liorzou *et al.* 2016). New crosses are thus difficult to drive, as the typical rose scent must include specific terpenes and phenylpropanoids among others, but not DMT. Unfortunately, coming back to old cultivars as genitors could have huge effects on diseases, recurrent blooming, and vase life.

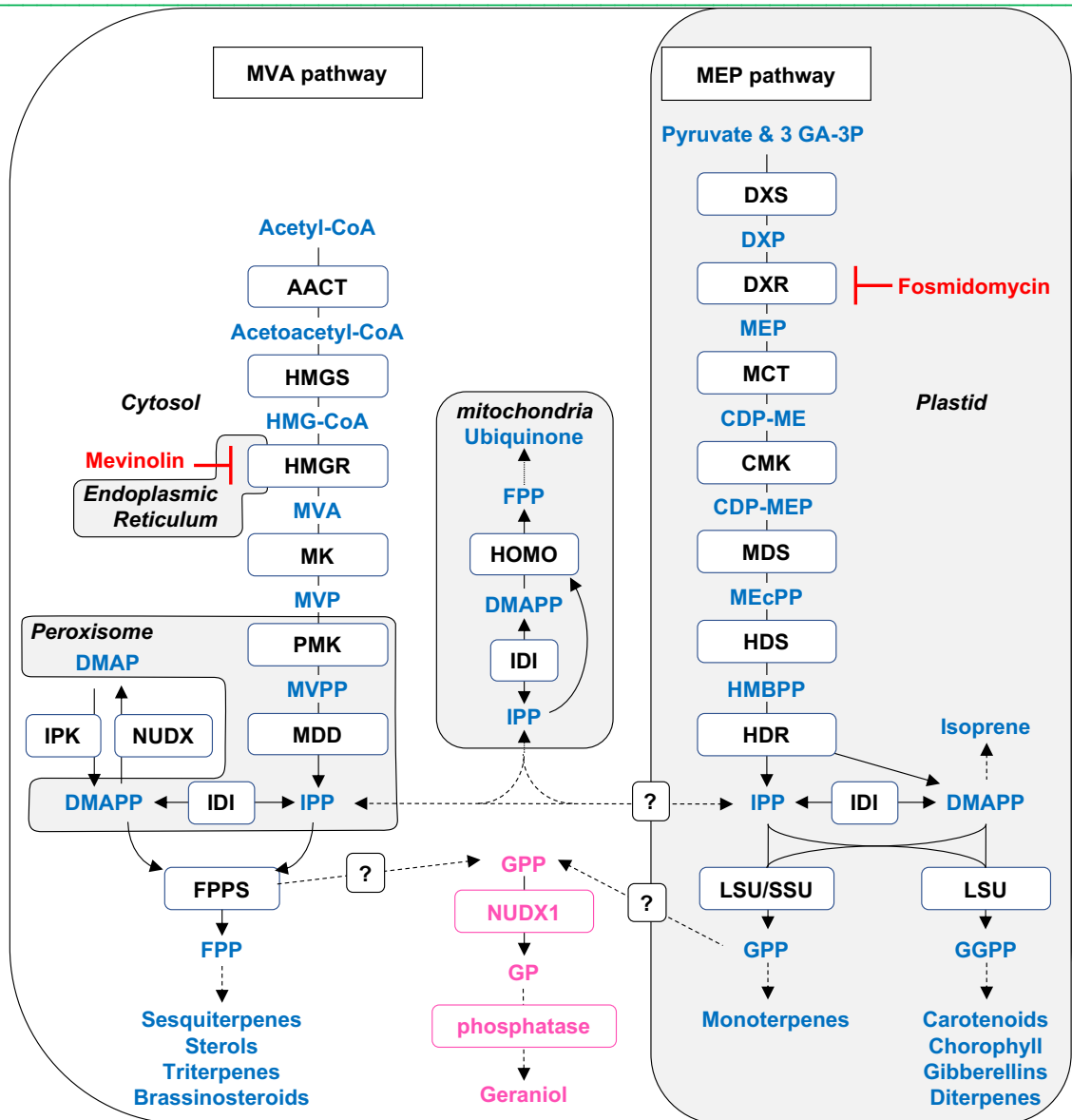
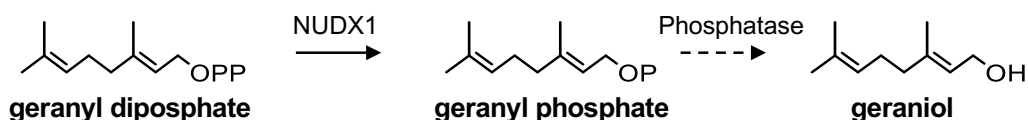


Figure 3 | isoprenoid biosynthetic pathway in plant cell. Adapted from Pulido *et al.* 2012. Enzyme name are in squared and products are in blue, In red are represented specific inhibitors. Dashed arrows represent multiple enzymatic step or unknown enzyme. AACT, Acetoacetyl-CoA thiolase; DMAP, dimethylallyl phosphate; DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-D-xylulose5-phosphate; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; CDP-ME, 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol; CDP-MEP, 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol 2-phosphate; CMK, 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase; FPP, farnesyl diphosphate; FPPS, farnesyl diphosphate synthase; GA-3P, glyceraldehyde 3-phosphate; GP, geranyl phosphate ; GPP, geranyl diphosphate; GGPP, geranylgeranyl diphosphate; HDR, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate reductase; HDS, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate synthase; HMBPP, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate; HMGR, 3-hydroxy-3-methylglutaryl CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl CoA synthase; HOMO, GPPS homodimeric; IPK, isopentenyl phosphate kinase; IPP, isopentenyl diphosphate; LSU, large sub-unit; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; MDD, 5-diphosphomevalonate decarboxylase; IDI, isopentenyl diphosphate isomerase; MEcPP, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2-C- methyl-D-erythritol 4-phosphate; MK, mevalonate kinase; MVA, mevalonic acid; MVP, 5-phosphomevalonate; MVPP, 5-diphosphomevalonate; NUDX, nudix hydrolase; PMK, 5-phosphomevalonate kinase; SSU, small sub-unit.

Rose petal VOCs are derived from many classes of compounds and are not restricted to phenolic methyl ethers. Phenylpropanoids/benzenoids and terpenoids are the most important VOCs family in rose petals (Baldermann *et al.* 2009). Phenylpropanoids/benzenoids are derived from aromatic aminoacids. In *Rosa*, they contain eugenol, benzyl alcohol or 2-phenylethanol for example. Surprisingly, in rose petals, three different biosynthetic pathways have been discovered for 2-phenylethanol (Tieman *et al.* 2006; Sakai *et al.* 2007; Chen *et al.* 2011; Sheng *et al.* 2018 ; Rocchia *et al.* 2019), the most important phenylpropanoid for the typical scent of rose. Terpenoids also shows biochemical originality in roses. They are polymers of the prenyl building blocks IPP (isopentenyl diphosphate) and DMAPP (dimethylallyl diphosphate). These compounds are synthesized by two different pathways (Figure 3): the cytosolic mevalonate pathway (MVA pathway), and the plastidial 2-C-methyl-D-erythritol-4-phosphate pathway (MEP pathway) (Lichtenthaler, 2001). The MVA pathway is involved in sesquiterpenes, triterpenes, dolichol, sterols and brassinosteroids synthesis. It always begins by the polymerisation of two units of IPP and one unit of DMAPP to give farnesyl-diphosphate (FPP). The MEP pathway is involved in monoterpenes, diterpenes, chlorophyll, carotenoids and gibberellins synthesis (Lichtenthaler *et al.* 2001). It always begins by the polymerisation of one unit of IPP and one unit of DMAPP to give geranyl-diphosphate (GPP), or three units of IPP and one unit of DMAPP to give geranylgeranyl-diphosphate (GGPP) (Figure 3). Rare exceptions exist in some species and one biochemical pathway of the rose is one of them. Indeed, this is the case of geraniol and its biochemical derivatives (nerol, β -citronellol, and their acetate and aldehyde derivatives), which are very important compounds in rose essential oil. Indeed, in lot of species (Table 2), geraniol is synthesized directly from the plastidial GPP by a terpene synthase (TPS), which removes the two phosphates. Surprisingly, in rose, geraniol is not synthesized by such enzyme, but by two enzymes, the first one, a cytosolic Nudix hydrolase of class I, removes one phosphate, and the second one, an unknown phosphatase, removes the second phosphate (Figure 4). Nudix hydrolases of class I are enzymes that cleave nucleoside diphosphate linked to some moiety X, like 8-oxo-7,8-dihydro-2'-deoxyguanosine 5'-triphosphate (8-oxo-dGTP), an oxidized deoxyribonucleotide that could create a mutation if inserted into DNA during replication (Yoshimura *et al.* 2007). Numerous articles have discussed the role of Nudix hydrolases of class I in bacteria, fungi, mammals, and plants, but none of them had ever considered a role in biosynthesis of perfumes before the article of Magnard *et al.* (2015). Since this article, and since the beginning of this thesis work, other Nudix hydrolases of class I involved in the biosynthesis of other terpenes, or in the regulation of the IPP concentration itself, have been discovered in some species (Henry *et al.* 2018; Li *et al.* 2020; Sun *et al.* 2020). This suggests that this enzyme specialization could have occurred several times in angiosperm evolution bringing up the question of the phylogenetic origin of Nudix hydrolase of class I specialization in *Rosa*, and perhaps in *Rosaceae*. Furthermore, at the beginning of this work, we discovered in our laboratory, that the Nudix hydrolase of class I gene (*RcNUDX1*) was in multiple copies in the genome of Old Blush even though the homologous *AtNUDX1* was in single copy in the model plant *Arabidopsis thaliana* (Henry *et al.* 2018). As a consequence, the question of duplications and diversification was superimposed on that of the specialization.

Table 2 | Characterized geraniol synthases in plants. Only at least *in vitro* tested enzymatic activity are presented.

| Species | Name | References | Years |
|-------------------------------|----------|---------------|-------|
| <i>Ocimum basilicum</i> | ObGES | Lijima et al. | 2004 |
| <i>Cinnamomum tenuipilum</i> | CtGES | Yang et al. | 2005 |
| <i>Perilla frutescens</i> | PfTPS-PL | Ito and Honda | 2006 |
| <i>Catharanthus roseus</i> | CrGES | Simkin et al. | 2012 |
| <i>Valeriana officinalis</i> | VoGES | Dong et al. | 2013 |
| <i>Lippia dulcis</i> | LdGES | Dong et al. | 2013 |
| <i>Daucus carota</i> | DcTPS1 | Yahyaa et al. | 2016 |
| <i>Camptotheca acuminata</i> | CaGES | Chen et al. | 2016 |
| <i>Caladenia plicata</i> | CpGES1 | Xu et al. | 2017 |
| <i>Pelargonium x hybridum</i> | PhGES | Blerot et al. | 2018 |
| <i>Prunus dulcis</i> | PdTPS3 | Nawade et al. | 2019 |
| <i>Dendrobium officinale</i> | DoGES1 | Zhao et al. | 2020 |

**Figure 4 | Geraniol dependent Nudix hydrolase I pathway in rose.** Geranyl diphosphate is dephosphorylated by NUDX1 to produce GP. Dotted line represents an uncharacterized enzyme.

In this work, we asked the following groups of questions:

What is the exact copy number overview of *RcNUDX1* in the genome of Old Blush, and in the genome of sequenced *Rosaceae*? Is it possible to estimate the copy number of *NUDX1* in non-sequenced species of *Rosa*?

What is the functional copy of *RcNUDX1* in rose petals? When did evolve the specialization of *NUDX1*, in the genus *Rosa* or in the *Rosaceae*? What could hypothetically explain this specialization?

As *RcNUDX1* is cytosolic, where does the GPP come from? Is it synthesized by the plastidial MEP pathway and exported by an unknown mechanism, or is it synthesized by a novel cytosolic pathway?

Does this export mechanism of GPP, or this new biosynthesis pathway, have appear in *Rosa* or in the *Rosaceae*?

After a state of art of the literature on enzymatic processes involved in VOCs biosynthesis, this thesis present three chapters. The first one is published in the Plant Journal, and is about the specialization of *NUDX1-2c* and its role in *E,E*-farnesol biosynthesis in roses. The second one is published in Molecular Biology and Evolution. It concerns the origin and the evolution of geraniol production by *NUDX1-1a* in rose. The third one, not yet published is about the origin of cytosolic GPP in rose petals by changing in the function of *FFPS1*. We also add a review in annexes, in which we helped for the GC analyses of the Table.

I- Rose fragrance biosynthetic pathways

Since several decades, considerable efforts have been made to understand VOCs biosynthesis in flowers (for a review see Muhlemann *et al.* 2014). In rose, many enzymes are already known for phenolic methyl ethers, phenylpropanoids and terpenes biosynthesis, but others are still unknown like those concerning oxylipins or benzenoids for example (Table 3).

Table 3 | characterized genes involved in VOCs biosynthesis in different rose species or cultivars. Compounds are organized by their chemical family.

| Family | Compound | Species | Gene | Ref |
|-------------------------|-----------------------|--|------------------------------|----------------------|
| Terpenes | | | | |
| Monoterpenes | | | | |
| | geraniol | <i>R. x hybrida</i> cv 'Papa Meilland' | <i>NUDX1^a</i> | Magnard et al. 2018 |
| | geranial | <i>R. x damascena</i> | <i>AAT1^b</i> | Shalit et al. 2003 |
| | linalool | Old Blush | <i>LINS^c</i> | Magnard et al. 2018 |
| Sesquiterpenes | | | | |
| | germacrene D | <i>R. x hybrida</i> cv 'Fragrant Cloud' | <i>GDS^d</i> | Guterman et al. 2002 |
| | nerolidol | Old Blush | <i>LIN-NERS1^e</i> | Magnard et al. 2018 |
| | | Old Blush | <i>LIN-NERS2^e</i> | Magnard et al. 2018 |
| Apocarotenoids | | | | |
| | β-ionone | <i>R. x damascena</i> | <i>CCD4^f</i> | Huang et al. 2009b |
| | | <i>R. x damascena</i> | <i>CCD1^f</i> | Huang et al. 2009a |
| | 3-hydroxy-β-ionone | <i>R. x damascena</i> | <i>CCD1^f</i> | Huang et al. 2009a |
| Phenylpropanoids | | | | |
| | 2 phenylethanol | <i>R. x hybrida</i> cv 'Fragrant Cloud' | <i>PAAS^g</i> | Farhi et al. 2010 |
| | | <i>R. wichurana</i> | <i>PAAS^g</i> | Roccia et al. 2019 |
| | | <i>R. x damascena</i> | <i>PAR^h</i> | Chen et al. 2010 |
| | | <i>R. x hybrida</i> cv 'Hyves Piaget' | <i>AAATⁱ</i> | Hirata et al. 2012 |
| | | <i>R. x hybrida</i> cv 'Hoh-Jun' | <i>AAATⁱ</i> | Hirata et al. 2012 |
| | | <i>R. x damascena</i> | <i>AAATⁱ</i> | Hirata et al. 2012 |
| | | <i>R. x hybrida</i> cv 'Yves Piaget' | <i>PPDC^j</i> | Hirata et al. 2016 |
| | 2-phenylethyl acetate | <i>R. x damascena</i> | <i>AAT1^b</i> | Shalit et al. 2003 |
| | eugenol | Old Blush | <i>RcEGS1^k</i> | Yan et al. 2018 |
| | methyleugenol | Old Blush | <i>RcOMT1^l</i> | Wu et al. 2003 |
| Benzenoids | | | | |
| | TMB | Old Blush | <i>POMT^m</i> | Wu et al. 2004 |
| | | Old Blush | <i>OOMT1ⁿ</i> | Scalliet et al. 2002 |
| | | <i>R. x hybrida</i> cv 'Golden Gate' | <i>OOMT2ⁿ</i> | Lavid et al. 2002 |
| | DMT | <i>R. x hybrida</i> cv 'Lady Hillingdon' | <i>OOMT1ⁿ</i> | Scalliet et al. 2002 |
| | | Old Blush | <i>OOMT1ⁿ</i> | Scalliet et al. 2006 |
| | | Old Blush | <i>OOMT2ⁿ</i> | Scalliet et al. 2006 |
| | | <i>R. x hybrida</i> cv 'Golden Gate' | <i>OOMT1ⁿ</i> | Lavid et al. 2002 |
| | | <i>R. x hybrida</i> cv 'Golden Gate' | <i>OOMT2ⁿ</i> | Lavid et al. 2002 |

^a nudix hydrolase

^b alcohol acetyltransferases

^c linalool synthase

^d germacrene D synthase

^e linalool-nerolidol synthase

^f carotenoid cleavage dioxygenase

^g phenylacetaldehyde synthase

^h phenylacetaldehyde reductase

ⁱ aromatic amino acid aminotransferase

^j phenylpyruvic acid decarboxylase

^k eugenol synthase

^l O-methyltransferase

^m phloroglucinol O-methyltransferase

ⁿ orcinol O-methyltransferase

I-1. Biosynthesis of phenolic methyl ethers

Phenolic methyl ethers (DMT and TMB) originated from roses of the *Chinenses* section. This family of VOCs is defined as a phenol with hydroxyl groups. DMT gives a particular scent to some species and cultivars of this group, the so called “tea scent”. Phenolic methyl ethers are produced by O-methyltransferases (OMT) which transfer methyl groups from orcinol or phloroglucinol to produce DMT and TMB respectively (Figure 5; Scalliet *et al.* 2002; 2006). More precisely in the case of DMT, a duplication of OOMT (orcinol O-methyltransferase) in *Chinenses*' roses has led to a specific activity in the biosynthesis pathway of DMT. OOMT1 is responsible for the first methylation of orcinol and OOMT2 is responsible for the second methylation resulting to the production of DMT. A single amino acid polymorphism between OOMT1 and OOMT2 results in change of substrate specificity (Scalliet *et al.* 2008). The biosynthesis of TMB needs one more step in the biosynthesis process because there are three methylations. First a phloroglucinol O-methyltransferase (POMT) methylates the phloroglucinol to produce 3,5-dihydroxyanisole. Then OOMT1 and OOMT2 finalize the double methylation to produce TMB (Scalliet *et al.* 2002; Wu *et al.* 2004).

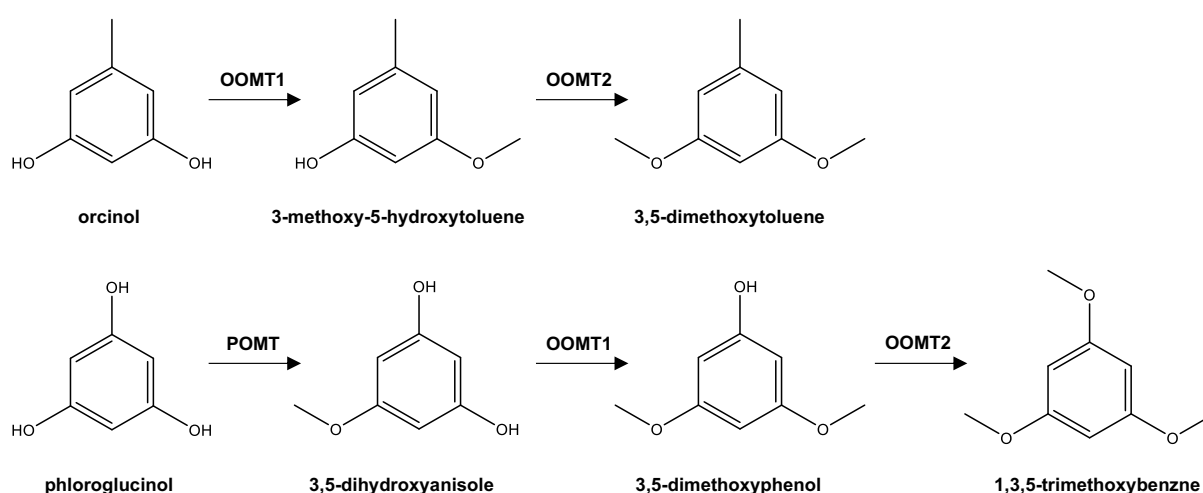


Figure 5 | Phenolic methyl ethers biosynthesis pathway in rose. OOMT1 and OOMT2 are implicated in both 3,5-dimethoxytoluene and 1,3,5-trimethoxybenzene. OOMT, orcinol O-methyltransferase; POMT, phloroglucinol O-methyltransferase.

I-2. Biosynthesis of phenylpropanoids

Phenylpropanoids are derived from phenylalanine and tyrosine amino acids. In rose this VOCs family contains 2-phenylethanol, eugenol, methyleugenol, 2-phenylethyl acetate, and 4-vinylphenol for example. 2-Phenylethanol is describe as a molecule with a sweet rose odour, the same as rose water in which it is mostly extracted during distillation. Unfortunately, it is not present in the rose model, Old Blush, that has been sequenced. Nevertheless, numerous pathways have been proposed in other rose varieties. These pathways always begin by L-phenylalanine as a precursor, giving 2-phenylacetaldehyde in one enzymatic step or in two different steps, and then 2-phenylethanol in one enzymatic step. Thus, several biosynthetic routes exist to transform L-phenylalanine into 2-phenylacetaldehyde in rose (Figure 6).

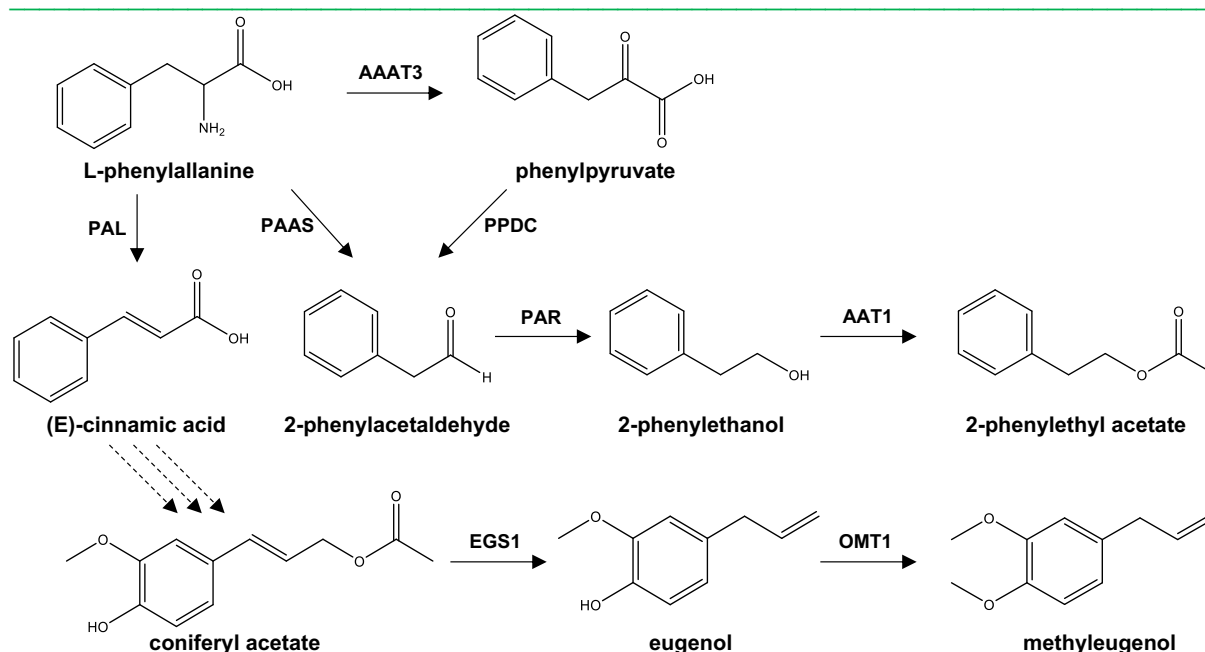


Figure 6 | phenylpropanoids biosynthesis pathway in rose. Two pathways are described in rose to produce 2-phenylacetaldehyde, precursor of 2-phenylethanol. Dotted arrows represent uncharacterized multiple enzymatic steps. AAT1, alcohol acetyltransferases; AAAT3, aromatic amino acid aminotransferase; EGS1, eugenol synthase; OMT1, O-methyltransferase; PAAS, phenylacetaldehyde synthase; PAL, phenylalanine ammonia lyase; PAR, phenylacetaldehyde reductase; PPDC, phenylpyruvic acid decarboxylase.

In *Petunia x hybrida* cv 'Mitchell', *R. x hybrida* cv. 'Fragrant Cloud' and cv 'H190', and *Populus trichocarpa* (poplar) genotype 'Muhle Larsen', a bifunctional phenylacetaldehyde synthase (PAAS) can catalyze both decarboxylation and oxidative deamination reactions and convert L-phenylalanine to 2-phenylacetaldehyde (Kaminaga *et al.* 2006; Farhi *et al.* 2010; Günther *et al.* 2019). Genetic analysis of 2-phenylethanol production in a segregating rose population has shown that this trait depends on specific PAAS alleles (Roccia *et al.* 2019). Aromatic aldehyde synthases (AAS) with the same catalytic properties as PAAS are also active in flowers or leaves of *A. thaliana*, depending on the ecotype (Gutensohn *et al.* 2011). Interestingly, one particular amino acid, present in a catalytic loop of these enzymes, determines whether they act as bifunctional aromatic aldehyde synthases (decarboxylation-deamination), or only as monofunctional aromatic amino acid decarboxylases, like the one found in *Solanum lycopersicum* and *P. trichocarpa* in a different route also producing 2-phenylethanol (Tieman *et al.* 2006). A mutation of this amino acid is sufficient to switch enzyme activities (Torrens-Spence *et al.* 2013; Günther *et al.* 2019). In *P. trichocarpa*, genes coding for AAS and PAAS are clustered on the same chromosome, which suggests that they may have evolved from a common ancestor by gene duplication and neofunctionalization.

In addition, another pathway leading to 2-phenylacetaldehyde biosynthesis from L-phenylalanine also exists in rose (Figure 6). In *R. damascena*, *R. x hybrida* cv. 'Hoh-Jun', and *R. x hybrida* cv. 'Yves Piaget', it has been demonstrated that L-phenylalanine is first converted into phenylpyruvic acid by an aromatic

amino acid aminotransferase, and then undergoes decarboxylation to form 2-phenylacetaldehyde by a phenylpyruvic acid decarboxylase (Hirata *et al.* 2012; Hirata *et al.* 2016). Similarly, in the fruits of *Cucumis melo*, L-phenylalanine is also converted by an amino acid transaminase to phenylpyruvic acid (Gonda *et al.* 2010). Experiments with transgenic petunia plants producing high levels of phenylpyruvate suggest that such an aminotransferase pathway could also be active in *P. x hybrida* petals (Oliva *et al.* 2017). Interestingly in rose, the route B becomes active only in summer under high-temperature conditions, in contrast to the route A, which is active throughout the year, regardless of temperature, and produces a nearly constant amount of 2-phenylethanol. The seasonally induced pathway may have originated in wild rose species as a heat adaptation when roses spread to low latitudes and/or low altitudes. Alternatively, this pathway may have been selected during cultivation to produce blossoms in summer. (Hirata *et al.* 2016).

Finally, in all species, the conversion from 2-phenylacetaldehyde to 2-phenylethanol is supposed to be catalyzed by a phenylacetaldehyde reductase, characterized in rose, poplar and tomato (*S. lycopersicum*) (Tieman *et al.* 2007; Chen *et al.* 2011; Günther *et al.* 2019). 2-phenylethanol may be further converted into ester derivatives, as shown in rose (Shalit *et al.* 2003).

Eugenol biosynthetic pathway has also been studied in rose (Figure 6). Surprisingly a eugenol synthase gene (*RcEGS1*) has been cloned in petals of Old Blush despite the absence of eugenol production in this cultivar. *RcEGS1* probably uses coniferyl acetate, a phenylalanine derivative, as substrate, because of their close homology. To confirm the role of this gene, virus induced gene silencing (VIGS) experiment was performed on *R. x hybrida* cv. 'Yunxiang' where this compound represents 12 % of total volatiles. Reduction of *RcEGS1* expression led to a decrease in production of eugenol in petals (Yan *et al.* 2018). Further modification can occur on this compound and especially methylation. In *R. chinensis spontanea*, methyleugenol and isomethyleugenol are produced by a eugenol O-methyltransferase, *RcOMT1*, that can efficiently methylates eugenol and isoeugenol (Wu *et al.* 2004). *RcOMT1*, compared to other OMT has its maximum activity on eugenol (100 %) and isoeugenol (75 %).

I-3. Biosynthesis of terpenes and terpenoids

Majority of VOCs that are produced by roses are terpenes or terpenes derivatives named terpenoids. They all come from two non-volatile compounds: IPP and DMAPP, which could be considered as two building blocks, each containing a five-carbons chain with two inorganic phosphates. Their polymerisation head-to-head or head-to-tail, and their subsequent modifications lead to more than 30,000 compounds in plants (Connolly and Hill, 1991). Plants can produce these two building blocks by two distinct pathways: the cytosolic MVA pathway, and the plastidial MEP pathway. In each cellular compartment, the polymerisation of IPP and DMAPP by different short chain isoprenyl-diphosphate synthase (IDS) lead to different precursors like the cytosolic FPP and the plastidial GPP and GGPP, for the most common. Each precursor is then used by different TPS to give cyclic terpenes, which could be modified by other enzymes to give terpenoids (Figure 3). Each precursor is the starting point of a terpene pathway: GPP for monoterpenes (C₅H₈)₂, FPP for sesquiterpenes (C₅H₈)₃ and triterpenes (C₅H₈)₆, GGPP for diterpenes (C₅H₈)₄, tetraterpenes (C₅H₈)₈ and carotenoids, for example. It is noticeable, that some

TPSs do not build the terpene ring, but give directly rise to acyclic terpenes with an alcohol group, like geraniol (C₁₀H₁₈O) for monoterpenes, or farnesol isomers (C₁₅H₂₆O) for sesquiterpenes for example.

Thanks to the genome assembly of Old Blush, between 73 and 48 TPSs were found depending on the algorithms and authors' decision to validate a gene as a TPS after BLAST analyses (Raymond *et al.* 2018; Yan *et al.* 2022). Among these TPSs, only five to ten are expressed in petals of open flowers. These TPSs can explain a part of terpene production by Old Blush such as monoterpenes or sesquiterpenes but few have been characterized.

Linalool belongs to the acyclic monoterpene family. In rose petals, it is produced with an enantiomeric specificity, 3S or 3R. In Old Blush, *R. x hybrida* cv 'Lady Hillington', and *R. x hybrida* cv 'The Mc Cartney rose', linalool is only the (-)-(3R) enantiomer. It is produced with GPP as precursor by a plastidial (-)-(3R)-linalool synthase. There are also two linalool/nerolidol synthases untargeted to plastids. *In vitro*, they produce the acyclic nerolidol with FPP as precursor, or (+)-(3R)-linalool with GPP as precursor. However, in the cytosol, these enzymes only produce very few amount of nerolidol probably because it may have reduced access to FPP (Magnard *et al.* 2018). Germacrene D in rose petals is one of the major emitted sesquiterpene. This VOC is produced by a sesquiterpene synthase, germacrene D synthase (GDS) or TPS1 and was the first rose TPS cloned and characterized (Guterman *et al.* 2002).

Apocarotenoids are also an important trait in rose fragrance. These compounds are always in weak amounts but, according to perfumers, some of them, like ionones, are very important because they give the violet note to the rose scent. Apocarotenoids derived from the degradation of carotenoids, themselves synthesized with GGPP as precursor (Figure 3). They are thus also named norisoprenoids. Due to the diversity of carotenoids, numerous products can be formed. Carotenoid cleavage dioxygenase 1 and 4 (CCD1 and CCD4) are the major enzymes responsible of degradation of carotenoids to produce apocarotenoid volatiles with 13 carbons. They cleave symmetrically carotenoids inside the carbon chain, most of time in 9,10 (9',10') double carbon bound. CCD4 is highly expressed in petals of open and senescent flower (Raymond *et al.* 2018) and is able to produce β-ionone *in vitro* with β-carotene but is more efficient with 8'-apo-β-caroten-8'-al (Huang *et al.* 2009b). The *R. x damascena* CDD1 have also been characterized and can produce β-ionone and 3-hydroxy-β-ionone depending on substrate respectively β-carotene or zeaxanthin (Huang *et al.* 2009a). But, CCD1 is probably less implicated than RdCCD4, because it lacks chloroplast transfer peptide, and carotenoids are produced and stored in plastids. In Old Blush petals, the major apocarotenoids are β-ionol and dihydro-β-ionol. Enzymes responsible for their biosynthesis are not yet characterized, and the full comprehension of ionol biosynthesis in rose is still unclear.

II- Diversity of enzymes producing terpenes in plants

II-1. The TPS family

Proteins sequence of TPSs share at least 40 % identity and can be divided into six subfamilies from TPS-a to TPS-h (Chen *et al.* 2011; Table 4). They form a multiple gene family inside genomes but all originate from a common ancestor which was present before angiosperms and the gymnosperms (Bohlmann *et al.* 1998). Beside the identity that can be shared by TPSs, products of these enzymes cannot be predictable only on the basis of phylogeny and homology. They have to be completely characterized *in vitro* for their enzymatic activities. For example, fully characterization of all the TPSs of *S. lycopersicum* was done recently (Zhou and Pichersky, 2020). Nevertheless, phylogeny can give clues to know what a specific TPS produces. TPS-a are sesquiterpene synthases, whereas TPS-b are monoterpene synthases. TPS-b are restricted to gymnosperm. TPS-g produce exclusively acyclic terpenes because of the lack of a domain (RRx₈W) at the beginning of the protein allowing the isomerization-cyclization reaction step (Williams *et al.* 1998; Chen *et al.* 2011). Some TPS can act as both mono- or sesquiterpenes synthase. This is the case for α -zingiberene synthase of *Ocimum basilicum*, which produces α -zingiberene, 7-*epi*-sesquithujene, α -bergamotene, β -bisabolene, and β -sesquiphellandrene with FPP as substrate, but α -thujene, α -terpinene, and terpinolene with GPP (Davidovich-Rikanati *et al.* 2008). This is also the case for linalool-nerolidol synthases from rose (Magnard *et al.* 2008), *Fragaria vesca* (Aharoni *et al.* 2004), *Antirrhinum majus* (Nagegowda *et al.* 2008) *Plectranthus amboinicus* (Ashaari *et al.* 2020), which produce linalool or nerolidol with, respectively, GPP or FPP. Furthermore, the presence of a predictable signal peptide in the protein sequence could also give clues on the enzyme substrate, as GPP and GGPP are plastidial, but FPP not. For example, monoterpene synthases are proteins often longer than sesquiterpene synthases because of the presence of a plastidial signal peptide (Bouvier *et al.* 2000; Degenhardt *et al.* 2009). It is the case of the *Valeriana officinalis* GES, which is clearly localized in plastids and in stromules (Dong *et al.* 2013). This localization is in agreement the one of the final step enzymes of the MEP pathway described in stromules of *Catharanthus roseus* cells (Guirimand *et al.* 2020).

Table 4 | Family of TPSs and their function in different organism groups. Adapted from Chen *et al.* 2011.

| Subfamily | Groups | Functions | Distribution |
|-----------|---------|---|---------------------------------------|
| TPS-a | TPS-a-1 | Sesquiterpenes | Dicots |
| | TPS-a-2 | Sesquiterpenes | Monocots |
| TPS-b | | Monoterpenes, Isoprene | Angiosperms |
| TPS-c | | Diterpenes, copalyl synthase/Kaurene synthase, copalyl synthase | Land plants |
| TPS-d | TPS-d-1 | Sesquiterpenes, monoterpenes | Gymnosperms |
| | TPS-d-2 | Sesquiterpenes | Gymnosperms |
| | TPS-d-3 | Sesquiterpenes, Diterpenes | Gymnosperms |
| TPS-e/f | | Monoterpenes, Sesquiterpenes, Diterpenes, KS | Vascular plants |
| TPS-g | | Monoterpenes, Sesquiterpenes, Diterpenes | Angiosperms |
| TPS-h | | Putative bifunctional ^a Diterpenes | <i>Selaginella moellendorffii</i> |

^aBifunctional diterpenes synthase convert GGPP into CPP then in diterpene whereas monofunctional convert GGPP into diterpenes without CPP intermediate

Mechanisms of monoterpenes or sesquiterpenes formation have been widely studied and reviewed, as resumed by Croteau (1987) and Degenhardt *et al.* (2009). TPSs do not act as phosphatases that directly hydrolyse the inorganic pyrophosphate of the substrates, IPP and DMAPP. Instead, these enzymes produce intermediate reaction compounds giving a carbocation. This carbocation can be cyclized and then form another carbocation (Figure 7). These mechanisms allow cyclization of monoterpenes and allow also the formation of hydroxy groups on other carbon than those where the pyrophosphate was bound (in the case of GPP, the carbon number 1). If the reaction is not complete, the product is released from the enzymatic site in the form of an alcohol, like geraniol, linalool, or farnesol for example. Indeed, in *O. basilicum*, it has been shown that the hydroxy group of geraniol is added to the intermediate carbocation formed by a GES, not by a phosphatase (Iijima *et al.* 2004).

For sesquiterpenes, the numerous possibilities of carbocation formation and cyclisation between different carbon position lead to a high diversity of sesquiterpenes (Figure 7). The only cofactor used by TPSs are metal divalent ions (Mg^{2+} or Mn^{2+}) needed for the carbocation formation, because it allows organisation between the pyrophosphate group and the active site (Tarshis *et al.* 1994; 1996).

II-2. Other enzymes producing terpenes and terpenoids

The wide diversity of terpenes produced by TPSs can be increased by chemical modifications, named decorations, giving numerous terpenoids, probably several tens of thousands. By all the putative decorations catalyzed by enzymes, CYP450 (cytochrome P450) are probably the best example to present the diversity of product derived from terpenes. Beside the numbers of products raised by CYP450 an example can be the production of 8-hydroxy-geraniol by AtCYP76C4 by hydroxylation of geraniol or linalool oxides by AtCYP76C1 in flowers (Höfer *et al.* 2013; Boachon *et al.* 2015). CYP450 is one of the most extended gene family in plants that can modify terpenes, but it is not the only one. Continuing on geraniol modification, alcohol dehydrogenase (ADH) can produce geranial by oxidation of geraniol in *Zingiber officinale* (Lijima *et al.* 2014). Geraniol can also be acetylated via acetyltransferase in *R. x damascena* flowers by RhAAT1 producing geranyl acetate (Shalit *et al.* 2003). A more complex biosynthesis pathway of monoterpene in plant starting from geraniol is described in the orchid *Caladenia plicata* (Xu *et al.* 2017). Geraniol is produced by a classical GES (CpGES1), then, in three subsequent reactions, β -citronellol is produced. Firstly, geranial is formed by oxidation of geraniol by an ADH (CpADH3). A geranial reductase (GER1) reduces geranial into β -citronellal. Finally, CpADH3 produce β -citronellol by reduction of β -citronellal. Also, some other enzymes have a role in direct terpenes production such as CCDs that produce apocarotenoids. For example, in *S. lycopersicum*, cleavage of apo-8'-lycopenal, a lycopene derivative, by SICCD1B allows the production of neral and geranial in fruits (Ilg *et al.* 2014).

More recently, a new enzyme family has been discovered to be implicated in terpene production by dephosphorylation of prenyl-diphosphates. These enzymes are part of the Nudix hydrolase family.

II-3. The NUDX1 family

Nudix hydrolases are widely spread in all leaving organisms, from bacteria to eukaryotes. In the model plant *A. thaliana*, there are 28 *NUDX* genes (*AtNUDX1* being one of them), but in other plants, this number can vary, e.g. 32 in *S. lycopersicum* and 20 in *Oryza sativa* (Yoshimura and Shigeoka, 2015). In *A. thaliana*, *NUDX* genes are numbered from 1 to 27, except *AtDcp2* described before the *NUDX* analysis in this species (Yoshimura and Shigeoka, 2015). *AtDcp2* enzyme has an essential role in embryonic development by regulation of mRNA turnover. This enzyme hydrolyses capped mRNA by m^7 -GPPP to produce m^7 -GDP (Iwasaki *et al.* 2007; Gunawardana *et al.* 2007). Nudx hydrolases have a wide spectrum of substrates such as NADH, ADP-ribose, GDP-mannose, Malocyl-Coa, 8-oxo-(d)GTP... depending on the enzyme number. Each enzymes have their preferential substrate but not all the substrates are known. They all possess a Nudix motif (GX₅EX₇REUXEEXGU, where U is an aliphatic, hydrophobic residue) (Yoshimura and Shigeoka, 2015). This motif allows a nucleophilic substitution by water at the internal β -phosphate atom (Weber *et al.* 1992). Phosphate is bound to active site by hydrogen bound organised with divalent cations (Mg²⁺ or Mn²⁺) with the Nudix motif and the carbon moiety is surrounded by a hydrophobic amino acid (Liu *et al.* 2018; Jemth *et al.* 2019).

Beside the Nudx enzyme family with a catalytical phosphatase activity, some other proteins have a structure near from this enzyme without Nudix motif and so without phosphatase activity. As a

consequence, they are considered to belong to a super-family of Nudix enzymes. This is the case of isopentenyl-diphosphate isomerase (IDI) which can isomerise IPP into DMAPP. The opposite reaction is less efficient. IDI protonated IPP or DMAPP to produce a carbocation and then re-protonated the carbocation to produce the electrophil isomer of the initial isopentenyl use for the reaction (Street *et al.* 1994). In plants, IDI are localized in different organelles. The shortest isoform of *A. thaliana* AtIDI1 is targeted to the peroxisome and allow the production of DMAPP from the MVA pathway (Sapir-Mir *et al.* 2008). The longest isoform of AtIDI1 is targeted to plastids, and the AtIDI2 protein is targeted to mitochondria (Phillips *et al.* 2008).

Initially, AtNUDX1 was described to dephosphorylate 8-oxo-(d)GTP (Yoshimura *et al.* 2007). This oxidized ribonucleotide is produced by reactive oxygen species induced by numerous biotic and abiotic stress. It can pair with adenine and cytosine and thus incorporated into DNA or RNA, which enhance mutagenesis by replicational and translational errors. By dephosphorylation of this oxidized ribonucleotide, AtNUDX1 prevents such mutations and acts as a “cell sanitizer”. *AtNUDX1* is the homologue of *Escherichia coli* *MutT* (Yoshimura *et al.* 2007), and *Homo sapiens* *MTH1* (Gad *et al.* 2014).

In rose, Magnard *et al.* (2015) published for the first time the implication of a NUDX1 hydrolysing GPP to produce GP in rose petals, leading to geraniol by an uncharacterized phosphatase. The *RhNUDX1* gene is homologous to *AtNUDX1* and is highly expressed in rose petals of modern garden roses such as *R. x hybrida* cv. ‘Papa Meilland’. *RhNUDX1* shows a drastically reduce GPP hydrolase activity by mutation of the R₃₄A and Y₉₄A amino acids which are implicated in the coordination of divalent cations (Liu *et al.* 2018). Since this work, other NUDX1 have been implicated in terpenoid metabolism. In *Tanacetum cinerariifolium* a plastid-localized NUDX1 dephosphorylates *E*-chrysanthemyl-diphosphate (ChPP) into chrysanthemol, en route to production of pyrethrins (Li *et al.* 2020). In *A. thaliana*, AtNUDX1 and AtNUDX3 dephosphorylate IPP and DMAPP into IP and DMAP (Henry *et al.* 2018), but another enzyme, named isopentenyl-phosphate kinase, phosphorylates IP and DMAP into IPP and DMAPP (Henry *et al.* 2015 and 2018). AtNUDX1 is probably more implicated in IPP dephosphorylation than in 8-oxo-(d)GTP dephosphorylation because of the Km values (Henry *et al.* 2018; Jemth *et al.* 2019). This could be a regulation loop of the IPP and DMAPP homeostasy in cells.

In vitro, most of NUDX1 proteins can hydrolysed numerous substrates. Such as DMAPP, IPP, GPP, FPP, NPP, LPP... (Henry *et al.* 2018; Li *et al.* 2020), but the question about their effective role in plant remain unclear. For example, RcNUDX1-1a can dephosphorylates FPP into FP *in vitro*, but there is no farnesol in Old Blush petals (Sun *et al.* 2020). As a consequence, the *in vitro* activity of NUDX1 is not a proof of *in planta* activity.

III- Biosynthesis of terpene precursors in plants

III-1. The MVA and MEP pathways

In plants, IPP and DMAPP are produced by two different pathways, the MVA and the MEP pathways. The MVA pathway is found in all living organism and comprises six enzymatic steps starting with two acetyl-CoA. In 2001, Lichtenthaler summarized the discovery of the MVA pathway. In 1951, Lynen *et al.* demonstrated that acetyl-CoA is an intermediate of cholesterol biosynthesis. In 1956, Wolf *et al.* and Tavormina *et al.* described the mevalonate as an essential intermediate for cholesterol biosynthesis. Finally, in 1958, Lynen *et al.* and Chaykin *et al.* demonstrated that IPP is an intermediate of sterol biosynthesis. Konrad Bloch and Feodor Lynen obtained the Nobel Prize for this discovery in 1964. Since these works, biochemistry experiments have allowed to understand the organisation of the MVA pathway in numerous organisms such as mammals, yeast, bacteria and plants, without using cloning or heterologous expression methods. Nevertheless, since 1950 labelling experiments, and use of inhibitors of the MVA pathway, results suggested an alternative route in plants. For example, labelled mevalonolactone did not incorporate into carotenoids (Treharne *et al.* 1966), and uses of inhibitors of the MVA pathway decrease the production of sterols and ubiquinones, but not the production of plastoquinone and phyloquinone (Schindler *et al.* 1985). This type of results has led some authors to take an interest in other possible pathways and to discover the MEP pathway, also named non-mevalonate pathway or Rohmer's pathway (Disch *et al.* 1998; Rohmer *et al.* 1999; Lichtenthaler 1999). The MEP pathway include seven enzymatic steps starting from pyruvate and glyceraldehyde-3-phosphate. The cloned of the MEP pathway encoded the 1-deoxy-D-xylulose-5-phosphate synthase (DXS) from *Mentha x piperita* (Lange *et al.* 1998). Finally, in 2002 the complete elucidation of the MEP pathway was done in plants (Rodriguez-Concepcion and Boronat, 2002), and widely studied and reviewed (Hemmerlin *et al.* 2012; Vravenova *et al.* 2012; Nagegowda and Gupta 2020) (Figure 3).

The subcellular localization of enzyme in cells is an important key regulator for compounds biosynthesis because of substrate availability, multi enzyme complex formation, or localized chain reactions (reviewed by Bassard and Halkier, 2018). By enzyme targeting in different organelles such as plastid, mitochondria, nucleus, endoplasmic reticulum, some pathways can be active or not. The MVA pathway is described as being active in multiple sub-cellular compartments. First description of non-cytosolic enzyme in plant was in radish, in which 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) activity was observed in membrane rich fractions after gradient centrifugation of cell lysis (Bach *et al.* 1986). Further investigation has proved that HMGR is localized in endoplasmic reticulum membrane (review in Hemmerlin *et al.* 2012). In the protein sequence of HMGR, there are two or four transmembrane domains interspaced by a linker to the C-terminus of the protein with the catalytic domain oriented to the cytosol (Bach *et al.* 1999). Not all the MVA pathway is localized on membrane, some enzymes in the beginning of the pathway are localized in the cytosol or in peroxisomes. Indeed, in *A. thaliana* and *C. roseus*, phosphomevalonate kinase, mevalonate diphosphate decarboxylase, and an isoform of IDI are targeted to peroxisomes (Sapir-Mir *et al.* 2008; Simkin *et al.* 2011). These multiple sub-cellular localisations could indicate a channelling of intermediate reaction products. Unlike the MVA pathway, the MEP pathway is strictly localized in plastids (Kreuz and Kleinig, 1981; Heintze *et al.* 1990;

Lange *et al.* 1998; Hemmerlin *et al.* 2012). More precise subcellular localisation of enzymes of the MEP pathway was achieved in *C. roseus* chlorophyll-containing cells. First step of MEP pathway catalysed by DXS and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) are localized in plastid stroma, whereas final steps enzymes, 2-C-methyl-D-erythritol 4-phosphate synthase, 4-(cytidine 5'diphospho)-2-C-methyl-D-erythritol kinase, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase, and (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase are localized in stromules (Guirimand *et al.* 2020). This organelle specific localization can extend contact area between plastids and other compartments, as endoplasmic reticulum for example, for compound exchange and enzyme contacts.

III-2. Key regulators for IPP and DMAPP supply

IPP and DMAPP are at the crossroads of specialized and primary metabolites production. Their concentrations are highly regulated in the different cell compartments. We have already described the regulation loop of NUDX1 and IPK on IPP and DMAPP homeostasy, but other mechanisms of regulation have been described. Indeed, they are regulated both by transcriptional and post-transcriptional mechanisms in the MVA and MEP pathways, but also by crosstalk between these pathways.

All enzymes from the MEP pathway are transcriptionally regulated by light and DXS is the enzyme with the highest transcriptional regulation in the pathway (see review in Hemmerlin *et al.* 2012). DXS exists in two isoforms names classes 1 and 2. Class 1 DXS is used for primary metabolism as house keeping enzymes, and class 2 DXS are used for specialized metabolism. The expression of class 2 DXS genes is more regulated by stress, by organ specificity, or to modulate specialized metabolism in some tissues. For example, DXS is highly expressed in internal phloem parenchyma cells of *C. roseus* leaves, in which monoterpene indole alkaloids are synthesized (Burlat *et al.* 2004; Guirimand *et al.* 2020). DXS activity is also regulated by post-transcriptional mechanisms. For example, DXS activity is regulated by IPP concentration in *P. trichocarpa*. Adding IPP or DMAPP *in vitro* reduce activity by competition with thiamine-diphosphate, an essential cofactor (Banerjee *et al.* 2013). The medical fight against malaria has led to the testing of inhibitors of the MEP pathway in *Plasmodium falciparum*. Fosmidomycin is one of these inhibitors and is commonly used to inhibit DXR during laboratory experiments. This inhibition can be complemented by addition of 1-deoxy-D-xylulose (DOX). The mechanisms of inhibition of fosmidomycin mimics MEP, the reaction product of the DXR activity (Zeidler *et al.* 1998). For example, albino phenotype can be observed when fosmidomycin is applied to *A. thaliana* seedlings or to *S. lycopersicum* fruits. This phenotype is due to a reduction of carotenoids and chlorophyll biosynthesis in *A. thaliana* or lycopene biosynthesis in *S. lycopersicum* (Rodríguez-Concepción and Boronat, 2002).

In the MVA pathway the regulation of flux is done with the regulation of HMGR. Isoforms of HMGR can be differentially expressed in response to different stress, to different biological processes, or to organ specificity (see review in Hemmerlin *et al.* 2012). For example, *Ginkgo biloba* has three isoforms of HMGR. *GbHMGR2* and *GbHMGR3* are differentially expressed in tissues and respond to stress differentially. *GbHMGR2* is highly expressed 48 h after cold treatment compared to *GbHMGR3*. By contrast ethylene treatment up-regulated *GbHMGR3* four times more than *GbHMGR2*. The over

expression of *HMGR* expression is linked to an increase in terpene trilactone content derived from the MVA pathway (Rao *et al.* 2019). There are also proofs of feedback regulation. For example, over expression of *AtFPPS2* leads to a decrease in *AtHMGR* expression, whereas, reduced expression of *AtFPPS2* and reduced activity of *AtFPPS2* increase *AtHMGR* expression in *A. thaliana* (Keim *et al.* 2012). *HMGR* is also regulated by post-translational mechanisms such as phosphorylation / dephosphorylation. *AtHMGR* activity is negatively regulated by dephosphorylation induced by the Protein Phosphatase 2A (Leivar *et al.* 2011). Because of its key role in the regulation of MVA pathway in all organisms, *HMGR* has been targeted pharmaceutically to treat hypercholesterolemia in humans. Mevinolin, a compound of the statin family, is currently used to inhibit *HMGR* activity. Its inhibitor effect is due to its structure which looks like the HMG moiety of 3-hydroxy-3-methylglutaryl-coenzyme A. Thus, mevinolin can enter in the active site of *HMGR* and competitively inhibit the enzymatic reaction (Bach and Lichtenthaler, 1982 a; b).

Numerous reports have also confirmed a cross talk between the MEP and the MVA pathways. Indeed, it has been demonstrated that IPP, DMAPP, GPP, FPP, and GGPP can move across membranes of plastids and mitochondria (Hemmerlin *et al.* 2012). This phenomenon shows that there are compensations and regulations when the concentration of precursors decreases or increases in the cell organelles. Cross talk between the MVA and MEP pathways is thus an important way to regulate terpene precursors homeostasis. Unfortunately, no transporter has ever characterized in such transport.

Experiments on *Vitis vinifera* cell suspensions have established the basis of IPP import into plastids (Soler *et al.* 1993). In this work, IPP intake and saturation phenomenon respecting Michaelis kinetic was proposed to be an argument in favour of a transporter. Furthermore, intake of IPP was strongly inhibited by aminophenylethyl-diphosphate, a diphosphate analogue of IPP, but not by 2-phenylethanol (the same compound without phosphates). Moreover, protein-binding inhibitors such as glyceraldehyde 3-phosphate or DL-glyceraldehyde, reduce drastically respectively IPP import or IPP export into intact plastid (Bick and Lange, 2003; Wang *et al.* 2003; Yang *et al.* 2012). In another work, IPP was presumed to be transported into plastids by a uniporter transporter (Flügge and Gao, 2005). Overexpression of *AtIPK* into *Nicotiana benthamiana* leaves leads to an increase of plastid-synthesized monoterpenes such as linalool and β -ocimene. As cytosolic *AtIPK* increases the cytosolic IPP pool by phosphorylation of IP, a possible interpretation is that the excess of IPP was transported into plastids (Henry *et al.* 2015). Uptake of IPP into mitochondria for ubiquinones biosynthesis has also been indirectly prove by Lütke-Brinkhaus *et al.* (1984). Indeed, working with isolated mitochondria *in vitro*, they demonstrated that [2-¹⁴C]MVA-5-P was not incorporated into ubiquinone intermediates. They concluded that, *in planta*, IPP was imported into this organelle.

DMAPP, FPP, GPP can also cross the inner membrane of plastids with different rates, even if IPP does two times more quickly than the others (Bick and Lange, 2003). Over production of GPP in *S. lycopersicum* fruits led to a decrease of carotenoid concentration and an increase of monoterpenes concentration (Gutensohn *et al.* 2013). Indeed, the IPP and DMAPP pool of plastids, was probably used to produce GPP (thus monoterpenes) but not GGPP (thus carotenoids). Interestingly, in the same plant,

with or without over production of plastidial GPP, the over production of a bi-functional TPS into the cytosol increases the concentration of monoterpenes, proving that GPP was exported from plastids to cytosol (Davidovich-Rikanati *et al.* 2008; Gutensohn *et al.* 2013).

GGPP is also able to cross membranes. It is exported from plastids to cytosol and used by a geranylgeranyl transferase, belonging to the prenyl transferase family, to build geranyl-geranylated proteins. For example, a proof of export of plastidial GGPP was done on tobacco cells (Gerber *et al.* 2009), by inhibition of MEP pathway by fosmydomycine. In this experiment, cytosolic geranyl-geranylation of proteins was drastically reduced, proving that the MEP pathway was involved.

III-3. Active sites and functioning of IDSs

GPP, FPP, and GGPP are produced by IDS. These enzymes called GPPS, FPPS, and GGPPS (where S is for synthase activity) catalyse the polymerization of IPP and DMAPP, the end products of the MVA and MEP pathways. GPPS uses IPP and DMAPP to produce GPP, FPPS uses a second IPP to produce FPP, and GGPPS uses a third IPP to produce GGPP. GPP is the precursor of monoterpenoids, FPP, of steroids, cholesterol, farnesylated proteins, and sesquiterpenoids, and GGPP, of carotenoids, geranylgeranylated proteins, diterpenes, gibberellins, chlorophylls, and vitamin K1.

Most of the IDS act in *trans* (head-to-tail) to produce *E*-GPP, *E,E*-FPP or *E,E,E*-GGPP, which are precursors of the most abundant isoprenoids. Sequential condensation allows the elongation to C₁₀, C₁₅, C₂₀ products. IDSs start the reaction with one molecule of IPP (homoallylic substrate) and one molecule of DMAPP (allylic substrate) to produce GPP, FPPS can use a second IPP unit to produce FPP, and GGPPS a third IPP unit to produce GGPP. In the case of FPPS and GGPPS, firsts products (GPP and/or FPP) are conserved in the active site and become the allylic substrate during the sequential addition of IPP (Kellogg and Poulter 1997) (figure 8a).

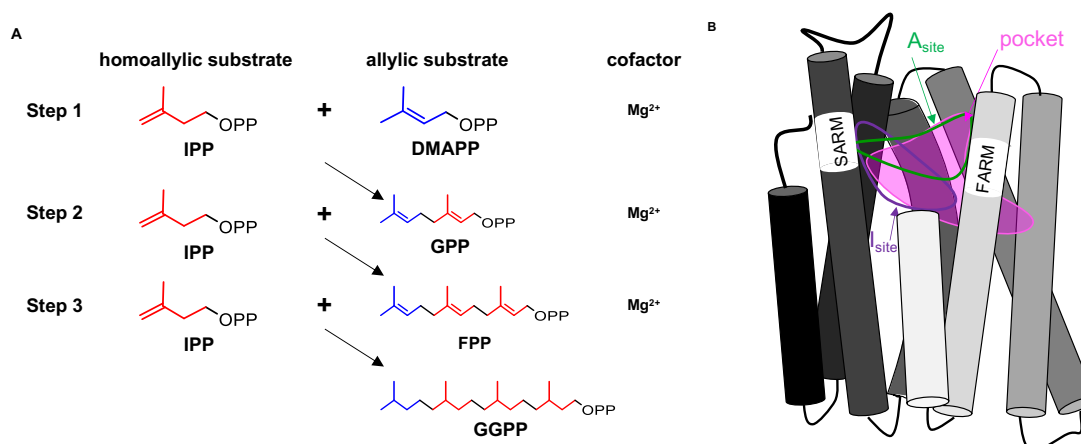


Figure 8 | mode of action of prenyl-diphosphate synthase. (A) sequential biosynthesis of prenyl diphosphate. Arrows indicate product formed for each step. Mg²⁺ as cofactor allow interaction between inorganic phosphate of substrate and amino acid from active site of enzyme. (B) schematic representation of one IDS sub-unit. For clarity some α helix are not presented. The elongation pocket is drawn in pink. The A-site corresponds to the place where allylic substrate binds and product carbon chain inserts. I site corresponds to the space where IPP binds. DMAPP, dimethylallyl diphosphate; FARM, first aspartate-rich motif; IPP, isopentenyl diphosphate; SARM, second aspartate-rich motif.

Some other IDSs have been identified and characterized to act in *cis* or *trans* (head-to-head or head-to-middle) allowing the production of *Z,Z*-FPP in *Solanum habrochaites* (Sallaud *et al.* 2009), lavandulyl-diphosphate in *Lavandula* (Demissie *et al.* 2013), neryl-diphosphate in *S. lycopersicum* (Schillmiller *et al.* 2009), chrisantemyl-diphosphate in *Chrysanthemum cinerariaefolium* (Rivera *et al.* 2001), and bornyl-diphosphate in *Salvia officinalis* and *Lavandula angustifolia* (Whittington *et al.* 2002; Despinasse *et al.* 2017). These irregular diphosphates are used by TPSs to produce irregular terpenes such as lavandulol, borneol, chrysanthemol (the precursor of pyrethrin), or *Z,Z*-farnesol for example.

The results of the sequential condensation of IPP and DMAPP of short-chain IDS is that they all produce GPP as final product or as intermediate reaction product. Thus, concerning the origin of cytosolic GPP in rose petals, we focused only on this class of enzyme and do not take interest in medium or long-chain IDS producing respectively C₃₀-C₃₅, C₄₀-C₅₀ prenyl-diphosphates, because they are not able to produce GPP as intermediates (Vandermoten *et al.* 2009; Wang *et al.* 2016).

Biosynthesis of prenyl-diphosphates by GPPS, FPPS or GGPPS, is performed by relatively identical enzymes. They have in common domains for substrate binding corresponding to the diphosphate substrate, such as IPP, DMAPP, GPP or FPP. These domains are called FARM and SARM (first and second aspartate-rich motif respectively). Amino acids in these conserved domains are essential to maintain substrates in the active site. Because aspartate is a polar amino acid, it interacts with divalent cation such as Mg²⁺. The recruitment of this cation helps to organise the diphosphate part of substrates in the active site and ensures its correct position (Ohnuma *et al.* 1996).

Regarding the mechanisms of sequential condensation of IPP to produce prenyl diphosphate the mechanism can be described in several steps (Figure 8b). Basically, after the first condensation of IPP and DMAPP to produce GPP, GPP move inside the active site and bind to the site where allylic substrate bind. This site is called the A site and is formed with the FARM and the SARM. In the next step, IPP bind to his binding zone name the I site (Hosfield *et al.* 2004). A second condensation step is involved in FPPS and GGPPS. Finally, for GGPPS, FPP produce by the second step of condensation move to the A site before last condensation of IPP. Substrate movement inside the active site needs space, and the deep of the hydrophobic elongation pocket regulates the final product length. An example of these mechanism can be explained with the case of FPPS. FPP carbon chain is too big to enter properly in the elongation pocket and the FPP cannot bind the A site. Finally, FPP is released from the active site because it cannot stay in the active site. This mechanism is called the flooring model (Wang *et al.* 2016). Product chain length is regulated by amino acids near to the elongation pocket. With the tertiary structure of protein, it as been proposed that a chain length determination domain is located four and five amino acids before the FARM (Tarshis *et al.* 1994). At these position amino acids can form some "floor" that can change the deep of the elongation pocket (Tarshis *et al.* 1994; 1996; Ohnuma *et al.* 1996).

During the different steps of IDS enzymatic cycle, numerous parts of proteins move to change the 3D structure of active site, to initiate each step in the right order (1st binding of allylic substrate, 2nd binding of IPP, 3rd condensation of both substrates). For example, the C-terminus of FPPS compose of charged amino acids that interact with IPP (Sond and Poulter, 1994) move after binding of IPP into I site to close

the active site and initiate the condensation of IPP and allylic substrate (Hosfield *et al.* 2004; Kavanagh *et al.* 2006; Park *et al.* 2017).

III-4. Dimeric structures of LSU and SSU

Short-chain IDS are essential for organisms, they can be easily identified by BLAST analysis on genomes. They are dimers of five different subunits, and can be classified into five subfamilies according to the way the subunits dimerize (Figure 9). The GPPS monomer can form a homodimer, named GPPS, which produce GPP and FPP. The FPPS monomer can also form a homodimer, named FPPS, which produces FPP. The large subunit (LSU) can form a homodimer, named GGPPS, which produces GGPP. The small subunit (SSU) can form a heterodimer with LSU, named GPPS, which mainly produces GPP. In recent works, authors distinguished a second small subunit (SSU-II) that can also form a heterodimer with LSU, named GPPS, but produces GPP and GGPP (Wang and Dixon, 2009). Other dimers are not functional.

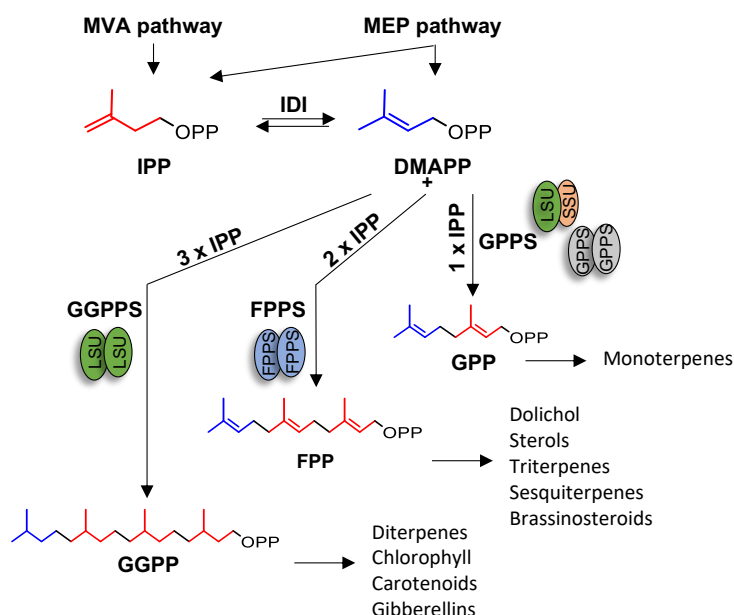


Figure 9 | biosynthesis of prenyl diphosphate in plant (adapted from Schmidt *et al.* 2009). Sub-unit of isoprenyl diphosphate synthase working in dimers are represented by different colors. IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GGPP, geranylgeranyl diphosphate; FPP, farnesyl diphosphate; GPP, geranyl diphosphate; LSU, large sub unit; SSU, small sub unit; FPPS, farnesyl diphosphate synthase; GPPS, homodimeric geranyl diphosphate synthase; IDI, Isopentenyl diphosphate isomerase

In plastid, LSU can form a homodimer, and then acts as a GGPPS (Dogbo and Camara, 1987), or can form a heterodimer with SSU, and acts as a GPPS (Burke *et al.* 1999). The activity of the homodimer is altered when the heterodimer is built. The result of the competition is in favor of GPPS. Concurrence between GPP and GGPP production occurs in plastid with the same source of IPP and DMAPP produced by the MEP pathway. Overexpression of SSU genes of *Lavandula x intermedia* or *A. majus* in *N. benthamiana* leads to a decrease in carotenoid, chlorophyll and gibberellin contents associated with a visible stunting phenotype (Orlova *et al.* 2010; Adal and Mahmoud 2020). These phenotypes confirm a decrease of GGPP availability in plastid when SSU is overexpressed. In *Humulus lupulus*,

producing lot of monoterpenes, *HtSSU* is highly expressed in trichomes (Wang and Dixon 2009). In *A. thaliana*, GPP production increased during anthesis by up-regulation of *AtSSU-II* expression. *AtSSU-II* can interact with *AtLSU* leading to GPP production and thus monoterpene production (Chen *et al.* 2015). In *A. majus* flowers, *AmSSU* expression is dependent of circadian cycle. *AmSSU* is also up regulated at protein level after anthesis and corelated with flower monoterpene production (Tholl *et al.* 2004). Regulation of SSU level is not only tissue-dependent, but also inducible by stress. For example, treatment with methyl jasmonate, used for stress simulation, enhance monoterpene indole alkaloid in *C. roseus*, *CrSSU* expression been seven times increased after 12 h (Rai *et al.* 2013). In *P. trichocarpa* two LSUs and one SSU-I are present in the genome. The heterodimer has been tested *in vitro* and [LSU1/SSU-I] or [LSU2/SSU-I] increases GPP production by four and three times respectively (Lackus *et al.* 2019). Furthermore, larvae-damage up-regulates *PtSSU-I* expression to increase herbivory-induced monoterpene in leaves by producing more GPP (Lackus *et al.* 2019). Thus, the regulation of SSU concentration in plastids is a way to tune the production of GPP and GGPPS in organs or during stress.

In another area, structural analyses have revealed that homodimer of SSU is inactive because of the lack of SARM. A conserved motif, CxxxC amino-acid (where x is a hydrophobic amino acid), is necessary for the interaction between LSU and SSU (Wang and Dixon 2009). By comparison of different LSUs and SSUs, they observed that one CxxxC is present on LSU, whereas two are present on SSU, allowing the interaction between proteins to produce heterodimer and enhance GPP biosynthesis. Product specify of the heterodimer can led to numerous options leading to an increase of GPP production with a conservation of GGPPS activity due to the LSU (Wang and Dixon 2009, Lackus *et al.* 2019), or completely abolish GGPP production (Tholl *et al.* 2004; Rai *et al.* 2013). These different productions are supposed to be type-dependent of SSU I or II (Barja *et al.* 2021). Furthermore, the heterodimer formation is not restricted to the species level and can be obtained by heterologous expression and chimerical heterodimer formation (Tholl *et al.* 2004; Wang and Dixon 2009). For example, heterologous expression of SSU from *A. majus* or *L. x intermedia* gives heterodimers in *N. benthamiana* (Orlova *et al.* 2009; Adal and Mahmoud, 2019).

Crystal structure of the heterodimer of GGPPS in *M. piperita* has revealed that it is in fact a heterotetramer including two heterodimers of LSU and SSU (Chang *et al.* 2010). Compared to other IDS homodimers, where the elongation pocket regulates the product chain length (flooring model), the authors explained the production of GPP of the heterotetramer by a two-chamber model. In this model, no physical clump is formed in the elongation pocket. Indeed, when SSU interacts with LSU, it changes the movement of a highly mobile part name active site cavity loop 2, found on the top of the active site of LSU. This part normally protects the allylic substrate from solvent during the catalytic reaction. Reducing the movement of active site cavity loop 2 enhance the formation of a “chamber” and let the active site in an “open” conformation. By reducing efficiency of FPP formation during the second step of GGPP biosynthesis, activity of the heterotetramer produces GPP because it leaves the active site by solvation before elongation into FPP or GGPP. This explains the decrease of GGPPS activity in favor of GGPPS activity. Because it is not a physical clump, this reduction of activity does not abolish completely the GGPPS

activity as in flooring model, and explains the residual GGPPS activity of the heterotetramer. In addition, directed mutagenesis on LSU reduces the movement of the active site cavity loop 2 and abolishes the GGPPS activity when the heterodimer is produced, leading to a heterotetramer producing only GPP (Hsieh *et al.* 2011).

III-5. Homodimeric structures of IDSs

Not all angiosperm species contain heteromeric GPPS protein but some of them have homomeric GPPS. The two monomers of GPPS have not the Cxxx motif allowing interaction with LSU. They have been identified and characterized in some species like *C. roseus* (Rai *et al.* 2013), *Picea abies* (Schmidt and Gershenzon, 2008) *S. lycopersicum* (van Schie, 2007), *Chimonanthus praecox* (Kamram *et al.* 2020) but also in monocotyledon as the orchid in *Phalaenopsis bellina* (Hsiao *et al.* 2008). These homodimeric GPPS can produce GPP, and thus monoterpene. For example, the expression pattern of *PbGGPPS* correlates with monoterpene production in flowers of different species (*P. bellina* versus *P. equestris*) (Hsiao *et al.* 2008). Recombinant *P. abies* homomeric GPPS (PaIDS2) produces only GPP *in vitro* with DMAPP and IPP (Schmidt and Gershenzon, 2008; Schmidt *et al.* 2010). In planta, the protein is predicted to be localized in plastids, and the gene responds to methyl jasmonate treatment: when it is over-expressed, it correlates with oleoresin production including monoterpenes.

Homodimeric form of GPPS are not only implicated in monoterpene production, but also in ubiquinone or gibberellin production some species. For example, in *S. lycopersicum esculentum*, VIGs experiment targeting the homodimeric *LeGGPPS* decreases its expression and reduce gibberellin content (Van Schie *et al.* 2007). The Norway spruce (*P. abies*) homodimeric GPPS (PaIDS3) which have a GPPS/GGPPS activity is predicted to be in mitochondria and is probably also implicated in ubiquinone formation (Schmidt and Gershenzon, 2008).

Finally, the case of the *A. thaliana* homomeric GPPS which is both localized in plastid and mitochondria has been widely studied and first described as GPPS (Bouvier *et al.* 2000) but later as polyprenyl-diphosphate synthase (Ducluzeau *et al.* 2012). This was confirmed because AtGGPPS can restore ubiquinone biosynthesis in yeast hexaprenyl-diphosphate synthase mutant (Ducluzeau *et al.* 2012). The final conclusion of these study is the product specificity of AtGGPPS dependent on IPP / allylic substrate ratio (Bouvier *et al.* 2000; Hsieh *et al.* 2011).

FPPSs have also been study in different organisms (archaea, bacteria, yeast, fungi, mammalian, plants) because of their essential role in sterol biosynthesis. Their structures have been discovered thanks to the study of mode of action of diverse FPPS inhibitors used as anti-cholesterol in humans. The first FPPS crystal structure was determined in avian (*Gallus gallus*) and is composed of 10 alpha helices (named A to J). To be active, FPPS needs to form a homodimer, because a part of the end of the elongation pocket is formed by the second FPPS monomer (Tarshis *et al.* 1994). Many studies have tried to reduce the steric hindrance of amino acid from the chain length determination domain to increase length of hydrophobic pocket and allow longer product formation (C_{20} , C_{25}) (Tarshis *et al.* 1994, 1996, Ohnuma *et al.* 1996). On the contrary, increasing the steric hindrance of these amino acids reduces the product carbon chain length (Narita *et al.* 1999; Stanley-Fernandez *et al.* 2000). Many works also have

reported impact on activity and/or product length specificity when some other amino acids are changed, probably because of change in 3D structure dynamics and substrate positioning. Indeed, conformation changes during the catalytic cycle such as C-terminus movement or some shift of loops near to the active site are needed to close the active site and protect the substrate from solvent (Gabelli *et al.* 2005; Kavanagh *et al.* 2006; Rondeau *et al.* 2006; Park *et al.* 2016).

FPPS are mostly localised in cytosol except CrFPPS which is targeted into peroxisome (Thabet *et al.* 2011). In lot of species, there are many copies of FPPS in genomes, and sometimes several isoforms. For example, in *A. thaliana*, *AtFPPS1* gene encodes two isoforms: one short producing a protein targeted to cytosol, and one long producing a protein targeted to mitochondria (Cunillera *et al.* 1997; Manzano *et al.* 2006).

FPPS can also be regulated at the gene expression level at the organ level or during stresses. For example, *AtFPPS2* is only highly expressed in petals of *A. thaliana* (Cunillera *et al.* 1996; Boachon *et al.* 2019), and HIFPPS in secretory glands of *H. lupulus* (Okada *et al.* 2001). In *Zea mays*, *ZmFPPS3* is up-regulated by wounding and elicitor treatment. This regulation allows substrate formation for sesquiterpene synthases, and thus produce sesquiterpenes for plant defence (Richter *et al.* 2015). This is also the case in *Withania somnifera*, in which *WsFPPS* up-regulation is described after methyl jasmonate or salicylic acid treatments, or even after mechanical wounding of leaves (Gupta *et al.* 2011). FPPS activity can also be regulated by FPP, DMAP or (Henry *et al.* 2015; Park *et al.* 2016).

In summary, terpenes production is highly regulated in plants both by transcriptional and post-transcriptional mechanisms of the MEP and MVA pathways, by a regulation loop of NUDX1 and IPK on IPP and IP, but also by reaction channelling and crosstalks between organelles. It can be regulated by organ specificity, and biotic and abiotic stresses. In such a network of regulations, the appearance of a new enzymatic function necessarily modifies the production of other essential molecules and possibly the function of other enzymes. In this work, we wanted to describe the biochemical adaptation by describing the evolution of the *NUDX1* gene family in *Rosaceae* and searching for the enzyme linked to a new production of cytosolic GPP. We thus obtained a putative scenario of functionality and evolution of *NUDX1* and *FPPS1* genes and enzymes.

General outline chapter 1:

Article 1 “**Functional diversification in the *Nudix hydrolase* gene family drives sesquiterpene biosynthesis in *Rosa x wichurana*”**”

Pulu SUN, Clément DEGUT, Stéphane RETY, Jean-Claude CAISSARD, Laurence HIBRAND-SAINT OYANT, Aurélie BONY, Saretta N. PARAMITA, Corentin CONART, Jean-Louis MAGNARD, Julien JEAUFFRE, Ahmed M. ABD-EL-HALIEM, Jordan MARIE-MAGDELAINE, Tatiana THOUROUDE, Raymonde BALTENWECK, Carine TISNÉ, Fabrice FOUCHER, Michel HARING, Philippe HUGUENEY, Robert C. SCHUURINK, Sylvie BAUDINO

The Plant Journal, Volume104, Issue1, September 2020, Pages 185-199

<https://doi.org/10.1111/tpj.14916>

Published: 08 July 2020

Since the discovery and the publication of NUDX1 implication in geraniol biosynthesis in rose petals by dephosphorylation of GPP into GP (Magnard *et al.* 2015), numerous *NUDX1* sequences have been cloned and sequenced in the BVpam laboratory. However, little was known about these sequences and their potential functions in rose scent. Thanks to a cross between OB and a hybrid of *R. x wichurana*, QTL analysis have shown that NUDX1 can also participate to (*E,E*)-farnesol biosynthesis. In this article 1, the biosynthesis activities of NUDX1-1a and NUDX1-2 (we will find it to be encoded by *NUDX1-2c* in Article 2) are discussed according to mendelian segregation of the gene and acyclic terpenes, and to structural biochemistry of the active site. Indeed, the expressed *RwNUDX1-2* was characterized, and investigations about protein sequence was performed to understand substrate specify for GPP or FPP. One major conclusion of Article 1 was that *NUDX1-1a* was expressed in OB petals and was involved in the production of geraniol, and that *NUDX1-2* was expressed in *R. x wichurana* petals and was involved in the production of (*E,E*)-farnesol.

In this work, I did the sequencing of *NUDX1* gDNAs including the introns, I have compared these sequences with that of the reference genomes (GDR, Jung *et al.* 2019), and I have assigned them to the different Nudx1 clades in *Rosa*.

Functional diversification in the *Nudix hydrolase* gene family drives sesquiterpene biosynthesis in *Rosa* × *wichurana*

Pulu Sun^{1,2} , Clément Dégut³ , Stéphane Réty⁴ , Jean-Claude Caissard¹ , Laurence Hibrand-Saint Oyant⁵ , Aurélie Bony¹, Saretta N. Paramita¹, Corentin Conart¹, Jean-Louis Magnard¹, Julien Jeauffre⁵, Ahmed M. Abd-El-Halim² , Jordan Marie-Magdelaine⁵, Tatiana Thouroude⁵, Raymonde Baltenweck⁶, Carine Tisné³ , Fabrice Foucher⁵ , Michel Haring², Philippe Hugueney⁶ , Robert C. Schuurink² , and Sylvie Baudino^{1*} 

¹Univ Lyon, UJM-Saint-Etienne, CNRS, BVPam FRE 3727, Saint-Etienne F-42023, France,

²Green Life Sciences Research Cluster, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, Amsterdam 1098 XH, The Netherlands,

³Expression Génétique Microbienne, UMR 8261, CNRS, Université de Paris, Institut de Biologie Physico-Chimique (IBPC), Paris 75005, France,

⁴Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, LBMC, 46 Allée d'Italie Site Jacques Monod, Lyon F-69007, France,

⁵IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, Beaucauzé 49071, France, and

⁶Université de Strasbourg, INRAE, SVQV UMR-A 1131, Colmar F-68000, France

Received 23 April 2020; revised 3 June 2020; accepted 24 June 2020; published online 8 July 2020.

*For correspondence (e-mail sylvie.baudino@univ-st-etienne.fr).

SUMMARY

Roses use a non-canonical pathway involving a *Nudix* hydrolase, RhNUDX1, to synthesize their monoterpenes, especially geraniol. Here we report the characterization of another expressed *NUDX1* gene from the rose cultivar *Rosa x wichurana*, *RwNUDX1-2*. In order to study the function of the *RwNUDX1-2* protein, we analyzed the volatile profiles of an F₁ progeny generated by crossing *R. chinensis* cv. 'Old Blush' with *R. x wichurana*. A correlation test of the volatilomes with gene expression data revealed that *RwNUDX1-2* is involved in the biosynthesis of a group of sesquiterpenoids, especially *E,E*-farnesol, in addition to other sesquiterpenes. *In vitro* enzyme assays and heterologous *in planta* functional characterization of the *RwNUDX1-2* gene corroborated this result. A quantitative trait locus (QTL) analysis was performed using the data of *E,E*-farnesol contents in the progeny and a genetic map was constructed based on gene markers. The *RwNUDX1-2* gene co-localized with the QTL for *E,E*-farnesol content, thereby confirming its function in sesquiterpenoid biosynthesis in *R. x wichurana*. Finally, in order to understand the structural bases for the substrate specificity of rose *NUDX* proteins, the RhNUDX1 protein was crystallized, and its structure was refined to 1.7 Å. By molecular modeling of different rose *NUDX1* protein complexes with their respective substrates, a structural basis for substrate discrimination by rose *NUDX1* proteins is proposed.

Keywords: *Nudix* hydrolase, rose scent, sesquiterpenes, farnesol, volatile compounds, quantitative trait locus, *Nudix* structure.

INTRODUCTION

Rose is one of the most economically important flowers, with thousands of cultivars mainly used as cut flowers, as garden ornamentals, and for the perfume industry. Fragrance is a very important rose trait that contributes to its commercial value, beside flower shape and petal color (Smulders *et al.*, 2019). Hundreds of rose volatile compounds have been identified so far, including terpenoids, phenylpropanoids, and lipid-derived volatiles (Shalit *et al.*, 2004). Many of the rose volatiles are commonly used in the

perfume and cosmetic industry (Schwab *et al.*, 2008). For example, geraniol, one of the major monoterpene alcohols (C₁₀ terpenoid) from rose petals, is responsible for their sweet floral rose smell (Chen and Viljoen, 2010). In many plant species, monoterpene biosynthesis relies on plastid-localized terpene synthases (TPSs) such as geraniol synthase (GES) (Iijima *et al.*, 2004; Yang *et al.*, 2005; Ito and Honda, 2007; Masumoto *et al.*, 2010; Dong *et al.*, 2013; Simkin *et al.*, 2013). However, no TPS with GES activity has been characterized in roses to date. Instead, rose flowers

use an alternative TPS-independent pathway to produce geraniol, involving a diphosphohydrolase belonging to the Nudix enzyme family. The cytosolic Nudix hydrolase RhNUDX1 converts geranyl diphosphate into geranyl monophosphate, which is then hydrolyzed to geraniol by petal-derived phosphatase activity (Magnard *et al.*, 2015).

Nudix hydrolases have been identified in many species, including archaea, bacteria, eukaryotes, and viruses (Gunawardana *et al.*, 2009). They were originally defined as housecleaning enzymes, eliminating toxic metabolites from the cells (Bessman *et al.*, 1996). Nudix hydrolases constitute a superfamily of pyrophosphatases catalyzing the hydrolysis of nucleoside diphosphates linked to different X moieties (Bessman *et al.*, 1996). Known substrates of Nudix hydrolases include nucleoside triphosphates (dNTPs) and their oxidized derivatives, such as 7,8-dihydro-8-oxo-deoxyguanosine triphosphate (8-oxo-dGTP), nucleotide sugars and alcohols, dinucleoside polyphosphates, dinucleotide coenzymes, and capped mRNAs (McLennan, 2006). Members of the Nudix hydrolase superfamily all share a conserved Nudix box. This Nudix motif is formed by a loop- α helix-loop structure, and provides binding sites for divalent cations (usually Mg^{2+} or Mn^{2+}) that play a crucial role in catalysis. The presence of these cations is required for the activity of the Nudix proteins. The substrate specificity and the catalytic reaction mechanism are also determined by regions outside of the Nudix motif (McLennan, 2006; Gunawardana *et al.*, 2009). For example, enzymes acting on Coenzyme A share a motif that is located outside of the Nudix box and is involved in substrate recognition (Kupke *et al.*, 2009). These regions, together with the Nudix motif, form an $\alpha/\beta/\alpha$ sandwich structure, which is also known as the Nudix fold (Mildvan *et al.*, 2005). This fold is shared by the isopentenyl diphosphate isomerases, which, together with Nudix hydrolases and other proteins, form a larger group. This group was previously called the Nudix suprafamily (McLennan, 2006), but is sometimes also referred to as the Nudix superfamily or the Nudix homology clan (Srouji *et al.*, 2017).

The functions of most Nudix hydrolases remain unclear in plants. For example, Nudix hydrolase 1 from *Arabidopsis thaliana* (AtNUDX1) was first described as an NADH pyrophosphatase (Dobrzanska *et al.*, 2002). Under specific physiological conditions, AtNUDX1 was later shown to be involved in folate biosynthesis, using dihydroneopterin triphosphate as a substrate (Klaus *et al.*, 2005). AtNUDX1 was also proposed to be involved in the elimination of harmful nucleoside- and deoxynucleoside-triphosphate derivatives such as 8-oxo-dGTP, similar to the MutT protein from *Escherichia coli* (Ogawa *et al.*, 2005; Yoshimura *et al.*, 2007). Recently, AtNUDX1 was shown to play a role in the regulation of terpene biosynthesis, by acting on isopentenyl diphosphate (IPP) precursors (Henry *et al.*, 2018).

The rose Nudix hydrolase RhNUDX1 has been shown to be involved in the biosynthesis of the monoterpene geraniol, which is a major petal volatile in many scented roses (Magnard *et al.*, 2015). However, besides monoterpenols, the petals of some rose species, such as *Rosa x wichurana*, emit significant amounts of sesquiterpenes (Roccia *et al.*, 2019). Using a combination of molecular, genetic, and structural approaches, we characterized a novel NUDX1 protein from *R. x wichurana*, RwnUDX1-2. Unlike RhNUDX1, which is involved in monoterpene biosynthesis, RwnUDX1-2 uses farnesyl diphosphate (FPP) to produce farnesyl monophosphate (FP) as a precursor of *E,E*-farnesol and other sesquiterpenoids. Furthermore, we solved the crystal structure of the RhNUDX1 protein and used molecular modeling to provide structural bases for explaining the different substrate specificities of NUDX1 enzymes from different rose cultivars and from *A. thaliana*.

RESULTS

NUDX1 genes in rose belong to a complex gene family organized in three clades: NUDX1-1 to NUDX1-3

In order to further investigate the functions of Nudix hydrolases 1 in roses, we generated RNA-seq reads from four different rose cultivars: *R. chinensis* cv. 'Old Blush' (OB), *R. x wichurana* (Rw), and two individuals (OW9035 and OW9047) obtained from a crossing between OB and Rw. Next, we pooled the generated reads together, used them to assemble a transcriptome *de novo*, and searched it to identify independent NUDX1 mRNA transcripts. Subsequently, we used the generated transcriptome as a reference to map the RNA-seq reads from each sample and to quantify gene expression. Four sequences were retrieved that were annotated as NUDX1 (Table S1). Among these sequences, one corresponded to the sequence previously characterized from *R. x hybrida* cv. 'Papa Meiland' (PM, RhNUDX1 in Magnard *et al.*, 2015) and was highly expressed in the rose petals from OB (fragments per kilobase of transcript per million fragments mapped reads [FPKM]: 9347.2) and OW9047 (FPKM: 6687.3). We named it NUDX1-1. Another sequence was highly expressed in Rw (FPKM: 1433.0) and OW9035 (FPKM: 1332.6), but differed from the first sequence and was named NUDX1-2. It was noticeable that NUDX1-1 and NUDX1-2 were not highly expressed together in these four cultivars (Table S1). Two other sequences were partial and corresponded to another NUDX1 (NUDX1-3) but were weakly expressed in the rose petals of the four individuals, as judged by the read counts in the RNA-seq data. A search in the recently published genomes was used to retrieve all NUDX1 sequences in OB (Hibrand Saint-Oyant *et al.*, 2018; Raymond *et al.*, 2018). NUDX1-1 was present in five highly similar copies on chromosome 2 of the homozygous genomes (collectively named *RcNUDX1-1a*, see Figure S1) and one less similar

copy on chromosome 4 (*RcNUDX1-1b*). *NUDX1-2* was present in two copies in the homozygous genomes (*RcNUDX1-2a* and *RcNUDX1-2b*). A search in the heterozygous genome allowed us to detect a third copy (*RcNUDX1-2c*), missing in both homozygous genomes. *NUDX1-3* was present in one copy in the homozygous genomes (*RcNUDX1-3*). Using genomic DNA from *Rw* as template, several corresponding sequences were amplified and named *RwNUDX1-1*, *RwNUDX1-2a*, *RwNUDX1-2b*, *RwNUDX1-2c*, *RwNUDX1-2c'*, and *RwNUDX1-3* (see Table S2 for primers). All *NUDX1* genes were predicted to encode proteins of around 150 amino acids that contained the characteristic Nudix box (Figure 1a) (Bessman *et al.*, 1996). Percentages of identity of these proteins are presented in Table S3. *RwNUDX1-1* was identical to *RcNUDX1-1b*, *RwNUDX1-3* was identical to *RcNUDX1-3*,

and *RwNUDX1-2c* and *c'* were very similar to *RcNUDX1-2c*. Alignment of rose *NUDX1* protein sequences with *AtNUDX1* and building of a phylogenetic tree confirmed that the sequences could be separated into three well-supported clades, named *NUDX1-1*, *NUDX1-2*, and *NUDX1-3* (Figure 1b). *RhNUDX1*, *RcNUDX1-1a*, *RcNUDX1-1b*, and *RwNUDX1-1* are closely grouped together, as are *RcNUDX1-2a*, *b*, and *c* and *RwNUDX1-2a*, *b*, *c*, and *c'*. *RcNUDX1-3* and *RwNUDX1-3* are in another cluster, to which *AtNUDX1* is the closest. As *NUDX1-1* and *NUDX1-2* were highly expressed in petals (Table S1), we decided to focus on these two *NUDX1* sequences.

***NUDX1-1a* expression is correlated with the production of geraniol and other monoterpenoids while *NUDX1-2* expression is correlated with the production of *E,E*-farnesol and other sesquiterpenoid compounds**

OB and *Rw* have distinct scent profiles: at the open flower stage (stage four according to Bergougnoux *et al.*, 2007), *OB* produced mainly 1,3,5-trimethoxybenzene, geraniol, dihydro- β -ionol, and germacrene D, while *Rw* was rich in 2-phenylethanol (2PE) and *E,E*-farnesol (Table 1). In order to study the potential functions of *NUDX1* genes, the expressions of *NUDX1-1* and *NUDX1-2* were analyzed in *OB* and *Rw* and in an *F*₁ progeny from crosses between *OB* and *Rw* genotypes. This mapping population (*OW*) consists of a full-sib family of 151 hybrids (Hibrand Saint-Oyant *et al.*, 2018). A small subset of this rose population, 18 out of 151 rose plants, including the parents *OB* and *Rw*, was subjected to quantitative reverse transcriptase-PCR (qRT-PCR) to determine the transcript levels of *NUDX1-1* and *NUDX1-2* across the *OW* population at stage 4. It was already known that *RhNUDX1* was responsible for the production of geraniol in modern roses (Magnard *et al.*, 2015). The *NUDX1-1* gene was highly expressed in *OB* and in 10 hybrids of the progeny (*OW9007*, *OW9011*, *OW9013*, *OW9021*, *OW9024*, *OW9047*, *OW9074*, *OW9099*, *OW9149*, and *OW9204*), while *NUDX1-2* was expressed in *Rw* and in 10 hybrids of the progeny (*OW9007*, *OW9013*, *OW9018*, *OW9021*, *OW9024*, *OW9035*, *OW9037*, *OW9082*, *OW9099*, and *OW9149*) (Figure 2a). Sequencing of the PCR products showed that only *RcNUDX1-1a* and *RwNUDX1-2c* were expressed. No correlation was found between the expression of *NUDX1-1* and *NUDX1-2*, suggesting that they have independent functions. Indeed, *NUDX1-1* and *NUDX1-2* were expressed independently in different individuals, or together as, for example, in *OW9099* and others. In addition, the expression levels of *NUDX1-2* were generally lower than the expression levels of *NUDX1-1* (Wilcoxon rank-sum test, *P* < 0.01). Differential expression of *NUDX1-2* across the hybrids, with a known volatile-associated function of *RhNUDX1* (Magnard *et al.*, 2015), could indicate that *NUDX1-2* was also associated with the production of one or more volatiles in rose.

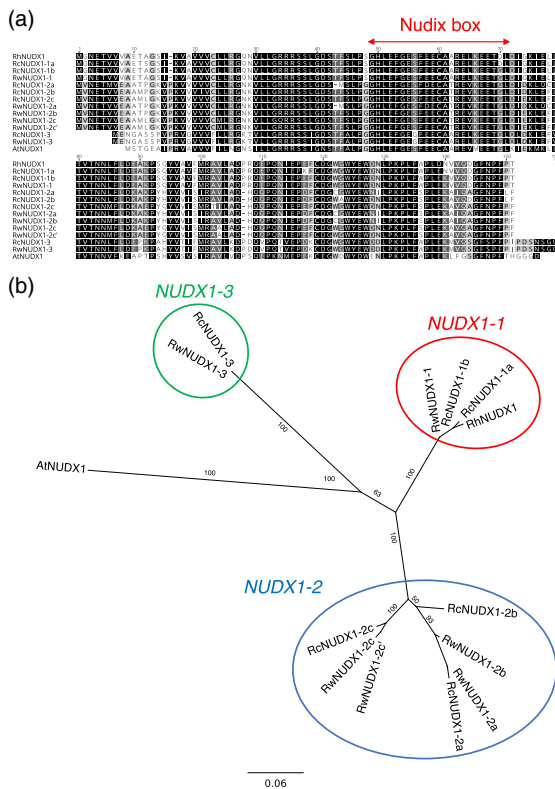


Figure 1. (a) Alignment of amino acid sequences of *R. x hybrida* cv. 'Papa Meiland' (*RhNUDX1*), *R. chinensis* cv. 'Old Blush' (*RcNUDX1-1a*, *RcNUDX1-1b*, *RcNUDX1-2a*, *RcNUDX1-2b*, *RcNUDX1-2c*, and *RcNUDX1-3*), *R. x wichurana* (*RwNUDX1-1*, *RwNUDX1-2a*, *RwNUDX1-2b*, *RwNUDX1-2c*, *RwNUDX1-2c'*, and *RwNUDX1-3*), and *A. thaliana* (*AtNUDX1*). The Nudix box corresponding to the consensus sequence Gx5Ex7REUxEExGU (x, any amino acid; U, bulky hydrophobic amino acid, normally Ile, Leu, or Val) is indicated. (b) Unrooted neighbor joining tree displaying the relationships of rose *NUDX1* proteins with *AtNUDX1* from *A. thaliana* (*AtNUDX1*). Alignment and tree were constructed using Geneious® software (version 10.2.3, Biomatters, Auckland, New Zealand), and the bootstrap values are shown as a percentage from 1000 bootstraps replicates.

188 Pulu Sun et al.

Table 1 Major volatile compounds extracted from petals of *R. chinensis* cv. 'Old Blush' (OB) and *R. x wichurana* (Rw) at fully open flower stage (stage four) and analyzed by GC-MS

| Compounds | Cultivar | |
|---------------------|-------------------|-------|
| | OB | Rw |
| Z-3-Hexenyl acetate | 5.55 ^a | nd |
| E-2-Hexenal | 1.21 | 0.48 |
| Z-3-Hexenol | nd | 0.77 |
| Geraniol | 10.96 | nd |
| Geranial | 4.43 | nd |
| Germacrene D | 8.42 | nd |
| δ-Cadinene | 1.24 | nd |
| E-β-Farnesene | nd | 0.75 |
| E,E-Farnesol | nd | 2.17 |
| Farnesyl acetate | nd | 0.83 |
| 2PE | nd | 73.04 |
| 2PEA | nd | 0.53 |
| TMB | 27.71 | nd |
| Dihydro-β-ionol | 9.90 | nd |

TMB, 1,3,5-trimethoxybenzene; 2PE, 2-phenylethanol; 2PEA, 2-phenylethyl acetate; nd, not detected.

^aValues represent the relative proportion of the total amount (average value of 7–9 different replicates from both 2014 and 2015).

To study this hypothesis, the correlation between the expression levels of the *NUDX1* genes and the volatile profiles of these 16 selected rose hybrids and their parents was analyzed. Spearman's rank correlation test was performed on the expression data of *NUDX1-1* and *NUDX1-2* with the volatile compounds data from the 18 samples. GC-MS analysis results and correlations are shown in Data S1. The volatile compounds that showed significant correlation with either *NUDX1-1* or *NUDX1-2* expression levels are presented in Figure 2b. *NUDX1-1* expression had positive correlation with the following monoterpenes: neral ($P \leq 0.001$), geranial ($P \leq 0.001$), β-myrcene ($P \leq 0.001$), geraniol ($P \leq 0.001$), Z-β-ocimene ($P \leq 0.001$), E-β-ocimene ($P \leq 0.01$), and limonene ($P \leq 0.05$). *NUDX1-2* expression showed positive correlation with the following sesquiterpenes: E,E-farnesol ($P \leq 0.0001$), E,E-farnesal ($P \leq 0.001$), E-β-farnesene ($P \leq 0.001$), E,E-α-farnesene ($P \leq 0.001$), Z,E-α-farnesene ($P \leq 0.001$), E-nerolidol ($P \leq 0.01$), allorfarnesene ($P \leq 0.01$), α-bisabolene ($P \leq 0.05$), and farnesyl acetate ($P \leq 0.05$). The strongest correlation was observed with E,E-farnesol (Data S1). This result showed that *NUDX1-2* is potentially involved in the production of sesquiterpenoids in rose, in particular E,E-farnesol.

Quantitative trait locus (QTL) mapping for E,E-farnesol content and mapping of the *NUDX1-2* gene

We detected a major QTL for geraniol biosynthesis in the OW progeny on the female linkage group 2, which colocalized with *RhNUDX1* (Magnard et al., 2015). In order to further support the hypothesis that *RwNUDX1-2* was

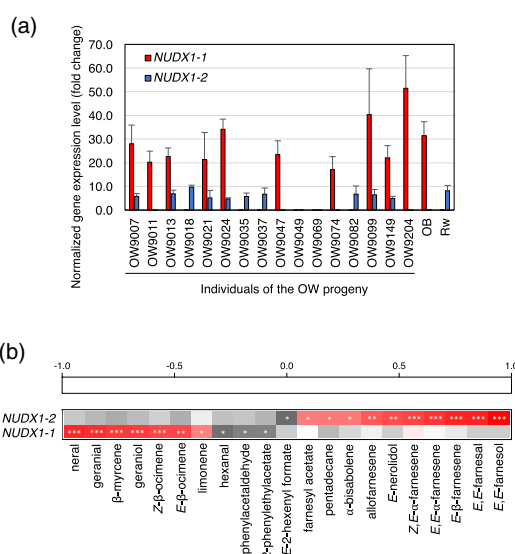


Figure 2. (a) Gene expression levels of *NUDX1-1* and *NUDX1-2* in 16 F₁ individuals from a cross between *R. chinensis* cv. 'Old Blush' (OB) and *R. x wichurana* (Rw) and their parents. Transcript levels of *NUDX1-1* and *NUDX1-2* were normalized to three housekeeping genes coding for α-tubulin, elongation factor 1-α, and translationally controlled tumor protein according to (Dubois et al., 2012). Error bars indicate standard deviation (SD) obtained from two independent biological replicates with at least two technical replicates each. A significant difference was found between the medians of the expression levels of *NUDX1-1a* and *NUDX1-2* (Wilcoxon rank-sum test, $P < 0.01$). (b) Heatmap correlation (Spearman's rank correlation from -1 to +1) of *NUDX1-1* and *NUDX1-2* expressions with some of the volatile scent compounds that showed significant correlation. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$. The complete correlation heatmap is available in Data S1.

involved in the production of E,E-farnesol and other sesquiterpenoids, we analyzed the distribution of the volatiles produced in petals of all the progeny of the cross between OB and Rw. Petals from 148 and 132 individuals of the progeny were collected in 2014 and 2015, respectively. Petals were subjected to hexane extraction and extracts were analyzed by GC-MS (Data S2). In total, 104 compounds were identified among 153 individuals (including the parents of the population, OB and Rw), but only 100 compounds were recovered in the samples collected in 2015. The compounds with the highest correlation coefficient with *NUDX1-2* expression (Figure 2b, E,E-farnesol, E,E-farnesal, E-β-farnesene, E,E-α-farnesene, and Z,E-α-farnesene) were chosen for further analyses. E,E-farnesol and other sesquiterpenoid contents appeared to segregate in the progeny (Figure S2). Due to the non-normal distribution of the variance residue of data, a log transformation was applied to make the variance residue distribution normal. The raw data were analyzed first by the non-parametric Kruskal–Wallis rank-sum test (KW) (Table S4). Interval mapping analysis was then performed on the

log-transformed data, for the chromosome regions on which QTLs were detected by KW both in 2014 and 2015 (Table S5). A step size of 1 cM was chosen to find regions with potential QTL effects, i.e., where the LOD score was greater than the threshold. QTLs were detected on linkage group B7 for all the five compound contents (Table S5 and Figure S3). For example, in 2014, *E,E*-farnesol biosynthesis was detected on the male linkage group B7 (LG B7) at position 51.173 cM (with the marker Rh12GR_21458_519) with a LOD score higher than 20, explaining 49.9% of the observed variation in *E,E*-farnesol content (Figure 3b). Next, we developed a genetic marker for *RwNUDX1-2*, and mapped the gene on the LG B7 at the same position of the Rh12GR_21458_519 marker (Figure 3a). In the reference sequence (homozygous genome from OB, Hibrand Saint-Oyant *et al.*, 2018; Raymond *et al.*, 2018), no *NUDX1-2* sequence was detected in this region (Figure S4). In the heterozygous sequence of OB (Raymond *et al.*, 2018), *RcNUDX1-2c* (RcHt_S2031.3) is located on the scaffold 2031. The scaffold 2031 is syntenic with a region of chromosome 7 (between 36.2 and 37.0 Mb, Figure S4) at a position close to the peak of the QTL (marker Rh12GR_21458_159 located at position 4.38 Mb of the chromosome 7). These results clearly demonstrate that *NUDX1-2c* co-localized with the QTL for *E-E*-farnesol production.

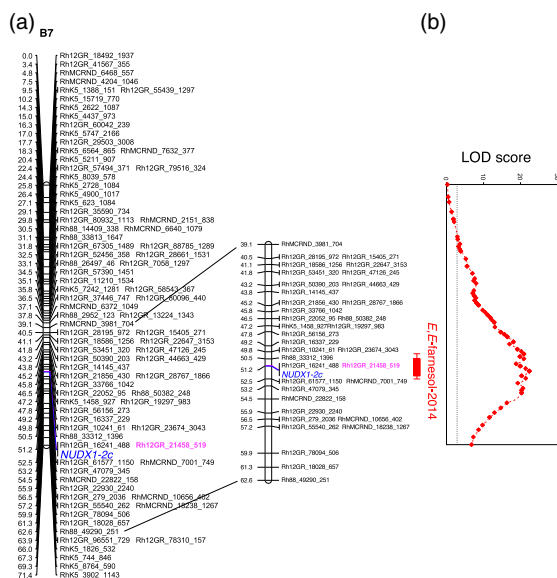


Figure 3. (a) Genetic map of the male linkage group 7, based on SNP and *RwNUDX1-2* markers. The genetic region where the QTL for *E,E*-farnesol production is located is enlarged to see the position of the markers. (b) A LOD score curve obtained from interval mapping analysis of *E,E*-farnesol contents in flowers on the male linkage group B7. The dashed line indicates the genome-wide significant threshold based on a permutation test. The marker framed in pink (Rh12GR_21458_519) indicates the highest LOD obtained in this analysis. The unit for the locus axis is centimorgan (cM). Volatile contents were analyzed in 2014.

Functional characterization of recombinant NUDX1 proteins in vitro and in vivo

To test the function of *NUDX1-2*, *in vitro* protein assays and *in vivo* transient transformation assays were performed. For the protein assay, five *NUDX1* proteins were expressed and purified using a bacterial Rosetta[®] system, including two *NUDX1-1* sequences from OB (*RcNUDX1-1a* and *RcNUDX1-1b*), one from *Rw* (*RwNUDX1-2c*), one from *PM* (*RhNUDX1*), and one from *A. thaliana* (*AtNUDX1*). Primers for gene cloning into expression vectors are listed in Table S2. Each purified protein was incubated with five potential substrates: GPP, FPP, IPP, dGTP, and 8-oxo-dGTP; the results are presented in Table 2.

All *NUDX1* proteins could convert GPP, FPP, and IPP into related products, but only *AtNUDX1* could use dGTP and 8-oxo-dGTP as substrates (Table 2). Among all *NUDX1-1* proteins, *RcNUDX1-1b* and *AtNUDX1* exhibited higher k_{cat}/K_M values on both GPP and FPP than those of the other rose proteins, indicating that *AtNUDX1* and *RcNUDX1-1b* could use these substrates with higher efficiency *in vitro*. Most *NUDX1-1* proteins had similar k_{cat}/K_M values for GPP and FPP, indicating that they had similar reaction efficiencies on these two substrates *in vitro*. For *RwNUDX1-2c*, however, the k_{cat}/K_M value for FPP was 140 times higher than that for GPP, indicating that *NUDX1-2* preferred FPP over GPP *in vitro*. In addition, all proteins accepted IPP as substrate. Unlike *AtNUDX1*, rose *NUDX1* proteins did not accept dGTP and 8-oxo-dGTP as substrates. However, due to the liquid chromatography LC-MS method used for activity measurements, kinetics parameters could not be determined for IPP and 8-oxo-dGTP substrates.

In order to obtain more evidence for the function of *RwNUDX1-2*, transient transformation of *RwNUDX1-2c* (*35S:RwNUDX1-2c*) in *Nicotiana benthamiana* leaves was conducted, in parallel with transient transformation of *RhNUDX1* (*35S:RhNUDX1*, positive control) and *GFP* (*35S:GFP*, negative control). Three days after transformation, infiltrated leaves were collected and freeze-dried, followed by extraction and analysis of geraniol and farnesol glycosides using ultrahigh-performance LC-MS. Very small quantities of geraniol and farnesol glycosides were found in the leaf samples that were infiltrated with *35S:GFP* construct. A significant amount of geraniol glycosides was detected in the leaf samples that were infiltrated with *35S:RhNUDX1* construct (Student *t* test, $P \leq 0.001$) (Figure 4). The amounts of farnesol glycosides were of the same order as that of the control leaves. Interestingly, significant amounts of both geraniol and farnesol glycosides were found in the leaf samples that were infiltrated with *35S:RwNUDX1-2c* (Student *t* test, $P \leq 0.001$), indicating that the *RwNUDX1-2* protein is involved in the production of farnesol and geraniol *in vivo*.

190 *Pulu Sun et al.*

Table 2 Kinetic parameters of NUDX1 proteins: RcNUDX1-1a, RcNUDX1-1b, RhNUDX1, RwnUDX1-2c, and AtNUDX1, with several potential substrates

| Protein | Substrates | Activities ^a | K_M (M) | k_{cat} (sec ⁻¹) | k_{cat}/K_M (M ⁻¹ sec ⁻¹) |
|------------|------------|-------------------------|---|--------------------------------|--|
| RcNUDX1-1a | GPP | + | 1.96×10^{-6} (1.36×10^{-7}) | 0.20 (0.09) | 1.02×10^5 |
| | FPP | + | 4.00×10^{-6} (1.01×10^{-6}) | 0.73 (0.48) | 1.82×10^5 |
| | IPP | + | n.c. ^b | n.c. | n.c. |
| | dGTP | - | n.d. ^c | n.d. | n.d. |
| | 8-oxo-dGTP | - | n.d. | n.d. | n.d. |
| RcNUDX1-1b | GPP | + | 2.61×10^{-7} (1.10×10^{-7}) | 0.30 (0.09) | 1.15×10^6 |
| | FPP | + | 5.55×10^{-7} (6.86×10^{-8}) | 1.37 (0.81) | 2.47×10^6 |
| | IPP | + | n.c. | n.c. | n.c. |
| | dGTP | - | n.d. | n.d. | n.d. |
| | 8-oxo-dGTP | - | n.d. | n.d. | n.d. |
| RhNUDX1 | GPP | + | 1.13×10^{-6} (4.13×10^{-7}) | 0.29 (0.11) | 2.57×10^5 |
| | FPP | + | 1.54×10^{-6} (1.88×10^{-7}) | 1.21 (0.74) | 7.86×10^5 |
| | IPP | + | n.c. | n.c. | n.c. |
| | dGTP | - | n.d. | n.d. | n.d. |
| | 8-oxo-dGTP | - | n.d. | n.d. | n.d. |
| RwnUDX1-2 | GPP | + | 8.78×10^{-6} (9.01×10^{-7}) | 0.21 (0.09) | 2.39×10^4 |
| | FPP | + | 4.95×10^{-7} (7.90×10^{-8}) | 1.67 (1.51) | 3.37×10^6 |
| | IPP | + | n.c. | n.c. | n.c. |
| | dGTP | - | n.d. | n.d. | n.d. |
| | 8-oxo-dGTP | - | n.d. | n.d. | n.d. |
| AtNUDX1 | GPP | + | 1.38×10^{-7} (4.86×10^{-8}) | 0.26 (0.10) | 1.88×10^6 |
| | FPP | + | 4.84×10^{-7} (1.05×10^{-7}) | 1.91 (0.45) | 3.95×10^6 |
| | IPP | + | n.c. | n.c. | n.c. |
| | dGTP | + | n.c. | n.c. | n.c. |
| | 8-oxo-dGTP | + | n.c. | n.c. | n.c. |

Data are presented as the means of three to six replicates using native protein, and the standard deviation (SD) is indicated between brackets. FPP, farnesyl diphosphate; GPP, geranyl diphosphate; dGTP, deoxyguanosine triphosphate; IPP, isopentenyl diphosphate; 8-oxo-dGTP, 7,8-dihydro-8-oxo-deoxyguanosine triphosphate.

^aActivities indicate the interaction between the protein and the substrate, + means interaction was detected and - means no interaction was detected.

^bn.c., not calculated, which means that there was a detectable activity but data could not be used for calculation.

^cn.d., not detected, which means that activity was too low to be detected with given substrates.

Determination of the crystal structure of the rose NUDX1-1 protein

As demonstrated above, RhNUDX1, RcNUDX1-1a, RcNUDX1-1b, RwnUDX1-2c, and AtNUDX1 can interact with FPP, GPP, and IPP (Table 2). However, rose NUDX1 seem to have lost their function on oxidized nucleotides (e.g., 8-oxo-dGTP). Moreover, NUDX1-2 seems to be more active on FPP than on GPP, which is not the case for the other rose proteins. In order to study the NUDX1-substrate interactions and to find out if the 3D structure was responsible for the different substrate specificities, the crystallization of rose NUDX1 protein was carried out. NUDX1 protein requires cations for the catalytic reaction. Divalent cations were therefore excluded during crystallization assays to inactivate the enzyme and to capture the substrate bound to the enzyme. The structure of RhNUDX1, the first structure of a Nudix enzyme from rose, was thus solved in the absence and in the presence of the GPP substrate (Figure 5a). The RhNUDX1 structure without any substrate was refined at 1.7 Å resolution (Table S6) with

one molecule per asymmetric unit in the protein crystal. In contrast to AtNUDX1, RhNUDX1 seems to be monomeric in solution, as the analysis using PISA (Krissinel and Henrick, 2007) of intermolecular contacts in the crystal did not identify assemblies of higher order. RhNUDX1 harbors the Nudix fold ($\alpha/\beta/\alpha$ sandwich structure). The GPP-bound structure of RhNUDX1 was obtained by flash-soaking crystals of RhNUDX1 in a solution of GPP and was refined at 1.45 Å resolution (Table S6). No structural difference was detected when comparing the structures of RhNUDX1 with and without GPP (root mean square deviation [RMSD] of C α , 0.19 Å) (Figure 5a). The X-loop (amino acids from position 86 to 90), which connects β 5 and β 6 strands and which is less conserved in the NUDX1 family, has been previously proposed to carry the substrate specificity (Srouji *et al.*, 2017). This loop seems to be more dynamic in the presence of GPP and does not participate in the binding of GPP (Figure 5a). The GPP substrate showed a well-defined electron density map in the active site of RhNUDX1 (Figure 5b).

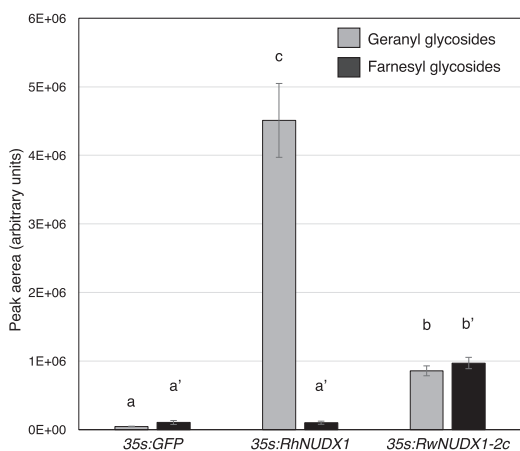


Figure 4. Patterns of terpenol glycoside accumulation following transient expression of *NUDX1* genes in *N. benthamiana*. Leaves of *N. benthamiana* were transformed with a *NUDX1* construct (35S:RwNUDX1-2c or 35:RhNUDX1) or with the 35S:GFP control. For each construct, eight independent biological replicates were used to quantify relative amounts of terpenol glycosides. Relative amounts are given as mean peak areas corresponding to the [C10H17]⁺ ion (*m/z* 137.1325) for geranyl glycosides and to the [C15H24]⁺ ion (*m/z* 205.1952) for farnesyl glycosides (expressed as arbitrary units); bars indicate the standard error. Means with different letters are significantly different (Student's *t* test, *P* < 0.01).

The structure of the RhNUDX1/GPP complex is remarkably similar to the structure of the AtNUDX1/GPP complex (PDB code 5GP0, Liu *et al.*, 2018), with an RMSD of C α atoms of 0.75 Å (Figure 5c). In these structures, GPP is nearly at the same position. Both structures were solved in the absence of cations (due to the absence of cations in the crystallization assays for RhNUDX1 and an E56A mutation in the Nudix motif of AtNUDX1 to prevent cation binding). In the RhNUDX1/GPP complex, the two phosphate groups of GPP make polar contacts with residues H49, Y94, S47, and R34 and the aliphatic chain of GPP sits in a hydrophobic pocket (Figure 6a). Although GPP is bound in the active site, it is not in a position to be hydrolyzed, as suggested by the structure of complexes of AtNUDX1/IPP (Henry *et al.*, 2018) and AtNUDX1/8-oxo-dGTP solved with Mg²⁺ bound in the active of the enzymes (Jemth *et al.*, 2019). Therefore, we turned to molecular modeling to further investigate and understand the substrate discrimination by rose NUDX1 proteins.

Modeling of rose NUDX1 proteins

From the newly determined RhNUDX1 crystal structure, we conducted molecular modeling on rose NUDX1 proteins, based on the available crystal structures of AtNUDX1 complexed with 8-oxo-dGTP (Jemth *et al.*, 2019), with IPP (Henry *et al.*, 2018), and with GPP (Liu *et al.*, 2018), of *E. coli* MutT complexed with 8-oxo-dGMP (Nakamura *et al.*, 2010), and of human MTH1 complexed with 8-oxo-

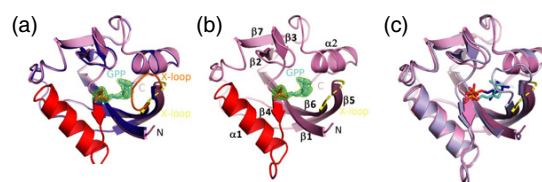


Figure 5. (a) Superimposition of the RhNUDX1/GPP complex (pink) and apo RhNUDX1 (blue). The X-loop for the apo structure is colored in orange. (b) Structure of the RhNUDX1/GPP complex. The electron density map contoured at 1.0 σ is shown around the GPP ligand colored in cyan. The secondary structure elements are numbered according to Figure S5, and the NUDX box (part of β 4-loop- α 1-loop) and the X-loop are colored in red and yellow, respectively. (c) Superimposition of RhNUDX1/GPP (pink) and AtNUDX1/GPP (PDB code 5GP0) (light blue). GPP is shown as sticks in cyan (RhNUDX1) and blue (AtNUDX1).

GMP (Svensson *et al.*, 2011). Models of others Rose NUDX1 were obtained by structural homology from crystal structures of NUDX1 proteins. The amino acid sequences of rose NUDX1 proteins were first aligned with AtNUDX1, MutT, and MTH1 according to sequence conservation and secondary structure and manually adjusted in order to have a correct structure-based sequence alignment (Figure S5). The AtNUDX1/IPP (PDB code 6DBZ) and AtNUDX1/8-oxo-dGTP (PDB code 6FL4) structures contain two Mg²⁺ ions in the active site, stabilizing the interaction with their substrates. These structures were used to position Mg²⁺ ions and ligands in the active sites of NUDX1 proteins. Molecular dynamics (MD) trajectories were calculated for 20 ns and were analyzed for C α RMSD, protein residue fluctuation (RMSD), ligand position RMSD, the distance between Mg²⁺ and the phosphate group, and predicted binding energy (Figure S6 and Table S7). The values calculated for the analysis of MD trajectories showed limited variations over time, suggesting that the structures modeled after energy minimization and 20 ns MD represent stable ligand-bound states. The representative structures, corresponding to the mean binding energy at the end of the MD trajectories, were compared.

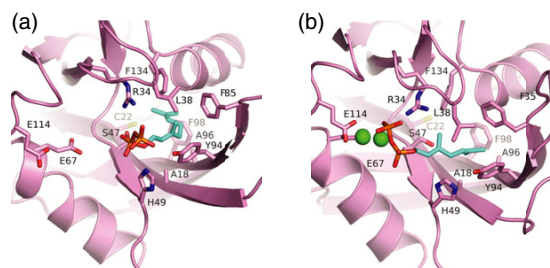


Figure 6. Comparison of GPP ligand position (a) in the crystal structure of RhNUDX1 and (b) in the model obtained by molecular modeling. The geranyl moiety is shown in cyan and the phosphate moiety in orange. Mg²⁺ ions are shown in green. The side chains of the residues of the ligand binding site are shown in sticks.

The analysis of averaged values produced during MD simulations (Table S7) show that AtNUDX1 exhibits the lowest predicted binding energy for 8-oxo-dGTP, whereas RwnNUDX1-2c shows the lowest predicted binding energy for FPP among the rose NUDX. This indicates that AtNUDX1 presents a higher affinity for 8-oxo-dGTP and RwnNUDX1-2C for FPP, which is in good agreement with the data in Table 2. MD models show a good fit with known crystal structures. Indeed, in models of AtNUDX1/8-oxo-dGTP and AtNUDX1/GPP, ligands are very close to their position in the crystal structures (PDB codes 6FL4 and 6DBZ, respectively) (Figure S6a,d). For the RhNUDX1/GPP model, although the ligand is nearly at the same position (Figure 6b and Figure S7e), GPP is moved closer to the Mg²⁺ coordination sites, confirming that in the crystal structures of RhNUDX1/GPP and AtNUDX1/GPP (5GP0), GPP is too far from the Mg²⁺ coordination sites for hydrolysis. Thus, MD provides a model for the RhNUDX1/Mg²⁺/GPP complex that is relevant for the hydrolysis of the phosphate group and compatible with geranyl chain position (Figure 6b).

The binding site pockets with their equivalent amino acids for AtNUDX1, RhNUDX1, and RwnNUDX1-2c are shown in Figure S7. In order to find differences in the active sites of these proteins that could explain their substrate preferences, only amino acids which are not conserved in these proteins are discussed below. Five residues that are not conserved and that can account for substrate preferences were identified in the substrate pocket of these proteins, namely, A11, F15, G40, S89, and S91 in AtNUDX1; A18, C22, S47, A96, and F98 in RhNUDX1; and V19, C23, G48, V97, and S99 in RwnNUDX1-2c. In addition, it has been shown in AtNUDX1 that N76 and S89, which are in the binding site pocket, play a role in 8-oxo-dGTP hydrolysis (Jemth *et al.*, 2019). The position corresponding to N76 is conserved in all NUDX studied (Figure S5, N83 in RhNUDX1). However, there is no serine at the position corresponding to S89 for rose enzymes. This position is occupied either by an alanine residue (A96 in RhNUDX1, RcNUDX1-1a, and RcNUDX1-1b) or a valine residue (V97 in RwnNUDX1-2c). This amino acid substitution may account for a lower activity towards 8-oxo-dGTP and could explain differences in substrate preferences. Indeed, S89 establishes a hydrogen bond with oxygen from 8-oxo-dGTP (Figure S7), which cannot happen with V97 or A96.

Moreover, the binding pocket is slightly different in RwnNUDX1-2c compared to other NUDX. RwnNUDX1-2c has a valine residue (V97) which is bulkier than the corresponding alanine present in other rose NUDX. We already mentioned that this position was shown to be important for interaction with 8-oxo-dGTP in AtNUDX1 (Jemth *et al.*, 2019). In RwnNUDX1-2c, V97 may render the binding pocket more hydrophobic. RwnNUDX1-2c also has a larger ligand

binding site and could fit the longer FPP aliphatic chain. The ligand binding pocket is narrower in AtNUDX1 and in RhNUDX1 due to the presence of F15 and F98, respectively, while these positions are occupied by C23 and S99 in RwnNUDX1-2c. Therefore, a subtle combination of amino acids in the substrate binding pocket may account for the observed substrate specificities.

DISCUSSION

Not so long ago, NUDX1 proteins were generally considered as ‘house cleaning’ enzymes (Bessman *et al.*, 1996; Yoshimura *et al.*, 2007; Bessman, 2019). Indeed, NUDX1 proteins from *E. coli* and humans have been shown to play major roles in removing oxidized nucleotides (Setoyama *et al.*, 2011; Gad *et al.*, 2014). The human protein MTH1 is critical for cancer cell survival where oxidative damage is very high, and targeting this protein is a highly promising anticancer strategy (Carter *et al.*, 2015). However, involvement of a NUDX1 protein in scent production has only been described in roses recently (Magnard *et al.*, 2015). Since this initial discovery, several studies have shown that in plants, these proteins may have evolved to fulfill other roles, apart from sanitizing the cell from oxidized nucleotides. In Arabidopsis, two Nudix hydrolases, among which AtNUDX1, were shown to play a role in terpene biosynthesis, by regulating the ratio of their precursors, IPP and dimethylallyl diphosphate (Henry *et al.*, 2018). It is not known if the two functions, sanitization of nucleotide pools and regulation of terpene biosynthesis, are both performed by AtNUDX1 *in vivo*, perhaps in different tissues. It was shown that AtNUDX1 activity on 8-oxo-dGTP is poor compared to that of human MTH1 (Jemth *et al.*, 2019). Nevertheless, AtNUDX1 is the only one of 25 Nudix hydrolases present in Arabidopsis able to act on mutagenic derivatives of dGTP and GTP and it is fully able to complement MutT mutation in *E. coli* (Ogawa *et al.*, 2005; Ogawa *et al.*, 2008).

In this work, we investigated the molecular bases of the peculiar scent composition of Rw, whose petals emit significant amounts of *E,E*-farnesol and other sesquiterpenes. To this end, we characterized F₁ individuals in a cross between OB and Rw, two rose cultivars with striking differences in scent composition. We found that only two genes were highly expressed in the petals of the parental lines OB and Rw, namely, *RcNUDX1-1a* and *RwnNUDX1-2c*, respectively. Expression of the *RcNUDX1-1a* gene correlated with the production of geraniol in the OW progeny. This finding was expected because *RcNUDX1-1a* was very similar to *RhNUDX1* from *R. x hybrida* cv. ‘Papa Meiland’, which has been shown to be responsible for the biosynthesis of monoterpenes, especially geraniol in this highly fragrant cultivar (Magnard *et al.*, 2015). Interestingly, expression of *RwnNUDX1-2c* correlated with the production of *E,E*-farnesol and other sesquiterpenoids in the OW progeny. A major QTL governing the production of *E,E*-

farnesol was detected, which co-localized with *RhNUDX1-2c*, providing strong evidence that polymorphism at the *NUDX1-2c* locus was responsible for the difference of *E,E*-farnesol production in the OW progeny. We also have shown that there is a link between *NUDX1-2* expression and the presence of linear sesquiterpenes such as *E,E*- α -farnesene and *E*- β -farnesene. As QTLs for these compounds co-localized with the QTL for farnesol, we can assume that this locus is involved in the biosynthesis of these compounds as well. We do not have mechanistic data to explain this finding. FP could be converted to these products by an uncharacterized pathway. Alternatively, these compounds could be degradation products arising from farnesol *in planta* or during the extraction process. A terpene synthase has been previously shown to be responsible for the production of the sesquiterpene germacrene D in roses (Guterman *et al.*, 2002). It is interesting to see that expression of *NUDX1-2* in the progeny did not correlate with all sesquiterpenes; for example, it was not correlated with the production of germacrene D (Data S1). Therefore, it seems that in rose petals, two different pathways may operate for the production of sesquiterpenoids.

In vitro assays showed that the recombinant rose NUDX1 proteins could hydrolyze GPP, FPP, and IPP into GP, FP, and IP respectively. GP has been shown to be further hydrolyzed by an unidentified phosphatase (Magnard *et al.*, 2015), which may also be the case for FP and IP. The RcNUDX1-1a and RcNUDX1-1b proteins did not show a marked preference for GPP or FPP *in vitro*. Conversely, RwNUDX1-2 used FPP more efficiently than GPP. Expression of *RhNUDX1* led to the accumulation of geranyl glycosides in tobacco leaves, without detectable biosynthesis of farnesyl glycosides, suggesting a strict selection of the GPP substrate *in vivo*. Conversely, expression of *RwNUDX1-2* led to accumulation of both farnesyl and geranyl glycosides, indicating a relaxed substrate preference *in vivo*, which may reflect differences in GPP and FPP substrates availability in *N. benthamiana* leaves. Indeed, GPP and FPP availabilities in the Rw rose cultivar could be major factors impacting RwNUDX1-2c substrate preference *in planta*. Nevertheless, experiments in *N. benthamiana* showed that the biosynthesis of significant amounts of *E,E*-farnesol derivatives *in vivo* was specific to the RwNUDX1-2 protein. Taken together, the activity of the RwNUDX1-2 *in vitro* and *in planta* and the co-localization of the corresponding gene with a QTL for *E,E*-farnesol accumulation are consistent with RwNUDX1-2 being involved in the biosynthesis of *E,E*-farnesol and other sesquiterpenoids in the petals of Rw.

E,E-farnesol is a sesquiterpene alcohol with a mild, delicate, sweet-oily odor (Lapczynski *et al.*, 2008). It is used as an ingredient for perfume and cosmetic products, such as deodorants and bath products. It may be present in high amounts in petals of some flowers such as *Actinidia*

chinensis (Green *et al.*, 2012). In plants, only five farnesol synthases have been characterized so far (Schnee *et al.*, 2002; Cheng *et al.*, 2007; Parveen *et al.*, 2015; Chen *et al.*, 2016; Rusdi *et al.*, 2018), and three of them produce a blend of sesquiterpenes when provided with the substrate FPP. In the present work, we show that, besides terpene synthases, *Nudix hydrolases* may be involved in the biosynthesis of sesquiterpenes, as shown previously in the case of monoterpenes (Magnard *et al.*, 2015).

Though several crystal structures of plant NUDX have been solved with various substrates, the structural basis for substrate preference is still under investigation. In this paper, MD helped to understand the catalytic mechanism of NUDX1. Key amino acids for substrate binding were identified. Based on the results from MD, two positions corresponding to S89 and S91 in AtNUDX1, A96 and F98 in RhNUDX1, and V97 and S99 in RwNUDX1-2c could be essential to explain RwNUDX1-2c preference for FPP. However, although the important S89 in AtNUDX1, which is H-bonded to 8-oxo-dGTP, is absent in the rose NUDX, the structural modeling could not fully explain why AtNUDX1 is the only enzyme which can hydrolyze 8-oxo-dGTP. Sequence comparison of RcNUDX1-1a, RcNUDX1-1b, RhNUDX1, RwNUDX1-2c, and AtNUDX1 also revealed differences in the X-loop region (86–91 with sequence LDEAKP in RhNUDX1) (Figure S5). This loop, which is well defined in the RhNUDX1 apo structure but which is disordered in RhNUDX1-GPP crystal structures, is one of the most flexible parts during MD simulations, as plotted on the protein C α fluctuation diagram (Figure S6). It may also interact with the substrate, but the precise role of the X-loop is not known. It has been suggested that modifications within this region could alter substrate specificity, thus allowing for altered substrate specificity and neofunctionalization (Srouji *et al.*, 2017). NUDX1 proteins can have very different binding mechanisms even if they use the same substrates. It has been shown that the substrate-interacting amino acids are completely different when comparing MTH1 and MutT (Svensson *et al.*, 2011), which makes the identification of key amino acids quite complicated. Although there are no specific residues in the X-loop which interact with the ligand, the flexibility and dynamics of this loop may still be important for ligand turnover.

We showed here that in Rw, the NUDX1-2 protein plays a role in the biosynthesis of sesquiterpenoid scent compounds, as a result of functional diversification in the *Nudix hydrolase* gene family. It remains to be established whether the NUDX1-dependent scent biosynthetic pathways are specific to roses, as they have not been characterized in other plant species so far. In many plants, such as rice (*Oryza sativa*) and maize (*Zea mays*), biosynthesis of farnesol is catalyzed by typical terpene synthases (Schnee *et al.*, 2002; Cheng *et al.*, 2007). It is interesting to note that farnesol is a precursor of the juvenile hormone in insects

194 Pulu Sun et al.

(Bellés *et al.*, 2005; De Loof and Schoofs, 2019). In some of these insects, it has been shown that farnesol is derived from FPP with the assistance of phosphatases from the haloalkanoic acid dehalogenase superfamily (Cao *et al.*, 2009; Nyati *et al.*, 2013). In recent years, the importance of the convergence of multiple unrelated metabolic pathways towards the biosynthesis of a single compound has been increasingly acknowledged (Sun *et al.*, 2016). The biosynthesis of farnesol is therefore another striking example of such metabolic convergence.

EXPERIMENTAL PROCEDURES

Plant material

A population of 151 F₁ progeny was generated by crossing two diploid roses (Hibrand Saint-Oyant *et al.*, 2018): *R. chinensis* cv. 'Old Blush' (OB, maternal plant) and *R. x wichurana* (Rw, paternal plant) originated from the Bagatelle garden (Paris, France). All these rose plants were grown outside at INRA Angers (Experimental Unit Hortis, Angers, France). Petals of all descendants were collected between May and July in 2014 and 2015. Progeny (148 and 132 individuals) were collected in 2014 and 2015, respectively. This difference in sample size is because some roses that were available in 2014 were dead in 2015 or too young to have flowers in 2014 and became capable of blooming in 2015. Collected petals were subjected to hexane extraction and extracts were analyzed by GC-MS. Transient transformation assays were performed on the leaves of 4-week-old *N. benthamiana* plants that were grown inside a climate room (21 ± 1°C, 16-h light period and 8-h dark period).

Total RNA extraction and primary cDNA synthesis

Sixteen out of the 156 progeny and the two parents were selected for total RNA extraction: OW9007, OW9011, OW9013, OW9018, OW9021, OW9024, OW9035, OW9037, OW9047, OW9049, OW9069, OW9074, OW9082, OW9099, OW9149, OW9204, OB, and Rw. The rose petals of these selected plants were collected and frozen immediately in liquid nitrogen. Prior to RNA extraction, frozen rose petals were ground to fine powder in liquid nitrogen using a sterilized mortar and pestle. Approximately 300–400 mg (3–4 tubes containing 100 mg ± 15%) of frozen ground petals was used in order to obtain sufficient amounts of total RNA. Total RNA was extracted using the NucleoSpin[®] RNA plant kit (MACHEREY-NAGEL, Düren, Germany) according to the manufacturer's instructions with slight modification as follows: before adding the lysis buffer (RAP), a spatula tip (around 5% of the sample weight) of polyvinylpyrrolidone (PVP-40, Sigma-Aldrich, St. Louis, USA) and ethylhexadecyldimethyl ammonium bromide (CTAB) (Sigma-Aldrich) was added directly to the ground petals, followed by 5 µl of β-mercaptoethanol (≥99.0%, Sigma-Aldrich). RNA was dissolved in the mixture at 60°C for 5 min and subsequently extracted using 1 volume of chloroform or chloroform:isoamyl alcohol (24:1) (Sigma-Aldrich). Supernatant was transferred to a new tube for further procedures. Prior to the loading, 0.5 volume of 100% ethanol was added in order to adjust the binding condition. The remaining washing procedures were performed according to the manufacturer's protocol, including decontamination of genomic DNA using DNase.

The quality of the RNA samples was evaluated by absorbance measurements using a NanoDrop 2000c (Thermo Fisher Scientific, Waltham, USA), and the integrity was determined by

electrophoretic analysis. All the RNA samples used for the qRT-PCR reactions had a 260/280 nm absorbance ratio between 1.7 and 2.2, while RNA samples with a 260/280-ratio close to 2 were generally qualified for subsequent reactions. To rule out any genomic DNA contamination in the RNA extracts, the RNA samples were subjected to PCR amplifications of the Tubulin and *NUDX1* gene (35 cycles). No visible amplifications of genomic DNA were detected from the RNA samples. The primary cDNA was synthesized from approximately 2 µg of total RNA using the SuperScript[®] III reverse transcriptase, oligo-dT, and RNasOUT[™] recombinant RNase inhibitor (all mentioned products above are from Thermo Fisher Scientific) in a reaction volume of 20 µl according to the manufacturer's procedure. The quality of the synthesized cDNA was tested using PCR with the primers of the housekeeping gene castor bean (*Ricinus communis*) translationally controlled tumor protein (*RcTCTP*). If an amplicon was found after 28 cycles of amplification, then the quality of cDNA sample was assumed good enough for further analysis.

qRT-PCR analysis

The reaction mixture of primary cDNA was diluted 125 times before qRT-PCR, which was carried out with a CFX96[™] Real-Time system equipped with C1000[™] Thermal Cycler (Bio-Rad, California, USA). The qRT-PCR reaction mixture consisted of 10 µl of SsoAdvanced[™] SYBR[®] Green Supermix (Bio-Rad), 2 µl of a pair of primers (1 µl for each primer, final primer concentration: 10 µM), 2 µl of diluted cDNA mixture, and DNase-free water to reach a total volume of 20 µl. The thermal cycling profile applied was 95°C for 5 sec, followed by 30 sec at 58, 60, or 64°C depending on the primer pair. In total 40 cycles were performed for each batch of samples. Within each qRT-PCR batch, two negative controls with water instead of cDNA were used. All qRT-PCR were carried out in two biological replicates, each of which was subjected to RNA extraction followed by qRT-PCR in two technical replicates. The resulting quantification cycle (C_q) values were the mean of four values from two biological and two technical replicates. These C_q values were automatically determined by the CFX96[™] Real-Time system with default settings. In order to evaluate the changes in expression levels of certain genes, three housekeeping genes that have consistent transcription levels across different samples were used as reference genes, i.e., genes that code for α-tubulin, elongation factor 1-alpha, and *TCTP* according to Dubois *et al.* (2012). To obtain the relative quantities of each amplified product in the samples (relative to the reference genes), the ΔΔC_t method was applied (Pfaffl, 2001). The melting curves of the amplified products were analyzed to determine the specificity of each qRT-PCR reaction, using the built-in standard method in the system. The primers that were used for RT-PCR and qRT-PCR are listed in Table S2.

RNA sequencing and assembly

Extracted total RNAs of OB, Rw, OW9035, and OW9047 hybrids with RNA integrity number (RIN) higher than 7 were used for RNA sequencing (Eurofins Genomics, Ebersberg, Germany) using an Illumina HiSeq 2500 sequencer with single-read module. The quality of the obtained RNA sequencing reads was assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and reads were trimmed to remove sequencing adapters using Trimmomatic (Bolger *et al.*, 2014). This was followed by reassessment using FastQC. The trimmed sequence reads of the above four samples were pooled together and used for *de novo* transcriptome assembly, using the software suite Trinity (Grabherr *et al.*, 2011). Next, we used the generated transcriptome as a

reference to map the reads from each individual sample and then estimate contig abundances using the RNA-Seq by Expectation Maximization (RSEM) package (Li and Dewey, 2011). This was followed by gene differential expression analysis in pairs using the edgeR package (Robinson *et al.*, 2010). The trimmed mean of M-values (TMM) method (Robinson and Oshlack, 2010) was applied for read counts normalization in FPKM values. Finally, the normalized data were centered by median and \log_2 -transformed to obtain relative expression levels.

QTL analyses and development of a genetic marker for *NUDX1-2*

QTL analysis was carried out using MapQTL[®] 5.0 (Van Ooijen, 2004). These QTL analyses were conducted using female and male SNP maps that have been developed using JoinMap 4.0 (Van Ooijen, 2006). Due to non-normality for most of the metabolites, the data were analyzed first by KW. Interval mapping analysis was performed with a step size of 1 cM to find regions with potential QTL effects, i.e., where the LOD score was greater than the threshold. A LOD threshold at which a QTL was declared significant was determined according to a genome-wide error rate of 0.05 over 1000 permutations (Churchill and Doerge, 1994). The percentage explained by the QTL (r^2) was also presented.

Based on the *RcNUDX1-2c* sequence (RcHt_2031.3), PCR primers (RhNUDX1-2_F1 and RhNUDX1-2_R1, Table S2) were developed and length polymorphism was detected: two DNA fragments (between 700 and 800 bp) for the locus in *Rw* and one DNA fragment for the same locus in *OB* (around 700 bp). PCR reactions were performed in 25 μ l volume with 10 ng genomic DNA, 1 \times Q5[®] buffer, 0.200 mM dNTPs, 0.5 μ M of each primer, and 0.02 U of Q5[®] High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, USA), with the following program: 98°C for 30 sec, 25 cycles (98°C for 10 sec, 60°C for 15 sec, and 72°C for 90 sec), and 72°C for 7 min. The PCR products were separated on agarose gel (2% w/v) during 2 h at 140 V and stained with ethidium bromide.

NUDX1 enzyme assay and kinetic parameters analysis using LC-MS

In order to determine the kinetic parameters of *NUDX1* proteins *in vitro*, *NUDX1* genes were amplified and cloned into destination vectors pHNGWA using the Gateway[®] technique, where the *NUDX1* protein was fused with the protein NusA (Busso *et al.*, 2005). The final constructs were transformed into the bacterial strain *Rosetta*[®] (Merck, Darmstadt, Germany). Native proteins (~16 kDa) were obtained by using TALON[®] metal affinity resins (Clontech Laboratories, California, USA) to cleave away the NusA tags under the reaction of thrombin. Purified native proteins (5–56 ng) were added to substrates GPP, FPP, IPP, dGTP, or 8-oxo-dGTP, ranging from 0.1 to 100 μ M. The proteins and the substrates were incubated in reaction buffer (50 mM Tris-HCl [pH 8.8], 5 mM MgCl₂, 14 mM β -mercaptoethanol, 10% glycerol, v/v) with a final volume of 100 μ l, for 15 min at 30°C with shaking. In parallel, for each assay, a similar negative control that only contained corresponding amounts of substrates was also included. The reactions were stopped by adding 10 μ l of EDTA (100 mM, pH 8) and 90 μ l of ethanol/0.5% NH₄OH solution, according to a previously published protocol (Magnard *et al.*, 2015).

The *NUDX1*-dependent production of GP or FP was calculated after subtracting the GP or FP amounts present in the corresponding negative control set; K_M and k_{cat} values were calculated by fitting the data (a series of GPP/FPP concentrations and the GP/FP conversion rates) to the Michaelis-Menten equation using R (R

Core Team, 2015). For substrate specificity studies and the analysis of reaction products, experiments were performed according to a previously published protocol (Magnard *et al.*, 2015) with slight adjustments as follows. Purified *NUDX1* proteins (500–1000 ng) were incubated with different substrates (GPP, FPP, IPP, and 8-oxo-dGTP) at a concentration of ~20 μ M in reaction buffer (100 μ l final volume) for 1 h at 30°C with shaking at 7 \times 100 min⁻¹. The parameters and equipment settings were given in a previously published study (Magnard *et al.*, 2015).

Expression of *NUDX1* in *N. benthamiana* and LC-MS analysis of terpenol glycosides

For transient expression assays, four constructs were transformed into *Agrobacterium tumefaciens* strain C58 (pMP90). These constructs allow, respectively, the expression of the following proteins: *RwNUDX1-2c* (*35S:RwNUDX1-2c*), *RhNUDX1* (Magnard *et al.*, 2015), GFP (*35S:GFP*), and the viral suppressor of gene silencing p19 (Voinnet *et al.*, 2003). The *A. tumefaciens* cultures were grown overnight and diluted in infiltration buffer (pH 5.6, 2% glucose, 5 g L⁻¹ MS salt, 10 mM MES, and 0.2 mM acetosyringone) to obtain a final OD₆₀₀ value of 0.6. Before infiltration, the cultures corresponding to *NUDX1* and GFP constructs were mixed with the suppressor construct culture in a ratio of 1:1. The bacterial culture mixtures were kept in the dark at room temperature for at least 1 h prior to infiltration. Four-week-old non-flowering *N. benthamiana* plants were used. Young leaves were infiltrated either with *A. tumefaciens* cultures harboring a *NUDX1* construct (*35S:RwNUDX1-2c* or *35:RhNUDX1*) or the *35S:GFP* control, all in combination with *35S:P19*. Three days after infiltration, 0.75 \pm 0.1 g of fresh transformed leaves were collected, immediately frozen in liquid nitrogen, and stored at -80°C before use. Frozen leaves were freeze-dried prior to extraction and analysis. Terpenol glycosides were extracted from powdered freeze-dried *N. benthamiana* leaves with methanol, using 20 μ l of methanol per mg of dry weight. The samples were subjected to sonication for 10 min in an ultrasonic bath (Elma, Singen, Germany) and centrifuged (11 292 g, 10 min, 4°C). The supernatants were recovered for analysis. Terpenol glycosides were analyzed using LC-MS as described previously (Magnard *et al.*, 2015). Terpenol glycosides were quantified using their respective characteristic ion: [C₁₀H₁₇]⁺ (m/z 137.1325) for geranyl glycosides and [C₁₅H₂₄]⁺ (m/z 205.1952) for farnesyl glycosides. The major geranyl glycosides were putatively identified as hexosyl-geraniol (C₁₆H₂₈O₆), malonyl-hexosyl-geraniol (C₁₉H₃₀O₉), and pentosyl-hexosyl-geraniol (C₂₁H₃₆O₁₀). Similarly, the main farnesyl glycosides were putatively identified as hexosyl-farnesol (C₂₁H₃₆O₆) and malonyl-hexosyl-farnesol (C₂₄H₃₈O₉). For each sample, total relative amounts of geranyl glycosides were obtained by summing the peak areas corresponding to the different geranyl glycosides; the same was done for farnesyl glycosides. For each construct, eight independent biological replicates were used.

GC-MS analysis of rose petal extracts

For volatile extraction, rose petals were collected between 8:30 and 10:30 in the morning. Each year, for each individual of the OW progeny, petals from different flowers on the same individual were harvested (from four to six independent flowers, depending on the number of petals per flower). The petals were mixed, and two to four technical replicates were prepared by placing 1 g of petals into a glass vial for volatile compound extraction. One gram of collected rose petals was incubated in 2 ml of hexane containing 5 mg L⁻¹ camphor as internal standard, at 4°C overnight. On the next day, the volatile extracts were transferred to a vial for GC-

MS analysis. Extracts were analyzed on an Agilent 6850 Network GC system equipped with a DB5 apolar capillary column (30 m × 0.25 mm) and coupled with a 7683B series injector and 5973 Network mass selective detector (all components from Agilent Technologies, Santa Clara, USA). The carrier gas was helium at a flow rate of 1.0 ml min⁻¹. The separation was performed with the following program: 40°C for 3 min, and subsequently a gradient of 3°C min⁻¹ was applied until the temperature reached 245°C. The injection volume was 2 µl with a split mode (split ratio 1:2) and the injector and detector temperatures were 250°C. The parameters for the mass spectrometer detector were set as follows: the mass scan range was 35–450 *m/z* and the ionization energy was 70 eV. The identification of volatile compounds was based on their retention time and the calculated Kovats retention index in combination with their mass spectrum matching with available databases (CNRS, Wiley 275, NIST08), using both Agilent software MSD ChemStation D.02.00.275 and an archive of mass spectra of essential oil components (Adam, 2007). The farnesol produced by *Rw* was extracted and sent to the company International Flavors & Fragrances (IFF LMR Naturals, Grasse, France) for isomer identification using the isomer-isolated farnesol standard. Data were analyzed by R software (R Core Team, 2015) to examine the correlations between data and determine the distribution of the compounds and the variance between replicated samples from the same genotype.

RhNUDX1 purification and crystal structure determination

A plasmid allowing the expression of RhNUDX1 protein in *E. coli* was constructed by inserting the *RhNUDX1* gene using the Gateway[®] technique in the pHGWA vector (Busso *et al.*, 2005), which adds a His₆ tag to the C-terminus of the protein. RhNUDX1 protein was overproduced by transforming *E. coli* Rosetta 2 (DE3) competent cells by this plasmid. A single colony was used to inoculate a 5-ml preculture in Lysogeny broth (LB) supplemented with 35 µg ml⁻¹ chloramphenicol and 100 µg ml⁻¹ ampicillin. The bacteria were then transferred to 1 L Terrific Broth supplemented with 0.8% glycerol, 35 µg ml⁻¹ chloramphenicol, and 100 µg ml⁻¹ ampicillin and grown at 18°C for 24 h. Bacteria were harvested by centrifugation and the cell pellet was suspended in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 5 mM dithiothreitol (DTT), and 1 mM phenylmethane sulfonyl fluoride (PMSF) and lysed by sonication. The lysate was applied to a Ni-NTA affinity column equilibrated in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 5 mM DTT, and 20 mM imidazole. RhNUDX1 protein was eluted with a linear gradient from 20 to 500 mM of imidazole in the buffer used to equilibrate the column. Fractions containing RhNUDX1 were collected and treated with 200 U of thrombin at 4°C for 18 h in order to cleave the His₆ tag. RhNUDX1 was further purified using a Superdex 200 gel filtration column and concentrated in Amicon[®] (Millipore, Guyancourt, France) to 2.5 mg ml⁻¹. Crystallization assays were performed on RhNUDX1 with and without its GPP substrate. Crystals were obtained in 200 mM ammonium sulfate, 100 mM sodium acetate (pH 4.6), and 25% PEG 400. Crystals were cryoprotected in 30% ethylene glycol and flash-frozen in liquid nitrogen. GPP-containing crystals were obtained by soaking the crystals 1 sec in 5 mM GPP without divalent cations before flash-freezing in liquid nitrogen. Diffraction data were collected on the synchrotron SOLEIL beamline Proxima-2 for RhNUDX1 and on the ESRF beamline ID29 for RhNUDX1/GPP. Data were integrated and scaled with DIALS (Winter *et al.*, 2018). Phases were determined by molecular replacement in PHASER (Phaser crystallographic software) (McCoy *et al.*, 2007) using the structure of MutT (PDB code 4KYX) as search model. Modeling and refinement were carried out using COOT (Emsley *et al.*, 2010) and Phenix (Afonine *et al.*, 2012).

Modeling of protein structures

Protein–ligand complexes for AtNUDX1 and RhNUDX1 were modeled using their respective crystal structures. Other Rose NUDX1 structures (RcNUDX1-1a, RcNUDX1-1b, and RwNUDX1-2c) were generated by homology modeling with the program MODELLER (Šali and Blundell, 1993). In order to remove bias from a single template, several structures were used as templates: the crystal structure of AtNUDX1 in complex with IPP (PDB code 6DBZ), 8-oxo-dGTP (PDB code 6FL4), and GPP (PDB code 5GP0), MutT and MTH1 in complex with 8-oxo-GMP (PDB codes 3A6T and 3ZR0, respectively), and RhNUDX1 in complex with GPP (this work, PDB code 6YPF). A structure-based alignment with the target sequence and the template structures was generated with MODELLER and manually corrected. In a first step, the ligand was added as a rigid molecule. Structures were generated with the automodel routine of MODELLER. The Mg²⁺ ions and ligand, when absent from the structure, were added and positioned according to AtNUDX1–Mg²⁺–ligand structures (6DBZ and 6FL4). Ligand topology was added in the MODELLER database. The protein–ligand complex was locally refined with MODELLER. The structures were sorted according to their Discrete Optimized Protein Energy (DOPE) score (Shen and Šali, 2006), and the structure with the lowest DOPE score was chosen as the prototype of the protein–ligand structure. In a second step, the structure of the protein–ligand complex, either GPP, FPP, or 8-oxo-dGTP, was refined by MD simulation. The coordinates and the topology of the ligands were generated with the CGenFF server (Vanommeslaeghe and MacKerell, 2012). Mg²⁺ ions were added in the active site, as they are important for binding and catalysis. The charge of the Mg²⁺ was calculated with AMBER (Case *et al.*, 2005). The structure was then subjected to energy minimization and 20 ns MD with GROMACS (version 2018.3) (Abraham *et al.*, 2015). Input parameter files for calculations with GROMACS using the CHARMM36 force field were generated with CHARMM-GUI (Lee *et al.*, 2016). The protein–ligand complex was equilibrated in explicit solvent with 150 mM KCl in a cubic box with periodic boundary conditions at 303 K and 1 bar. The MD trajectory was analyzed with the tools provided with GROMACS. The protein–ligand binding energy was evaluated with PRODIGY-LIG (Kurkcuoglu *et al.*, 2018). The molecular structure figures were prepared with PyMOL (Schrödinger, 2015).

ACCESSION NUMBERS

Sequence data from this article can be found in the GenBank/European Molecular Biology Laboratory databases (see accession numbers in Table S8). Transcriptomic data are available from the European Nucleotide Archive under accession number PRJEB37774. The structures of RhNUDX1 and RhNUDX1/GPP were deposited to the protein databank under the accession codes 6YPB and 6YPF, respectively.

ACKNOWLEDGEMENTS

This work was supported by grants from the Région Rhône-Alpes, CNRS GDR MediatEC (3658), and ANR Rosascent. We thank Florence Nicolé (UJM Saint-Etienne) for helping with statistical analysis. We thank F. Hache from the English department of Saint-Etienne University for text revisions. We also thank F. Gros for technical assistance (UJM Saint-Etienne). We thank Thuy-Thanh Truong (INRAE, Colmar) for help with the terpenol glycoside analysis. We thank the experiment unit Hortis (Beaucouzé, France) for

taking care of the OW progeny. We thank the staff at the synchrotron SOLEIL beamline Proxima-2 and at the ESRF beamline ID29 for support in data collection. We thank IFF-LMR Naturals (Grasse, France) for help with compound identification.

AUTHOR CONTRIBUTIONS

SB conceived the original screening and research plans; SB, RCS, PH, J-CC, CT, FF, and MH supervised the experiments. PS, CD, AB, and SR performed most of the experiments. J-LM, LH-SO, TT, JJ, JM-M, AME-H, CC, SNP, and RB performed some experiments and gene expression analyses and provided conceptual and/or technical assistance to PS, CD, and AB; SB, RCS, CT, SR, and PH designed the experiments; SB, PS, and PH wrote the article with contributions from all the authors.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest associated with the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Sequences in RNA-seq transcriptome with homologies to *NUDX1* sequences.

Table S2. List of PCR primers.

Table S3. Identity percentages of amino acid sequences among *NUDX1* proteins.

Table S4. Summary of QTLs for sesquiterpenoids detected with the non-parametric Kruskal–Wallis rank-sum test.

Table S5. Summary of QTLs for sesquiterpenoids detected with interval mapping.

Table S6. Data collection and refinement statistics.

Table S7. Summary of value analysis during 20 ns MD.

Table S8. Accession numbers of the *NUDX1* sequences.

Figure S1. Alignment of amino acid sequences of four *RcNUDX1-1a* copies.

Figure S2. Distribution in the OW progeny of volatile compound amounts.

Figure S3. Localization of the QTLs for *E,E*-farnesol and other sesquiterpenoid compounds.

Figure S4. Microsynteny analysis at the *NUDX1-2c* locus.

Figure S5. Amino acid sequence alignment and structure indication of *NUDX1* proteins.

Figure S6. Analysis of the 20 ns trajectories for *NUDX1* proteins.

Figure S7. Models of *NUDX1*–substrate interactions.

Data S1. qPCR data of 18 rose individuals and their corresponding major volatile compounds and correlation coefficient values generated by Spearman's correlation test.

Data S2. Major volatile compounds extracted from petals of 148 progeny and two parents as analyzed by GC-MS.

REFERENCES

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25.

Adam, R.P. (2007) *Identification of essential oil components by gas chromatography/mass spectrometry*, 4th edition. Carol Stream, IL: Allured Publ. Corporation.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Cryst. D*, 68, 352–367.

Bellés, X., Martín, D. and Piulachs, M.-D. (2005) The mevalonate pathway and the synthesis of juvenile hormone in insects. *Annu. Rev. Entomol.* 50, 181–199.

Bergougnoux, V., Caissard, J.C., Jullien, F., Magnard, J.L., Scalliet, G., Cock, J.M., Huguency, P. and Baudino, S. (2007) Both the adaxial and abaxial epidermal layers of the rose petal emit volatile scent compounds. *Planta*, 226, 853–866.

Bessman, M.J. (2019) A cryptic activity in the *Nudix hydrolase* superfamily. *Protein Sci.* 28, 1494–1500.

Bessman, M.J., Frick, D.N. and O'Handley, S.F. (1996) The MutT Proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.* 271, 25059–25062.

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

Busso, D., Delagoutte-Busso, B. and Moras, D. (2005) Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal. Biochem.* 343, 313–321.

Cao, L., Zhang, P. and Grant, D.F. (2009) An insect farnesyl phosphatase homologous to the N-terminal domain of soluble epoxide hydrolase. *Biochem. Biophys. Res. Commun.* 380, 188–192.

Carter, M., Jemth, A.S., Hagenkorf, A. et al. (2015) Crystal structure, biochemical and cellular activities demonstrate separate functions of MTH1 and MTH2. *Nat. Commun.* 6, 7871.

Case, D.A., Cheatham, T.E. III, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688.

Chen, W. and Viljoen, A.M. (2010) Geraniol - a review of a commercially important fragrance material. *S. Afr. J. Bot.* 76, 643–651.

Chen, X., Wang, Y., Sun, J., Wang, J., Xun, H. and Tang, F. (2016) Cloning, expression and functional characterization of two sesquiterpene synthase genes from moso bamboo (*Phyllostachys edulis*). *Protein Expr. Purif.* 120, 1–6.

Cheng, A.X., Xiang, C.Y., Li, J.X., Yang, C.Q., Hu, W.L., Wang, L.J., Lou, Y.G. and Chen, X.Y. (2007) The rice (E)-beta-caryophyllene synthase (OsTPS3) accounts for the major inducible volatile sesquiterpenes. *Phytochemistry*, 68, 1632–1641.

Churchill, G.A. and Doerge, R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, 138, 963–971.

De Loof, A. and Schoofs, L. (2019) Mode of action of Farnesol, the "noble unknown" in particular in Ca²⁺ homeostasis, and its juvenile hormones in evolutionary retrospect. *Front. Neurosci.* 13, 141.

Dobrzanska, M., Szurmak, B., Wyslouch-Cieszyńska, A. and Kraszewska, E. (2002) Cloning and characterization of the first member of the *Nudix* family from *Arabidopsis thaliana*. *J. Biol. Chem.* 277, 50482–50486.

Dong, L., Miettinen, K., Goedbloed, M., Verstappen, F.W.A., Voster, A., Jongma, M.A., Memelink, J., Krol, S.V.D. and Bouwmeester, H.J. (2013) Characterization of two geraniol synthases from *Valeriana officinalis* and *Lippia dulcis*: similar activity but difference in subcellular localization. *Metab. Eng.* 20, 198–211.

Dubois, A., Carrere, S., Raymond, O. et al. (2012) Transcriptome database resource and gene expression atlas for the rose. *BMC Genom.*, 13, 638–648.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of *Coot*. *Acta Cryst. D*, 66, 486–501.

Gad, H., Koolmeister, T., Jemth, A.-S. et al. (2014) MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature*, 508, 215–221.

Grabherr, M.G., Haas, B.J., Yassour, M. et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.

Green, S.A., Chen, X., Nieuwenhuizen, N.J., Matich, A.J., Wang, M.Y., Bunn, B.J., Yauk, Y.-K. and Atkinson, R.G. (2012) Identification, functional characterization, and regulation of the enzyme responsible for

- floral (*E*)-nerolidol biosynthesis in kiwifruit (*Actinidia chinensis*). *J. Exp. Bot.* **63**, 1951–1967.
- Gunawardana, D., Likić, V.A. and Gayler, K.R. (2009) A comprehensive bioinformatics analysis of the Nudix superfamily in *Arabidopsis thaliana*. *Comp. Funct. Genomics*, **2009**, 820381.
- Guterman, I., Shalit, M., Menda, N. et al. (2002) Rose scent: genomics approach to discovering novel floral fragrance-related genes. *Plant Cell*, **14**, 2325–2338.
- Henry, L.K., Thomas, S.T., Widhalm, J.R., Lynch, J.H., Davis, T.C., Kessler, S.A., Bohlmann, J., Noel, J.P. and Dudareva, N. (2018) Contribution of isopentenyl phosphate to plant terpenoid metabolism. *Nat. Plants*, **4**, 721–729.
- Hibrand Saint-Oyant, L., Ruttink, T., Hamama, L. et al. (2018) A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nat. Plants*, **4**, 473–484.
- Iijima, Y., Gang, D.R., Fridman, E., Lewinsohn, E. and Pichersky, E. (2004) Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiol.* **134**, 370–379.
- Ito, M. and Honda, G. (2007) Geraniol synthases from perilla and their taxonomical significance. *Phytochemistry*, **68**, 446–453.
- Jemth, A.S., Scaletti, E., Carter, M., Helleday, T. and Stenmark, P. (2019) Crystal structure and substrate specificity of the 8-oxo-dGTP hydrolase NUDT1 from *Arabidopsis thaliana*. *Biochemistry*, **58**, 887–899.
- Klaus, S.M.J., Wegkamp, A., Sybesma, W., Hugenholtz, J., Gregory, J.F. and Hanson, A.D. (2005) A Nudix enzyme removes pyrophosphate from dihydroneopterin triphosphate in the folate synthesis pathway of bacteria and plants. *J. Biol. Chem.* **280**, 5274–5280.
- Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
- Kupke, T., Caparrós-Martin, J.A., Malquichagua Salazar, K.J. and Culiáñez-Maciá, F.A. (2009) Biochemical and physiological characterization of *Arabidopsis thaliana* AtCoAse: a Nudix CoA hydrolyzing protein that improves plant development. *Physiol. Plant.* **135**, 365–378.
- Kurkuoglu, Z., Koukos, P.I., Citro, N. et al. (2018) Performance of HADDOCK and a simple contact-based protein-ligand binding affinity predictor in the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **32**, 175–185.
- Lapczynski, A., Bhatia, S.P., Letizia, C.S. and Api, A.M. (2008) Fragrance material review on farnesol. *Food Chem. Toxicol.* **46**, S149–S156.
- Lee, J., Cheng, X., Swails, J.M. et al. (2016) CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **12**, 405–413.
- Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
- Liu, J., Guan, Z., Liu, H., Qi, L., Zhang, D., Zou, T. and Yin, P. (2018) Structural insights into the substrate recognition mechanism of *Arabidopsis* GPP-bound NUDX1 for noncanonical monoterpene biosynthesis. *Mol. Plant*, **11**, 218–221.
- Magnard, J.-L., Roccia, A., Caissard, J.-C. et al. (2015) Biosynthesis of monoterpene scent compounds in roses. *Science*, **349**, 81–83.
- Masumoto, N., Korin, M. and Ito, M. (2010) Geraniol and linalool synthases from wild species of perilla. *Phytochemistry*, **71**, 1068–1075.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674.
- McLennan, A. (2006) The Nudix hydrolase superfamily. *Cell. Mol. Life Sci.* **63**, 123–143.
- Mildvan, A.S., Xia, Z., Azurmendi, H.F., Saraswat, V., Legler, P.M., Massiah, M.A., Gabelli, S.B., Bianchet, M.A., Kang, L.W. and Amzel, L.M. (2005) Structures and mechanisms of Nudix hydrolases. *Arch. Biochem. Biophys.* **433**, 129–143.
- Nakamura, T., Meshitsuka, S., Kitagawa, S. et al. (2010) Structural and dynamic features of the MutT protein in the recognition of nucleotides with the mutagenic 8-oxoguanine base. *J. Biol. Chem.* **285**, 444–452.
- Nyati, P., Nouzova, M., Rivera-Perez, C., Clifton, M.E., Mayor, J.G. and Noriega, F.G. (2013) Farnesyl phosphatase, a *Corpora allata* enzyme involved in juvenile hormone biosynthesis in *Aedes aegypti*. *PLoS One*, **8**, e71967.
- Ogawa, T., Ueda, Y., Yoshimura, K. and Shigeoka, S. (2005) Comprehensive analysis of cytosolic Nudix hydrolases in *Arabidopsis thaliana*. *J. Biol. Chem.* **280**, 25277–25283.
- Ogawa, T., Yoshimura, K., Miyake, H., Ishikawa, K., Ito, D., Tanabe, N. and Shigeoka, S. (2008) Molecular characterization of organelle-type Nudix hydrolases in *Arabidopsis*. *Plant Physiol.* **148**, 1412–1424.
- Parveen, I., Wang, M., Zhao, J. et al. (2015) Investigating sesquiterpene biosynthesis in *Ginkgo biloba*: molecular cloning and functional characterization of (*E*, *E*)-farnesol and α -bisabolene synthases. *Plant Mol. Biol.* **89**, 451–462.
- Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45.
- R Core Team. (2015) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raymond, O., Gouzy, J., Just, J. et al. (2018) The *Rosa* genome provides new insights into the domestication of modern roses. *Nat. Genet.* **50**, 772–777.
- Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
- Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25.
- Roccia, A., Hibrand-Saint Oyant, L., Cavel, E. et al. (2019) Biosynthesis of 2-phenylethanol in rose petals is linked to the expression of one allele of *RhPAAS*. *Plant Physiol.* **179**, 1064–1079.
- Rusdi, N.A., Goh, H.H., Sabri, S., Ramzi, A.B., Mohd Noor, N. and Baharum, S.N. (2018) Functional characterisation of new sesquiterpene synthase from the Malaysian herbal plant, *Polygonum minus*. *Molecules*, **23**, 1370.
- Šali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
- Schnee, C., Kollner, T.G., Gershenzon, J. and Degenhardt, J. (2002) The maize gene *terpene synthase 1* encodes a sesquiterpene synthase catalyzing the formation of (*E*)- β -farnesene, (*E*)-nerolidol, and (*E*, *E*)-farnesol after herbivore damage. *Plant Physiol.* **130**, 2049–2060.
- Schrödinger, L.L.C. (2015) The PyMOL Molecular Graphics System, Version 1.8. <https://pymol.org>.
- Schwab, W., Davidovich-Rikanati, R. and Lewinsohn, E. (2008) Biosynthesis of plant-derived flavor compounds. *Plant J.* **54**, 712–732.
- Setoyama, D., Ito, R., Takagi, Y. and Sekiguchi, M. (2011) Molecular actions of *Escherichia coli* MutT for control of spontaneous mutagenesis. *Mutat. Res.* **707**, 9–14.
- Shalit, M., Shafir, S., Larkov, O. et al. (2004) Volatile compounds emitted by rose cultivars: fragrance perception by man and honeybees. *Israel J. Plant Sci.* **52**, 245–255.
- Shen, M.Y. and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524.
- Simkin, A.J., Miettinen, K., Claudel, P. et al. (2013) Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpene branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry*, **85**, 36–43.
- Smulders, M.J.M., Arens, P., Bourke, P.M. et al. (2019) In the name of the rose: a roadmap for rose research in the genome era. *Hortic. Res.* **6**, 65.
- Srouji, J.R., Xu, A., Park, A., Kirsch, J.F. and Brenner, S.E. (2017) The evolution of function within the Nudix homology clan. *Proteins*, **85**, 775–811.
- Sun, P., Schuurink, R.C., Caissard, J.C., Huguene, P. and Baudino, S. (2016) My way: noncanonical biosynthesis pathways for plant volatiles. *Trends Plant Sci.* **21**, 884–894.
- Svensson, L.M., Jemth, A.-S., Desroses, M., Loseva, O., Helleday, T., Högbom, M. and Stenmark, P. (2011) Crystal structure of human MTH1 and the 8-oxo-dGMP product complex. *FEBS Lett.* **585**, 2617–2621.
- Van Ooijen, J.W. (2004) MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. Wageningen, The Netherlands: Plant Research International.
- Van Ooijen, J.W. (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Wageningen, The Netherlands: Plant Research International.
- Vanommeslaeghe, K. and MacKerell, A.D. Jr (2012) Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3144–3154.

Diversification in the Nudix hydrolase gene family 199

- Voinnet, O., Rivas, S., Mestre, P. and Baulcombe, D.** (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**, 949–956.
- Winter, G., Waterman, D.G., Parkhurst, J.M. et al.** (2018) DIALS: implementation and evaluation of a new integration package. *Acta Cryst. D*, **74**, 85–97.
- Yang, T., Li, J., Wang, H. and Zeng, Y.** (2005) A geraniol-synthase gene from *Cinnamomum tenuipilum*. *Phytochemistry*, **66**, 285–293.
- Yoshimura, K., Ogawa, T., Ueda, Y. and Shigeoka, S.** (2007) *AtNUDX1*, an 8-oxo-7,8-dihydro-2'-deoxyguanosine 5'-triphosphate pyrophosphohydrolase, is responsible for eliminating oxidized nucleotides in *Arabidopsis*. *Plant Cell Physiol.* **48**, 1438–1449.

General outline chapter 2:

Article 2 “Duplication and specialization of *NUDX1* in *Rosaceae* led to geraniol production in rose petals”

Corentin CONART, Nathanaelle SACLIER, Fabrice FOUCHER, Clément GOUBERT, Aurélie RIUS-BONY, Saretta N. PARAMITA, Sandrine MOJA, Tatiana THOUROUDE, Christophe DOUADY, Pulu SUN, Baptiste NAIRAUD, Denis SAINT-MARCOUX, Muriel BAHUT, Julien JEAUFFRE, Laurence HIBRAND SAINT-OYANT, Robert C. SCHUURINK, Jean-Louis MAGNARD, Benoît BOACHON, Natalia DUDAREVA, Sylvie BAUDINO, Jean-Claude CAISSARD

Molecular Biology and Evolution, Volume 39, Issue 2, February 2022, msac002



<https://doi.org/10.1093/molbev/msac002>

Published: 12 January 2022

After finding that *NUDX1* enzymes could be involved in the biosynthesis of different acyclic terpenes, and that *NUDX1* copies were numerous in the genome of garden roses in Article 1, several questions arose about their duplications: what is the organisation of *NUDX1* copies in the genome and what is the ancestral copy? When had they evolved, during domestication, or during evolution of some *Rosaceae*? What could explain the petal specificity of the expression? This Article 2 gives some answers. Indeed, *NUDX1* copies were mapped on the available genomes of OB and other *Rosaceae* (GDR, Jung *et al.* 2019), and their organisation in clusters has been verified by MinION sequencing. After this, a collection of more than 30 wild roses was used to investigate evolution of these copies in the genus *Rosa*, their expression and their duplication. It was then possible to conclude that geraniol production was due to a cluster of *NUDX1-1a* copies on chromosome 2. Finally, we also mapped transposable elements and interpreted their possible involvement in ancestral *NUDX1* duplications, and their role in *NUDX1-1a* petal specificity expression. We concluded with an evolutive scenario of the *NUDX1* gene family in *Rosaceae*.

In this work, I verified the mapping of *NUDX1* copies in *Rosaceae* by blastn on the online genomes (OB, *R. multiflora*, *F. vesca*, *P. micrantha*, *P. persica*, *M. x domestica*), and by MinION sequencing of OB, *R. moschata*, and *R. laevigata*. I also mapped all the transposable elements in these regions. I made all the qRT-PCR and qPCR experiments. I cloned the *NUDX1-1a* promoter, and made the GFP constructs and transient expression experiments. Finally, I studied the structure of the *NUDX1-1a* promoter in several wild species.

Duplication and Specialization of *NUDX1* in *Rosaceae* Led to Geraniol Production in Rose Petals

Corentin Conart,¹ Nathanaelle Saclier,² Fabrice Foucher ³, Clément Goubert,⁴ Aurélie Rius-Bony,¹ Saretta N. Paramita,¹ Sandrine Moja,¹ Tatiana Thouroude,³ Christophe Douady,^{2,5} Pulu Sun,⁶ Baptiste Nairaud,¹ Denis Saint-Marcoux,¹ Muriel Bahut,⁷ Julien Jeauffre,³ Laurence Hibrand Saint-Oyant,³ Robert C. Schuurink,⁶ Jean-Louis Magnard,¹ Benoît Boachon,¹ Natalia Dudareva,^{8,9} Sylvie Baudino ^{*},¹ and Jean-Claude Caissard¹

¹Université Lyon, Université Saint-Etienne, CNRS, UMR 5079, Laboratoire de Biotechnologies Végétales appliquées aux Plantes Aromatiques et Médicinales, Saint-Etienne, France

²Université Lyon, Université Claude Bernard Lyon 1, CNRS, UMR 5023, ENTPE, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, Villeurbanne, France

³Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, Angers, France

⁴Department of Human Genetics, McGill University Genome Center, Montreal, QC, Canada

⁵Institut Universitaire de France, Paris, France

⁶Green Life Sciences Research Cluster, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

⁷Univ Angers, SFR QUASAV, Angers, France

⁸Department of Biochemistry, Purdue University, West Lafayette, IN, USA

⁹Purdue Center for Plant Biology, Purdue University, West Lafayette, IN, USA

*Corresponding author: E-mail: Sylvie.Baudino@univ-st-etienne.fr.

Associate editor: Michael Purugganan

Abstract

Nudix hydrolases are conserved enzymes ubiquitously present in all kingdoms of life. Recent research revealed that several Nudix hydrolases are involved in terpenoid metabolism in plants. In modern roses, RhNUDX1 is responsible for formation of geraniol, a major compound of rose scent. Nevertheless, this compound is produced by monoterpene synthases in many geraniol-producing plants. As a consequence, this raised the question about the origin of RhNUDX1 function and the *NUDX1* gene evolution in *Rosaceae*, in wild roses or/and during the domestication process. Here, we showed that three distinct clades of *NUDX1* emerged in the *Rosoideae* subfamily (Nudx1-1 to Nudx1-3 clades), and two subclades evolved in the *Rosa* genus (Nudx1-1a and Nudx1-1b subclades). We also showed that the Nudx1-1b subclade was more ancient than the Nudx1-1a subclade, and that the *NUDX1-1a* gene emerged by a *trans*-duplication of the more ancient *NUDX1-1b* gene. After the transposition, *NUDX1-1a* was *cis*-duplicated, leading to a gene dosage effect on the production of geraniol in different species. Furthermore, the *NUDX1-1a* appearance was accompanied by the evolution of its promoter, most likely from a *Copia* retrotransposon origin, leading to its petal-specific expression. Thus, our data strongly suggest that the unique function of *NUDX1-1a* in geraniol formation was evolved naturally in the genus *Rosa* before domestication.

Key words: *Rosaceae*, *Rosa*, Nudix hydrolase, monoterpenes, *NUDX1* synteny.

Introduction

Rosa is a complex taxon with more than 150 intertwined species (Wissemann 2003). Only few (around 15) rose species have been domesticated by humans since Antiquity (fig. 1). In Knossos (1700 B.C.), roses were painted with only few petals like wild briars (fig. 1a), whereas in Rome and Pompei (79 A.C.) they were presented with dozens of petals (fig. 1b), meaning that the domestication process had already started. Indeed, over the past three centuries, domestication resulted in flowers with hundreds of petals often with a strong

fragrance. Some of the very ancient roses, approximately 1,000–2,000 years old, have come down to us as heritage roses (fig. 1c). This includes *Rosa chinensis* cv. “Old Blush” (Old Blush) from China, which is likely a natural hybrid between wild species (Raymond et al. 2018). This rose has been largely used by breeders, and many modern roses probably have Old Blush as an ancestor. Other heritage roses have also been used for horticultural selection and hybridization with other varieties (supplementary table S1, Supplementary Material online). As a result, modern roses are an extended combination

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

between alleles of different wild species, and alleles that appeared by spontaneous bud mutations.

One of the most important traits attracting humans to roses is their pleasant fragrance. Geraniol is one of the rose scent constituents, which contributes to the flower rosy note. In contrast to most plants, formation of this monoterpene in modern roses does not rely on a canonical biosynthetic pathway (Magnard et al. 2015) that involves a plastidial monoterpene synthase (Sun et al. 2016). Instead, a cytosolic Nudix hydrolase (RhNUDX1) converts geranyl diphosphate (GPP) to geranyl phosphate (GP), which in turn is dephosphorylated by uncharacterized phosphatase to geraniol.

Nudix hydrolases are conserved enzymes hydrolyzing nucleoside diphosphates linked to some moiety X. They are ubiquitously present in all kingdoms of life and were proposed to function as housecleaning enzymes involved in cell sanitation (McLennan 2013; Yoshimura and Shigeoka 2015; Srouji et al. 2017). However, recent research revealed that Nudix hydrolases can be involved in terpenoid metabolism in plants (Magnard et al. 2015; Henry et al. 2018; Li et al. 2020; Sun et al. 2020). Indeed, *Arabidopsis thaliana* Nudix hydrolase 1 (AtNUDX1) together with an isopentenyl kinase coordinately regulates the isopentenyl diphosphate amount destined for higher-order terpenoid biosynthesis (Henry et al. 2015, 2018). Although AtNUDX1 is also able to efficiently dephosphorylate GPP and farnesyl diphosphate (FPP) in vitro, no geraniol nor (*E,E*)-farnesol was detected in this species (Chen et al. 2003). In contrast, RwnNUDX1-2 from a cultivated hybrid of *R. wichurana* hydrolyzes specifically cytosolic FPP into farnesyl phosphate en route to (*E,E*)-farnesol formation (Sun et al. 2020). The fact that members of NUDX1 family could have diverse functions in different species raises the question about RhNUDX1 evolution, whether it is present only in cultivated modern roses, or was already evolved in wild *Rosa* and/or *Rosaceae* species.

Here, we investigated the origin of RhNUDX1 function. We analyzed the evolution of all *NUDX1* gene homologs, their genomic localization and synteny by comparing the recently published genomes of Old Blush (Hibrand Saint-Oyant et al. 2018; Raymond et al. 2018) and several closely related genomes in the *Rosaceae* family (fig. 1c). We also examined the transposable elements (TEs) surrounding these genes and proposed an evolutionary scenario of duplication and specialization of *NUDX1-1a*, the gene encoding the Nudix hydrolase responsible for the GPP hydrolysis in rose petals.

Results

RcNUDX1 Is Present in Multiple Copies in Old Blush, but Only *RcNUDX1-1a* Is Highly Expressed in Its Petals

Discovery of terpene synthase-independent pathway for geraniol biosynthesis in modern roses and the involvement of RhNUDX1 in its formation (Magnard et al. 2015) raised the question of how this trait was evolved. Thus, we have isolated the corresponding genomic sequence from *R. x hybrida* cv. 'Papa Meilland', which revealed that *RhNUDX1* contains a single intron (*RhNUDX1-rs* for reference sequence). This sequence was used for phylogenetic analysis of *NUDX1* genes in

Rosaceae family. A maximum likelihood tree (ML tree) rooted with the *A. thaliana* homolog, AtNUDX1, was constructed using genomic sequences of Old Blush, *Fragaria vesca*, *Malus x domestica*, and *Prunus persica*, available in the Genome Database for *Rosaceae* (GDR, www.rosaceae.org [Jung et al. 2019]; supplementary table S2, Supplementary Material online) as well as recently published *R. x wichurana* sequences (Sun et al. 2020) (fig. 2). For the readability of the ML tree, we did not use Old Blush sequences that were 100% identical between them (supplementary table S3, Supplementary Material online).

This ML tree revealed three well-resolved at nearly all node clades, numbered Nudx1-1 to Nudx1-3, and a lesser-supported clade named Nudx1-4. Two sequences (*Prupe.1G302800* and *MD13G1049100*) could not be assigned to a clade, and appeared on branches with low bootstraps. Interestingly, these branches and the Nudx1-4 clade include exclusively sequences of *M. x domestica* and *P. persica*, whereas the three other clades contain all the sequences of Old Blush, *R. x wichurana* and *F. vesca*. As *M. x domestica* and *P. persica* belong to *Amygdaloideae* subfamily and *Rosa* species and *F. vesca* belong to *Rosoideae* subfamily (Xiang et al. 2017) (fig. 1c), it suggests that duplications of the first ancestral *NUDX1* ortholog led to divergent sequences in the Nudx1-4 clade in *Amygdaloideae*, but to homologous sequences in well-supported Nudx1-1 to Nudx1-3 clades in *Rosoideae*. *RhNUDX1-rs*, which is involved in geraniol production in horticultural roses, was found in the Nudx1-1 clade. This clade also encompasses closely related *RcNUDX1-1* sequences from Old Blush with 97.1–97.6% identity to the reference *RhNUDX1-rs* (supplementary table S2, Supplementary Material online), indicating that they could be the result of very recent duplications of the same gene.

To gain insights in the evolution of these paralogs, we analyzed their genomic organization in three Old Blush genomes published in the GDR (supplementary table S2, Supplementary Material online). We also sequenced the Old Blush accession using MinION technology (supplementary table S5, Supplementary Material online). This technology increases the error rate in sequences, but allows to obtain very long reads without informatics assembly (Lu et al. 2016), thus to verify gene clusters on chromosomes 2 and 4, and also to detect alleles and null alleles on homologous chromosomes. Comparison of all these sequences allowed to draw a comprehensive map in Old Blush (fig. 3), and a synteny map in *Rosaceae* (fig. 4). Two clusters containing *NUDX1* paralogs were found in Old Blush genome. The first cluster on chromosome 4 included the more ancient gene, *RcNUDX1-3*, along with one copy of both *RcNUDX1-1b* and *RcNUDX1-2a*, but a pseudogene ^ψ*RcNUDX1-2a* with two STOP codons on the other homologous chromosome 4. The second cluster was on chromosome 2 and contained four nearly identical copies of *RcNUDX1-1a* and one pseudogene ^ψ*RcNUDX1-1a* with a STOP codon. The four copies are nearly identical showing 98.7% of DNA identities and 96.8–99.0% of protein identities (supplementary tables S3 and S4, Supplementary Material online). Surprisingly, the *RcNUDX1-1a* genes were totally absent on a second homologous chromosome 2,

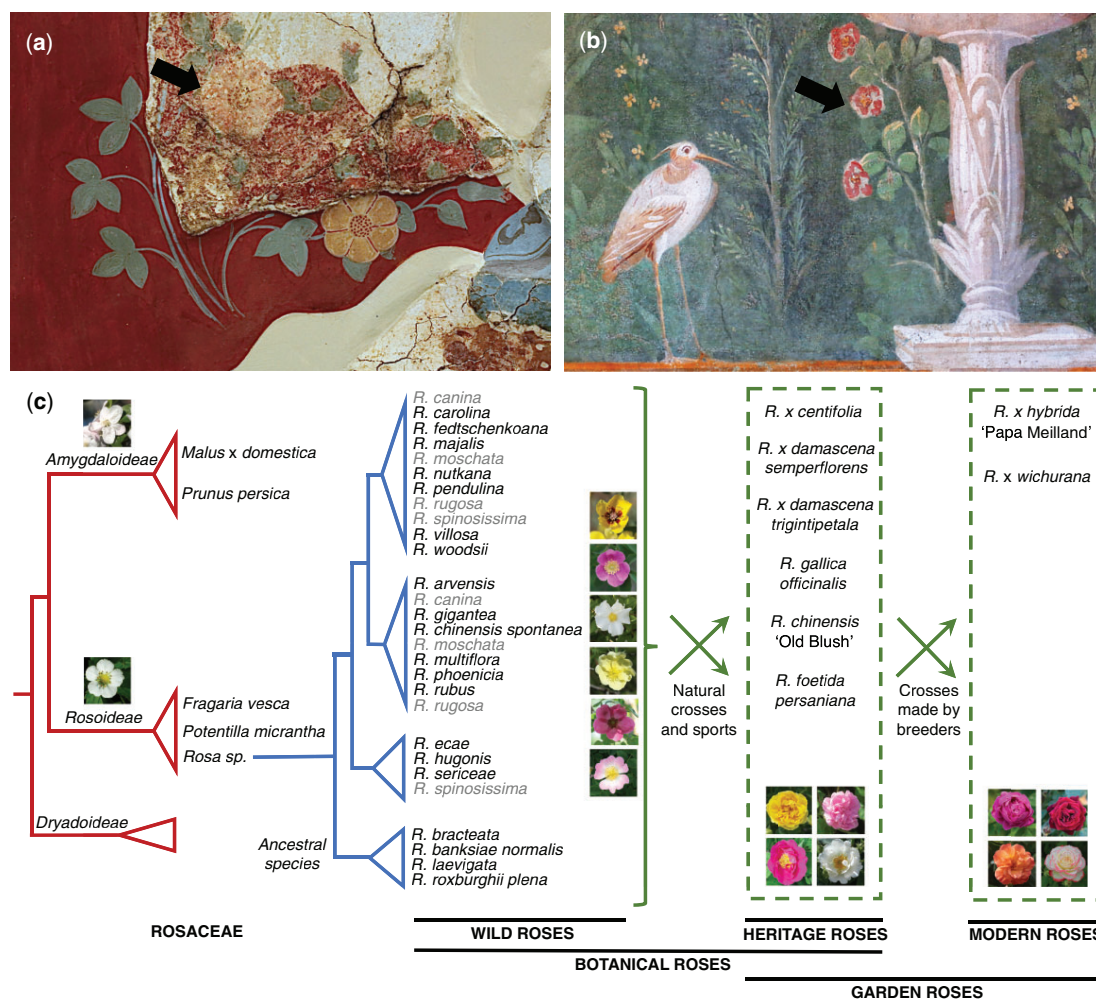


Fig. 1. Overview of the evolution of the *Rosaceae* family and of the *Rosa* genus. (a) Antique murals in Knossos (~1,700 B.C.). Arrow shows the original drawing of a wild rose (the other drawing was made during an irreversible restoration). (b) Antique murals in Pompei (~79 A.C.). Roses were painted with dozens of petals (arrow). (c) Synthetic phylogeny and evolution diagram obtained by simplification of data from Fougère-Danezan et al. (2015), Zhu et al. (2015), Xiang et al. (2017), Zhang et al. (2017), and Debray et al. (2019). Only species and varieties used or cited in our article are shown (supplementary table S1, Supplementary Material online). Some species are written in gray because their phylogenetic position is discussed (*R. moschata*, *R. rugosa*), or because they are allopolyploids (*R. canina*, *R. spinosissima*). *Rosa foetida* and *R. stellata mirifica* are not shown because of their unresolved position. Heritage roses also include some crosses made by breeders, which are not considered as botanical roses, and which are not shown here.

which thus correspond to a null allele ^{na}*RcNUDX1-1a*. Copies of *NUDX1-2* (*RcNUDX1-2b* and *RcNUDX1-2c*) were also found on chromosomes 6 and 7, respectively.

Comparisons of the two *NUDX1* clusters and the surrounding genes of the other *Rosaceae* (fig. 4; supplementary table S2, Supplementary Material online) revealed that the possible ancestral gene *NUDX1-3* has duplicated on chromosome 4 thus separating *Amygdaloideae* and *Rosoideae* subfamilies, and giving, respectively, sequences of the *Nudx1-4* clade, and *Nudx1-1* and *Nudx1-2* clades. Indeed, they were in the same microsyntenic region (fig. 4a). Furthermore, the two unresolved sequences *Prupe.1G302800* and

MD13G1049100 (fig. 2) were close to the homolog of the marker gene F, in similar position to *RcNUDX1-3* and its orthologs in *F. vesca* and *Potentilla micrantha*, implying that the ancestral gene had highly diverged between *Rosoideae* and *Amygdaloideae*. The other cluster, with the *RcNUDX1-1a* copies, was unique to Old Blush, indicating that it likely had evolved in very ancient roses at the beginning of the domestication process or in wild ancestors of Old Blush (fig. 4b).

Our previous RNAseq, QTL and correlation analyses (Magnard et al. 2015; Sun et al. 2020) performed mainly on modern hybrid roses, showed that *RcNUDX1-1a* was expressed in petals and responsible for the geraniol

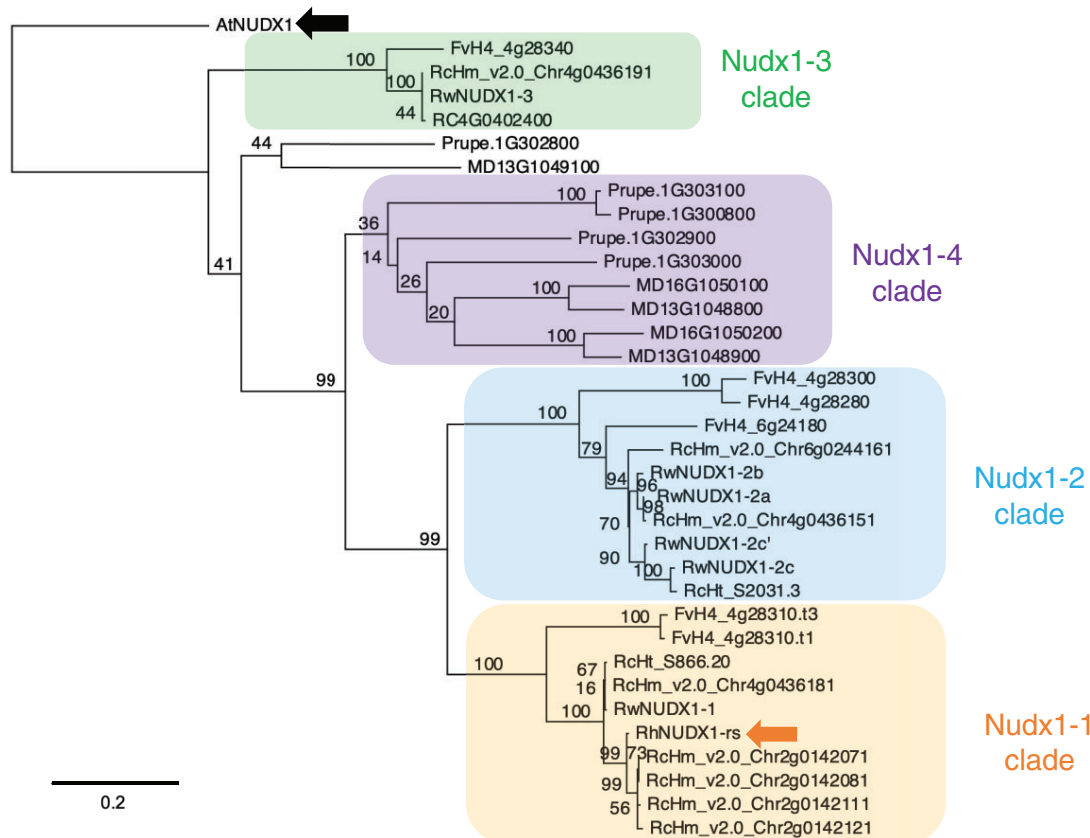


Fig. 2. ML tree of genomic sequences of *NUDX1* homologs in the *Rosaceae*. The tree was made with sequences of Sun et al. (2020), and with sequences obtained by BlastN (from ATG to STOP including the intron) in selected species of the GDR ([Align_Rosaceae_MLtree.fasta](#), [Supplementary Material](#) online). *AtNUDX1* gene was used to root the tree (large black arrow). *RhNUDX1-rs* was added for reference (large orange arrow). Clades were named according to Sun et al. (2020). Numbers correspond to bootstraps (%). Scale bar represents substitution per site.

production. On the other hand, the *RcNUDX1-1b* protein was active in vitro, but the *RcNUDX1-1b* gene was not expressed. We verified that it was also the case in a wild species by checking the in vitro activities of *RmNUDX1-1a* and *RmNUDX1-1b* proteins of the *Moschata* accession. These activities were quite similar to those of the corresponding Old Blush enzymes ([supplementary table S6](#), [Supplementary Material](#) online), suggesting that only the gene expression could be responsible of geraniol production in wild species. Thus, to determine whether the other *RcNUDX1-1a* homologs, *RcNUDX1-1b*, *RcNUDX1-2*, and *RcNUDX1-3*, were expressed in petal tissue, qRT-PCR analyses with gene-specific primers were performed ([supplementary table S7](#), [Supplementary Material](#) online). These analyses revealed that only *RcNUDX1-1a* transcripts indeed accumulate at high levels in Old Blush petals (60,000x more than *RcNUDX1-1b*), thus further suggesting that such mode of expression is rose specific and uniquely clustered *RcNUDX1-1a* paralogs are involved in the biosynthesis of geraniol ([supplementary fig. S1](#), [Supplementary Material](#) online).

Taken together, these results support that the *NUDX1-3* ancestral orthologs were duplicated many times in the *Rosaceae*. The ortholog was probably an ortholog of *AtNUDX1* that had likely the same function. Although genes within the Nudx1-1 and Nudx1-2 clades evolved in the subfamily *Rosoideae*, the *NUDX1-1a* paralogs emerged only in the genus *Rosa*. In addition, the high sequence similarity of the clustered *RcNUDX1-1a* paralogs with the characterized *RhNUDX1-rs*, as well as high level of expression, suggest that these paralogs are involved in the biosynthesis of geraniol in Old Blush. The presence of ^{na}*RcNUDX1-1a* opens the possibility that one of the potential wild parents of Old Blush did not have such cluster, and therefore the duplication of *RcNUDX1-1a* had occurred in wild species of the genus *Rosa*.

The *NUDX1-1a* Paralogs Are Specific to Wild Roses Producing Geraniol

To determine whether *RcNUDX1-1a* had already arisen in wild species of *Rosa* or evolved early during the domestication process, we performed GC-MS metabolic profiling of the

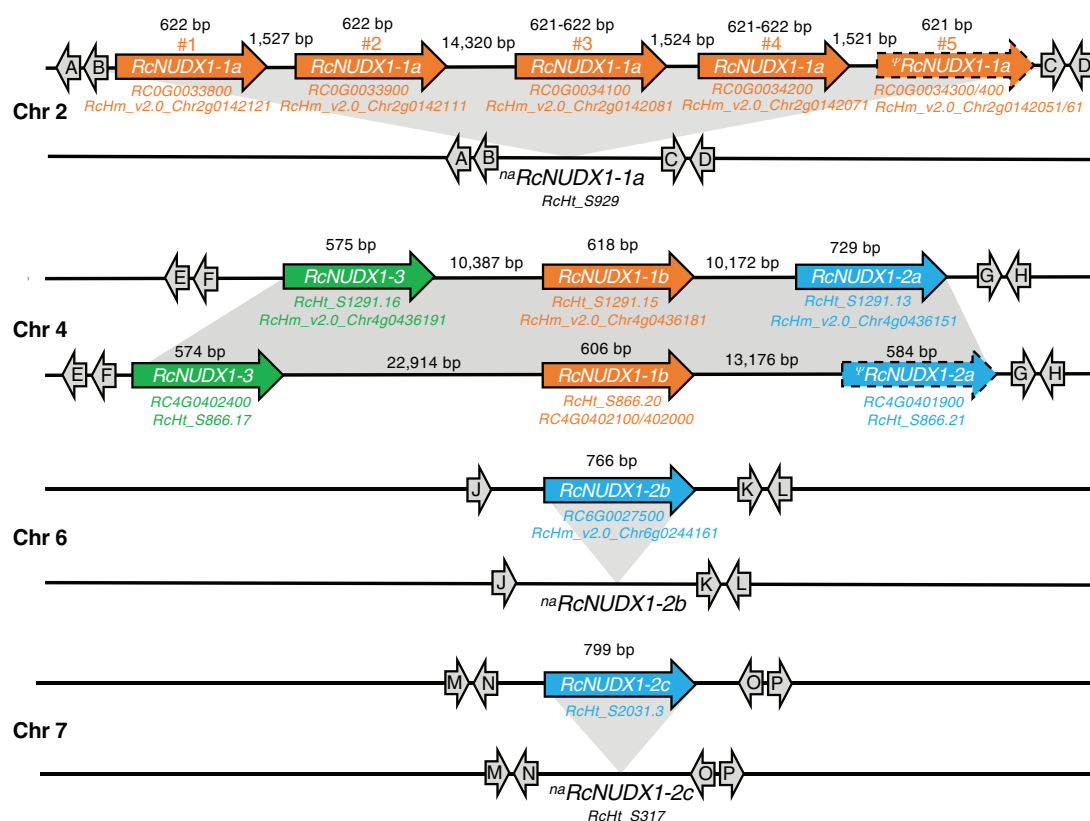


FIG. 3. Gene map of *RcNUDX1* in Old Blush. Each pair of homologous chromosomes are shown. Similar regions including *RcNUDX1* sequences are highlighted in gray between the two homologous chromosomes. Gene lengths, from the ATG codon to the STOP, including introns, and intergenic lengths are indicated. However, the picture does not respect these lengths. Gene numbers were obtained by making a systematic inventory of chromosomes on the three genomes of Old Blush published in the GDR and by comparison with our MinION long reads (supplementary tables S2, S5, and Align_OldBlush_DNAsequences.fasta, Supplementary Material online), but only sequence accessions useful for mapping are shown. Null alleles were confirmed on chromosomes 2 and 7 because scaffolds available in the GDR including both upstream and downstream regions were found. All null alleles were also confirmed by MinION sequencing (supplementary table S5, Supplementary Material online). Large orange arrows, genes from Nudx1-1 clade; large blue arrows, genes from Nudx1-2 clade; large green arrows, genes from Nudx1-3 clade. Copies of *RcNUDX1-1a* are arbitrarily numbered in orange on chromosome 2. Sequences with a dashed outline are pseudogenes including STOP codons. Chr, chromosomes. Marker genes (gray arrows) used for microsynteny are listed in supplementary table S14, Supplementary Material online. On chromosome 2, gene D was not found on scaffold *RcHt_S929* but useful in MinION reads. On chromosome 6, marker genes were not found around the null allele in the GDR, but MinION long reads included marker genes J, K, L, and *RcNUDX1-2b*, or its null allele (read numbers in supplementary Table S5, Supplementary Material online).

volatiles produced by petals along with analysis of the *RcNUDX1-1* homologs in a collection of 29 accessions of wild roses and six accessions of heritage roses (supplementary tables S1 and S8, Supplementary Material online). Their genomic DNAs and mRNAs were used to isolate and characterize full-length *NUDX1-1* sequences (table 1; supplementary table S9, Supplementary Material online). Due to the high sequence identity (89.5–91.6%, supplementary table S3, Supplementary Material online) between *RcNUDX1-1a* and *RcNUDX1-1b*, the primers were designed based on Old Blush sequences to amplify the region from ATG to STOP codons (supplementary table S7, Supplementary Material online). Sequencing of the obtained PCR products revealed that the

primers were specific for Nudx1-1 clade and did not amplify sequences of the Nudx1-2 and Nudx1-3 clades.

cDNAs were obtained from all species that emit geraniol except for *R. sericea* producing a very small amount of this compound (supplementary tables S8 and S9, Supplementary Material online). We also cloned cDNAs from *R. rubus* that does not produce geraniol, but these cDNAs were as close to *RcNUDX1-1a* as to *RcNUDX1-1b*. For most of the accessions, several genomic sequences (gDNA) of *NUDX1-1* were obtained. However, numerous gDNAs were attained for some species due to the ploidy level (table 1; see supplementary table S1 for ploidy levels, Supplementary Material online) and two species have only a single gDNA. Interestingly, in

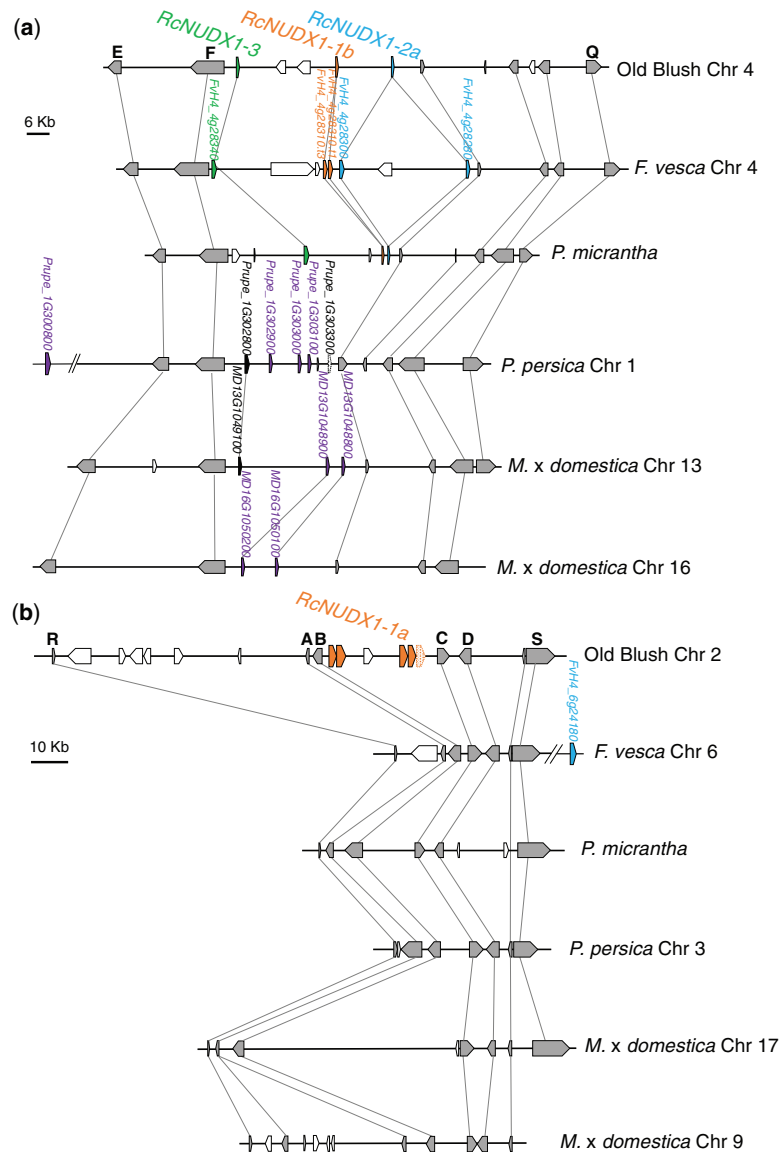


FIG. 4. Synteny map of the *Rosaceae* genomes. (a) Microsynteny of chromosome 4 of Old Blush in the cluster region of *RcNUDX1-1b*, *RcNUDX1-2a*, and *RcNUDX1-3*. (b) Microsynteny of chromosome 2 of Old Blush in the cluster region of *RcNUDX1-1a*. Chromosome numbers are indicated except for *P. micrantha* for which the genome was nonassembled in the GDR (supplementary table S2, Supplementary Material online). Large orange arrows, genes from Nudx1-1 clade; large blue arrows, genes from Nudx1-2 clade; large green arrows, genes from Nudx1-3 clade; large violet arrows, sequences of the Nudx1-4 clade; large black arrows, other *NUDX1* genes; large white arrows, unique genes; large gray arrows, genes used for microsynteny (marker genes are listed in supplementary table S14, Supplementary Material online). Accession numbers of *NUDX1* genes are in supplementary table S2, Supplementary Material online. There was no sequence of *NUDX1* in the microsyntenic regions of chromosomes 6 and 7. Distances between sequences and scales are approximative, and gene lengths are distorted to show the relative organization. Chr, chromosomes.

R. rubus, no *NUDX1-1* gDNA corresponding to the isolated cDNAs was detected.

All identified gDNAs contained one intron of variable size (supplementary table S2, Supplementary Material online),

and clustered in two groups on the ML tree (fig. 5; supplementary fig. S2, Supplementary Material online). The first group included the Old Blush *RcNUDX1-1a*, and thus was named Nudx1-1a subclade (orange names in supplementary

Table 1. Comparison of Geraniol Concentration and Expression of *NUDX1-1* Homologs in Wild and Heritage Roses.

| Accession Names ^a | Geraniol Concentration ($\mu\text{g/gFW}$) | qRT-PCR on <i>NUDX1-1</i> Homologs (a.u.) ^b | Number of cDNA Clones ^c | Number of gDNA ^d Clones ^c |
|------------------------------|---|---|------------------------------------|---|
| Arvensis_B | 0.0 (0.0) ^e | 0.1 (0.1) ^e | 0 | 2 |
| Banksiae | 0.0 (0.0) | 0.0 (0.0) | 0 | 4 |
| Bracteata | 0.0 (0.0) | 0.0 (0.0) | 0 | 1 |
| Chinensis | 0.0 (0.0) | 0.0 (0.0) | 0 | 2 |
| Gigantea | 0.0 (0.0) | 0.0 (0.0) | 0 | 2 |
| Laevigata | 0.0 (0.0) | 0.0 (0.0) | 0 | 1 |
| Mirifica | 0.0 (0.0) | 0.0 (0.0) | 0 | 5 |
| Roxburghii | 0.0 (0.0) | 0.0 (0.0) | 0 | 4 |
| Rubus | 0.0 (0.0) | 0.7 (0.0) | 3 | 6 |
| Sericea | 0.8 (0.3) | 0.0 (0.0) | 0 | 6 |
| Foetida | 3.0 (2.1) | 13.3 (12.6) | 2 | 3 |
| Persian_Yellow | 5.4 (1.9) | 34.8 (30.3) | 1 | 6 |
| Ecae | 5.8 (0.2) | 0.0 (0.0) | 1 | 3 |
| Hugonis_B | 17.9 (2.3) | 0.0 (0.0) | nd ^f | 2 |
| Canina | 22.9 (6.0) | 111.5 (1.6) | 1 | 15 |
| Phoenicia | 27.0 (4.4) | 155.2 (12.2) | 1 | 7 |
| Moschata | 29.5 (10.2) | 111.6 (5.8) | 3 | 11 |
| Fedtschenkoana | 37.8 (5.9) | 87.2 (3.3) | 1 | 6 |
| Rugosa | 44.4 (24.2) | 36.1 (24.0) | 1 | 12 |
| Centifolia | 45.3 (17.2) | 207.1 (131.9) | 3 | 14 |
| Arvensis_A | 53.9 (52.8) | 256.8 (139.7) | 2 | 5 |
| Gallica_B | 63.4 (3.9) | 91.5 (20.2) | 1 | 3 |
| Autumn_Damask | 84.1 (11.6) | 63.1 (18.8) | 3 | 7 |
| Hugonis_A | 89.8 (31.4) | 12.5 (0.3) | nd | 4 |
| Nutkana | 96.3 (26.3) | 374.1 (129.2) | 2 | 6 |
| Old_Blush | 99.8 (5.9) | 61.0 (12.0) | 1 | 2 |
| Pendulina | 104.4 (45.4) | 174.8 (82.9) | 2 | 3 |
| Villosa | 107.9 (10.4) | 128.0 (6.0) | 1 | 5 |
| Gallica_A | 108.7 (11.6) | 88.3 (21.7) | 2 | 6 |
| Damask_Kazanlik | 112.2 (39.5) | 43.1 (1.6) | 3 | 9 |
| Majalis | 112.6 (2.0) | 25.3 (1.8) | 2 | 7 |
| Carolina | 145.7 (40.7) | 339.4 (100.7) | 3 | 6 |
| Woodsii | 180.4 (2.8) | 19.8 (3.0) | 2 | 5 |
| Offinalis | 192.1 (42.5) | 112.5 (8.3) | 1 | 5 |
| Spinosissima | nd | nd | 1 | 14 |

^aFor the rose accession names, see [supplementary table S1, Supplementary Material](#) online.

^bAmplification with FP8-RP8 primers ([supplementary table S7, Supplementary Material](#) online).

^cCloning with FP7-RP7 primers ([supplementary table S7, Supplementary Material](#) online).

^dThe number of gDNA clones correspond to different genomic sequences from ATG to STOP codons ([supplementary table S9](#) and [Clones_gDNAs_cDNAs.fasta, Supplementary Material](#) online). They all included a single intron ([Clones_IntronExonStructure.fasta, Supplementary Material](#) online).

^eValues correspond to averages, and SD are given in parentheses. Extensive values are given in [supplementary tables S8](#) and [S10, Supplementary Material](#) online.

^fNot done.

[fig. S2, Supplementary Material](#) online), whereas the second group, named Nudx1-1b subclade, included the gDNAs which were closer to *RcNUDX1-1b* than to *RcNUDX1-1a* (red names in [supplementary fig. S2, Supplementary Material](#) online). A BlastN analysis of all the gDNAs ([supplementary table S9, Supplementary Material](#) online) revealed that most of the gDNAs on the ML tree share 88.7–99.8% identity with both the *RcNUDX1-1a* and *RcNUDX1-1b* sequences. Thus, gDNAs displaying identity more than 1% higher with *RcNUDX1-1a* than with *RcNUDX1-1b* were assigned to the Nudx1-1a subclade and vice versa ([supplementary fig. S2, Supplementary Material](#) online). Few gDNAs were as close to *RcNUDX1-1a* as to *RcNUDX1-1b* because they exhibit <1% identity in favor to either of two subclades (shown in black in [supplementary fig. S2, Supplementary Material](#) online). These sequences were often distant from all other gDNAs (long black branches in [supplementary fig. S2,](#)

[Supplementary Material](#) online) and could have thus diverged in these particular species. Some of them were located at the root of the tree suggesting that they could represent *NUDX1-1* ancestral sequences.

In contrast to Nudx1-1a subclade, Nudx1-1b subclade included all the gDNAs from the species that do not produce geraniol (blue stars in [fig. 5](#) and [table 1; supplementary table S8, Supplementary Material](#) online). Unlike *NUDX1-1a* gDNAs, which were clearly absent in 8 accessions, *NUDX1-1b* gDNAs were undetectable only in 2 accessions ([supplementary table S9, Supplementary Material](#) online). The gDNAs of Nudx1-1b subclade were closer to the root of the phylogenetic tree than those of the Nudx1-1a subclade. Thus, despite weak branch support of the ML tree, these data suggest an ancestral origin of the *NUDX1-1b* genes.

All cloned cDNAs were found to correspond to the ORF sequence found only in gDNAs belonging to the Nudx1-1a

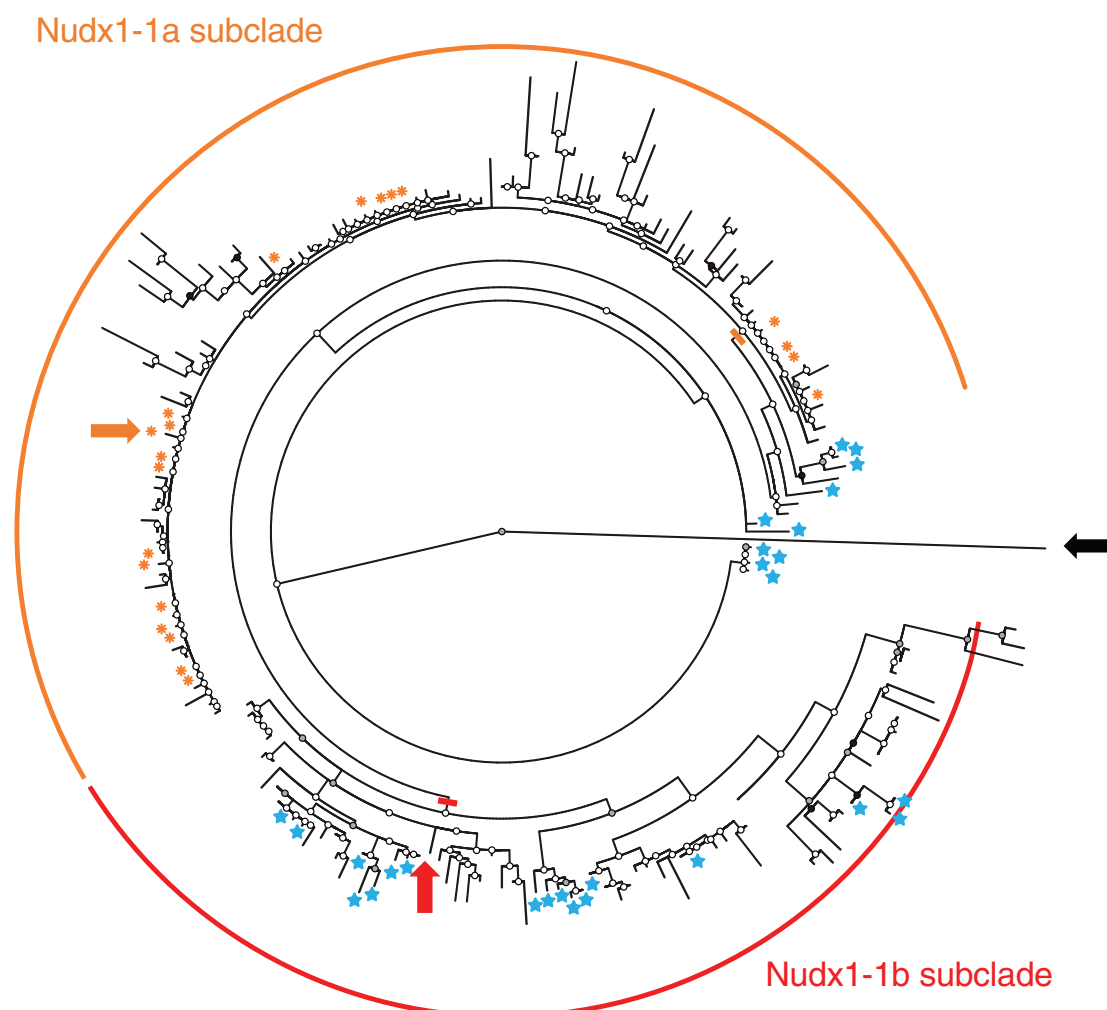


Fig. 5. ML tree of genomic sequences of the Nudx-1 clade. Orange asterisks indicate species in which a cDNA clone is the exact ORF of the gDNA (supplementary table S9 and Clones_IntronExonStructure.fasta, Supplementary Material online). Blue stars indicate species not producing geraniol (table 1; supplementary table S8, Supplementary Material online). Large orange and red arrows indicate, respectively, the *RcNUDX1-1a* and *RcNUDX1-1b* genes of Old Blush. White dots correspond to bootstraps <70%, gray dots, between 70% and 95%, and black dots, more than 95%. The tree is rooted with a sequence of *F. vesca* (large black arrow). For the extended tree see supplementary figure S2 and Align_OldBlush_MLtree.fasta, Supplementary Material online.

subclade (orange asterisks in fig. 5; supplementary table S9, Supplementary Material online), suggesting that only members of this clade are expressed. Next, we evaluated expression of *NUDX1-1* homologs in the petals of all 34 accessions (table 1; supplementary tables S1 and S10, Supplementary Material online) by qRT-PCR with consensus primers, which were capable of amplifying both *NUDX1-1a* and *NUDX1-1b* (supplementary table S7, Supplementary Material online). As no cDNAs belonging to the *NUDX1-1b* group were obtained, transcripts detected in this analysis correspond to *NUDX1-1a* homologs (table 1). *NUDX1-1* transcripts were barely detected in botanical species not producing geraniol. In contrast, *NUDX1-1* was expressed in all species producing geraniol

and for which genomic sequences corresponding to *NUDX1-1a* were obtained. The exceptions include two geraniol-producing species (accessions Hugonis B and Ecae) with very low *NUDX1-1* expression, and two low-geraniol producers (accessions Foetida and Persian Yellow) with substantial *NUDX1-1* expression (table 1). In the latter two species, low geraniol levels could be the result of substrate limitation, whereas in two former species another *NUDX1* homolog could be involved in geraniol production. We have recently shown the existence of specialization of different homologs as *RwNUDX1-2c* was active in *R. x wichurana*, but not in Old Blush (Sun et al. 2020). In botanical and heritage roses, *NUDX1-1a* expression was highly correlated (*P*-values <

0.001) with geraniol levels, as well as with the levels of acyclic monoterpenes (supplementary fig. S3 and table S11, Supplementary Material online). It was also positively correlated with the production of the acyclic sesquiterpenes (*E,E*)-farnesol, (*E,E*)- α -farnesene, and (*Z, E*)- α -farnesene as well as 2-phenylethanol. A negative correlation was found for 2-pentadecanone.

Thus, the presence of *NUDX1-1a* paralogs and its expression in some but not all botanical species as well as a positive correlation between *NUDX1-1a* expression and geraniol levels could indicate that the unique function of *NUDX1-1a* in geraniol production was evolved naturally in the genus *Rosa* before domestication.

Trans-Duplication of *NUDX1-1b* and Additional *cis*-Duplications Led to a *NUDX1-1a* Cluster in the Genus *Rosa*

Our data show that the ancestral *RcNUDX1-1b* gene homologs exist in many wild roses and in some other *Rosaceae* species, whereas *RcNUDX1-1a* homologs are only present in some wild roses mostly producing geraniol. This strongly suggested that *NUDX1-1a* homologs arose from *trans*-duplication of *NUDX1-1b* in wild roses, followed by *cis*-duplications on chromosome 2.

To understand the origin of the clustered *RcNUDX1-1a* paralogs on chromosome 2, we first performed a dot-plot analysis of nucleotide sequence similarity (supplementary fig. S4, Supplementary Material online). The identified repeated sequences (supplementary fig. S4a, Supplementary Material online) were then compared with the TEs annotated in the GDR (supplementary fig. S4b and table S12, Supplementary Material online) to draw a comprehensive map (fig. 6). This analysis revealed that all five copies of *RcNUDX1-1a* with their intergenic regions were nearly identical and contained the same TEs in the same order (fig. 6a). Each *NUDX1-1a* copy was surrounded by a fragment of the *Copia* R24588 retrotransposon (class I, RNA intermediate) at the 5'-end, and by two embedded Miniature Interspersed TEs (MITEs; Wicker et al. 2007) at the 3'-end (except for copy number 5). MITE G13554 itself was inserted into MITE P580.2030 (respectively, named in the GDR as *ms382250_RcHm_v2.0_Chr2_DXX-MITE_denovoRcHm_v2.0-B-G13554-Map6* and *ms580616_RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20*). The embedded MITEs in the second copy were interrupted by a long sequence containing genes, noncoding RNAs, and TEs (supplementary table S12, Supplementary Material online). Analysis of the four copies of these embedded MITEs revealed that they all have more than 80% of identity compared with their consensus sequences published in the GDR (supplementary table S12, Supplementary Material online), suggesting that the initial *RcNUDX1-1a* block may have then been duplicated in tandem after its initial insertion on chromosome 2.

To further analyze the origin of these block duplications, we searched for MITE G13554, MITE P580.2030, and *Copia* R24588 localizations around the *RcNUDX1* homologs on other chromosomes, and found two copies on chromosome

4 (supplementary table S12, Supplementary Material online). Analysis of available genomic sequences of the two rose haplotypes of the GDR revealed that *Copia* R24588 was absent on chromosome 4 of one annotated haplotype (Raymond et al. 2018), whereas it was found manually in the other (Hibrand Saint-Oyant et al. 2018). To compare the organization of the clusters on chromosomes 2 and 4 in different species, we also performed MinION sequencing of *Moschata* accession, which produces geraniol, and of *Laevigata* accession, an unscented rose species (supplementary table S13, Supplementary Material online). In *Moschata*, we found two copies of *RmNUDX1-1a* harboring the same organization of TEs as in Old Blush, but none in the accession *Laevigata* (fig. 6a). As *R. laevigata* is more ancient than *R. moschata*, which in turn is more ancient than *R. chinensis* cv. "Old Blush" (Fougère-Danezan et al. 2015; Debray et al. 2019), these results suggest that a series of duplications occurred during the evolution of the genus *Rosa*. Analysis of microsyntenic region of chromosome 4, that includes the cluster *RcNUDX1-3/RcNUDX1-1b/RcNUDX1-2a*, revealed a sequence *NUDX1-1b* directly upstream of the same MITE and *Copia* R24588 elements found in the chromosome 2 of Old Blush and *Moschata* (fig. 6b). Contrary to chromosome 2, the MITE P580.2030 was repeated in tandem and did not embed MITE G13554. The absence of the embedded MITE suggests that the *NUDX1* cluster on the chromosome 4 of Old Blush is a likely candidate for being the ancestral sequence from which *RcNUDX1-1a* blocks on chromosome 2 originate.

To determine whether in general *Rosa* species have multiple copies of *NUDX1-1a*, we estimated the copy number of *NUDX1-1* homologs in some wild roses using qPCR experiments on genomic DNA (Axelsson et al. 2013) (supplementary table S7, Supplementary Material online). Quantification was done for 12 wild species, and revealed that the number of *NUDX1-1a* copies ranged from three to ten in geraniol producing species and from two to five in species producing no geraniol (supplementary fig. S5, Supplementary Material online). These results clearly show that the number of *NUDX1-1* copies is indeed variable in rose species and overall higher in species producing geraniol.

Taken together, these results are consistent with a *trans*-duplication occurring in the genus *Rosa* between chromosome 4 and chromosome 2, and show that *NUDX1-1a* was a result of specialized duplication of *NUDX1-1b*. After this duplication, MITE G13554 was inserted into MITE P580.2030. The sequence block *Copia* R24588 *NUDX1-1a* with MITE P580.2030 [MITE G13554] at the beginning or at the end, was further duplicated in tandem in some wild roses producing geraniol.

Promoter Specificity and Gene Dosage Determine the High *NUDX1-1a* Expression Level in Petals

Our results indicate that the clustered *NUDX1-1a* paralogs arose from the duplication of the *NUDX1-1b* gene, which is not expressed in petals, raising the question of how tissue specificity and high levels of *NUDX1-1a* expression were achieved.

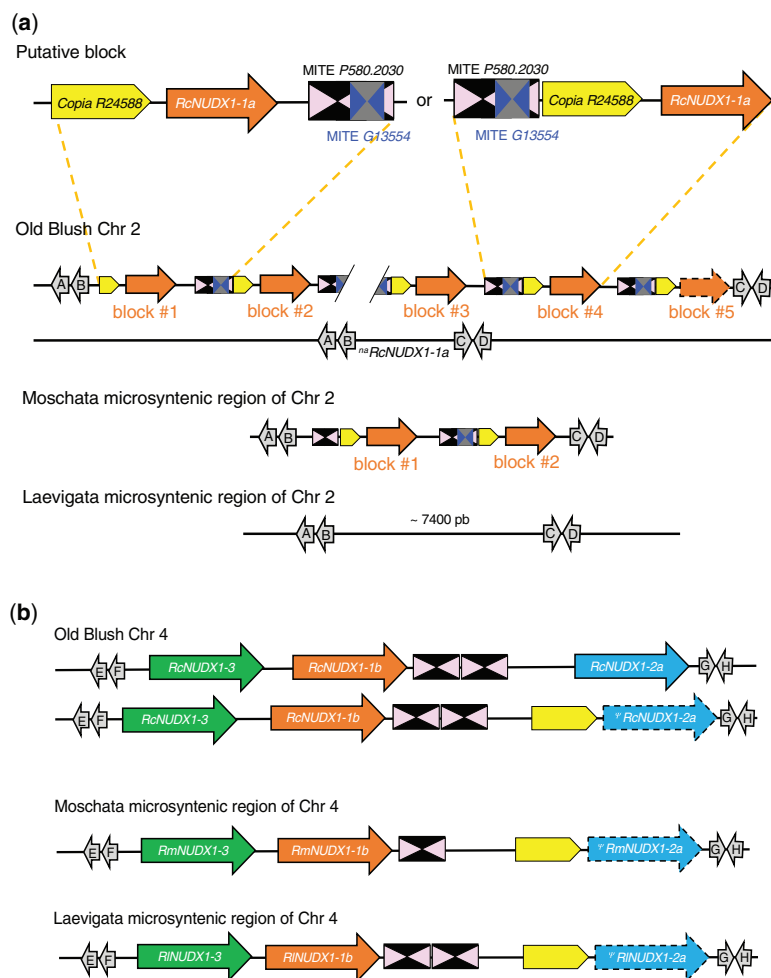


Fig. 6. Organization of the shared TEs around the *NUDX1-1a* and *NUDX1-1b* sequences in three accessions: Old Blush, Moschata, and Laevigata. (a) Chromosome 2 of Old Blush and corresponding microsyntenic regions of Moschata and Laevigata accessions. The cluster could be interpreted with two types of putative blocks (shown on top), which could then duplicate into five blocks. In the first hypothesis, MITEs are missing in block #5. In the second hypothesis, MITEs are missing in block #1. (b) Chromosome 4 of Old Blush and corresponding microsyntenic regions of Moschata and Laevigata accessions (MinION sequencing in [supplementary table S13, Supplementary Material](#) online). Only shared TEs are shown ([supplementary table S12, Supplementary Material](#) online). Large orange arrows, genes from *Nudx1-1* clade; large blue arrows, genes from *Nudx1-2* clade; large green arrows, genes of *Nudx1-3* clade; pink triangles, MITE P580.2030; dark blue triangles, MITE G13554; yellow arrow, *Copia* R24588; large gray arrows, marker genes used to find reads in the MinION database ([supplementary table S14, Supplementary Material](#) online). Distances between sequences are approximate and gene lengths and TE sizes are distorted to show the relative organization. Chr, chromosomes.

To answer this question, we first tested our hypothesis that a gene dosage affects *NUDX1-1a* expression in wild roses producing geraniol. Thus, we analyzed whether the number of *NUDX1-1* copies in the 13 already analyzed wild species ([supplementary fig. S5, Supplementary Material](#) online) correlates with the expression levels of *NUDX1-1* homologs ([table 1](#)). Indeed, the *NUDX1-1a* copy number positively correlated, although not linearly, with the expression of *NUDX1-1a* in rose petals ([fig. 7](#)). These results suggest that the number of duplication events leading to multiple copies of *NUDX1-1a* paralogs directly impacts its expression in petals. We did not try to find the exact expression level of each of the four copies of

RcNUDX1-1a, because of the very high DNA sequence identities in the exons ([Align_OldBlush_DNAsequences.fasta](#) and [Clones_IntronExonStructure.fasta](#), [Supplementary Material](#) online), which would make almost impossible qRT-PCR experiment, even with a High Melting Resolution technique ([Roccia et al. 2019](#)). It was also because of the same length and structure of their promoters (see below and [supplementary fig. S6, Supplementary Material](#) online) which could indicate a similar expression.

Next, to investigate the contribution of promoters to different expression levels of the *RcNUDX1-1a* and *b* paralogs, we searched for the presence of specific sequences or

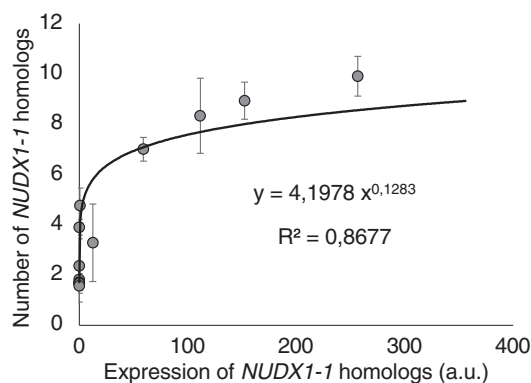


Fig. 7. Correlation between the expression of *NUDX1-1* homologs and the number of gene sequences in rose species. Expression of *NUDX1-1* was determined by qRT-PCR with FP8-RP8 primers, and FP5-RP5 and FP6-RP6 primers for reference genes (supplementary tables S7, and S10, Supplementary Material online). Number of gene sequences was estimated by qPCR with FP8-RP8 primers (supplementary fig. S5, Supplementary Material online). Error bars correspond to SD. a.u., arbitrary units.

structures upstream the coding sequences. In Old Blush, we manually identified four repeats of a conserved 38 bp sequence, designated as *box38* A to D. These repeats were identical in all five blocks of *RcNUDX1-1a*, #1 to #5, and always located 138 bp upstream the *RcNUDX1-1a* transcription starting site. Moreover, we found a 33-bp overlap between *box38* A and a fragment of the *Copia* R24588 localized at the 5'-end of each *NUDX1-1a* copy. In order to test the relationship between *Copia* R24588 and *box38*, the fragments of *Copia* R24588 and the *box38* repeats upstream of each copy of *RcNUDX1-1a* gene on the chromosome 2 were analyzed. The *Copia* R24588 fragments contained the consensus sequence published in the GDR and identified from the interspersed copies of *Copia* R24588 in the Old Blush genome. A search for short homologous sequences of *box38* in the Old Blush genome using BlastN and multiple sequence alignment (supplementary fig. S6, Supplementary Material online) confirmed that *box38* was the result of the 3'-end duplication of the *Copia* R24588 fragment (supplementary fig. S6a, Supplementary Material online). There were no other *box38* elements in the Old Blush genome, but only very short fragments were found in other TE, intron, and intergenic hits (supplementary fig. S6b, Supplementary Material online). The available online PlantCARE tool (Lescot et al. 2002), was unable to detect any known binding sites for transcription factors in the *box38* repeats, which does not exclude the existence of unknown ones. To go further, we performed another multiple sequence alignment using the *Copia* R24588 consensus sequence of the GDR. On this sequence, we aligned the following sequences: The *Copia* R24588 fragment upstream *RcNUDX1-1a* blocks on chromosome 2, and the *Copia* R24588 fragment upstream Ψ *RcNUDX1-2a* on chromosome 4 (fig. 8). The alignment clearly showed the origin of the promoter fragment (fig. 8a) in the complete consensus

map of *Copia* R24588, with *box38* A being the best aligned within the 3' long-terminal repeat (LTR) of *Copia* R24588 (fig. 8b). It also showed that *box38* B to D only exist upstream *RcNUDX1-1a* blocks (fig. 8c).

To find whether this pattern is conserved in botanical roses and important for the expression of *NUDX1-1a* in petals, we compared the upstream sequences of *NUDX1-1a* and *b* in a set of botanical roses producing and not producing geraniol (supplementary fig. S7, Supplementary Material online). Although the number of *box38* repeats varied in the wild roses, the 138 pb distance between the last *box38* sequence and the ATG codon of the *NUDX1-1a* was conserved (supplementary fig. S7a, Supplementary Material online). In contrast, none of the upstream region of *NUDX1-1b* contained any *Copia* R24588 sequence or *box38* repeats (supplementary fig. S7b, Supplementary Material online). One copy of the *box38* was also present in the *Copia* R24588 elements upstream Ψ *RcNUDX1-2a*, Ψ *RmNUDX1-2a*, and Ψ *RINUDX1-2a* pseudogenes on chromosome 4 suggesting that it could be more ancestral than those of chromosome 2.

All these results suggested a chronology of duplications: the *Copia* R24588 fragment of chromosome 4 was *trans*-duplicated on chromosome 2, the *box38* A was then *cis*-duplicated into four copies, and one of the putative blocks in figure 6a was *cis*-duplicated on chromosome 2. Furthermore, these results indicated that the promoter of *RcNUDX1-1a* seemed to be unique, and originated from a specialization of a fragment of the LTR of *Copia* R24588.

Finally, we analyzed the impact of the *box38* repeats and different TEs in the promoter region of *RcNUDX1-1a* on the specific expression of this paralog in rose petals (fig. 9). Reporter gene encoding the green fluorescence protein (GFP) was fused to the promoter region of *RcNUDX1-1a* of different lengths (fig. 9a). The longest *RcNUDX1-1a* promoter construct (*a1085:GFP*) included the entire 5'-region between MITEs and *RcNUDX1-1a* copy #4. The other constructs were made by removing the TEs one by one by PCR (supplementary table S7, Supplementary Material online). The *35S:GFP* used as a positive control displayed GFP fluorescence in parenchymous and epidermal cells (fig. 9b and c). No detectable GFP expression was found in rose petals transferred with the empty vector (fig. 9d and e) and the *RcNUDX1-1b* construct (1,529 pb upstream of the ATG codon, named *b1529:GFP* construct) used as a negative control (fig. 9f and g). GFP fluorescence was observed in rose petals expressing the three *RcNUDX1-1a* constructs, *a1085:GFP*, *a521:GFP*, and *a316:GFP* (fig. 9h–j). However, the removal of the *box38* repeats in the *a138:GFP* construct eliminated GFP expression (fig. 9k and l) suggesting that the *box38* repeats are essential for petal expression.

Overall, these data suggest that the appearance of the *NUDX1-1a* paralogs by the transposition of *NUDX1-1b* was accompanied by the evolution of its promoter, likely by duplication of sequence in the LTR region of *Copia* R24588, leading to the specific expression of this paralogs in petals. This could come from the promoter of an ancestral copy of *NUDX1-2* which already had the *box38* fragment.

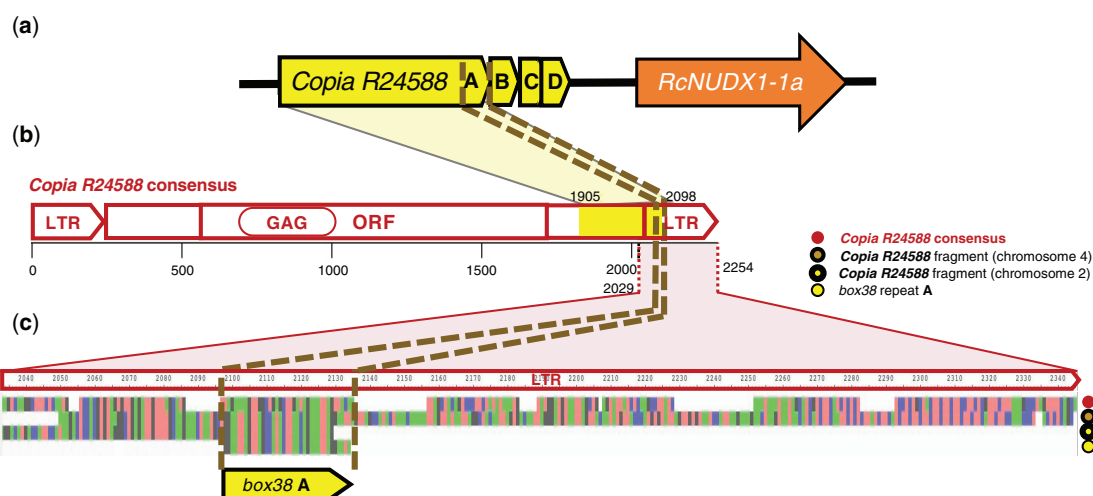


Fig. 8. Alignment interpretation of *box38* of chromosomes 2 and 4 of Old Blush genome. (a) An interpretative map of a block on chromosome 2 showing the localization of *Copia R24588* and *box38 A* fragment in the promoter of *RcNUDX1-1a*. (b) Manual annotation of *Copia R24588* consensus with the different regions of the retrotransposon. (c) Alignment (MAAFT) of *Copia R24588* consensus and upstream regions of *RcNUDX1-1a* on chromosome 2, and Ψ *RcNUDX1-2a* on chromosome 4 (alignment is given in [Align_CopiaLTR_Ch2and4.fasta](#), [Supplementary Material](#) online). This *Copia R24588* fragment aligns 4 bp further with the *box38* consensus (37/38 bp) than the fragments seen in the repeat blocks of chromosome 2, strengthening the LTR origin hypothesis for *box38*. Red circle, *Copia R24588* consensus of the GDR (Raymond et al. 2018; Jung et al. 2019); Brown circle, upstream region of Ψ *RcNUDX1-2a* on chromosome 4 (Jung et al. 2019; Hibrand Saint-Oyant et al. 2018); Yellow circle with thick black line, *Copia R24588* fragments (226 bp) located within *NUDX1-1a* block #1 on chromosome 2; yellow circle, corresponding *box38* repeat A; GAG, conserved capsid domain of the retrotransposon polyprotein; LTR, long-terminal repeat; ORF, open reading frame. Coordinates are in bp.

Discussion

Our analysis of the *NUDX1* genes in the *Rosaceae* family revealed that three clades (*Nudx1-1* to *Nudx1-3*) evolved in the *Rosoidae* subfamily (including *P. micrantha*, *F. vesca*, and *Rosa* species), and that two subclades (*Nudx1-1a* and *Nudx1-1b*) evolved in the *Rosa* genus (figs. 1, 2, and 5; [supplementary S2](#) and [table S2](#), [Supplementary Material](#) online). Considering *AtNUDX1* as an outgroup and *RhNUDX1-rs* from a modern garden rose, the *Nudx1-3* clade appeared to be more ancient than the others, and the *Nudx1-1a* subclade more recent. Comparative analysis of genetic maps of Old Blush, as a heritage rose producing geraniol, *Moschata*, as an accession of a wild rose producing geraniol, and *Laevigata*, as an accession of an unscented wild rose, allowed to access a global history of duplications in the *Rosoidae* subfamily (figs. 3, 4, and 6). The cluster *NUDX1-3/NUDX1-1b/NUDX1-2a* on chromosome 4 was found in *Rosoidae* accessions, suggesting a very old duplication of the putative ancestral *NUDX1-3* gene. In the *Amygdaloideae* subfamily (including *P. persica* and *M. x domestica*), their multiple copies in the same microsyntenic region (between marker genes F and Q in fig. 4) have significantly diverged, thus forming a different clade, *Nudx1-4* (fig. 2). In contrast, the cluster of *NUDX1-1a* copies on chromosome 2 is more recent, specific to some species of the *Rosa* genus and absent in ancestral species like *R. banksiae*, *R. roxburghii*, and *R. laevigata* (fig. 1c; [supplementary fig. S2](#), [Supplementary Material](#) online) (Fougère-Danezan et al. 2015; Debray et al. 2019). Moreover, the number of *NUDX1-1a* copies varies

depending on species, with two copies in the *Moschata* accession, and five copies in Old Blush (e.g., fig. 6; [supplementary fig. S5](#), [Supplementary Material](#) online). In Old Blush we identified two alleles on chromosome 2, one with five copies of *RcNUDX1-1a*, and the other with a null allele (fig. 3; [supplementary table S2](#), [Supplementary Material](#) online), which could confirm the previously predicted hybrid origin of this heritage rose (Raymond et al. 2018).

Our analysis of the TE landscape of *NUDX1-1* genes suggested a *trans*-duplication of a first paralog from chromosome 4 to 2, and then several *cis*-duplications of *NUDX1-1a* blocks including TEs in tandem (figs. 6 and 10; [supplementary fig. S4](#) and [table S12](#), [Supplementary Material](#) online). The presence of TEs in both the putative source of *NUDX1-1a* on chromosome 4 and duplication blocks on chromosome 2 raise the possibility of TE-mediated mechanisms. Indeed, sequence similarity between TE copies across the genome can be responsible for nonhomologous recombination and the relocation and rearrangement of genomic features between TE dense regions (Cerbin and Jiang 2018), as observed for other biosynthetic gene clusters in plants (Boutanaev and Osbourn 2018). Further extensive analysis of the repeat content in *Rosa* species and other *Rosaceae* will be required to test this hypothesis and other putative TE-derived mechanisms, such as Pack-MULE or retrotransposition (e.g., Jiang et al. 2004; Cerbin and Jiang 2018; Krasileva 2019).

RcNUDX1-1a copies 2, 3, and 4 were found on chromosome 2 as repeats of a sequence block *Copia R24588/RcNUDX1-1a* with MITE P580.2030 [MITE G13554] at the

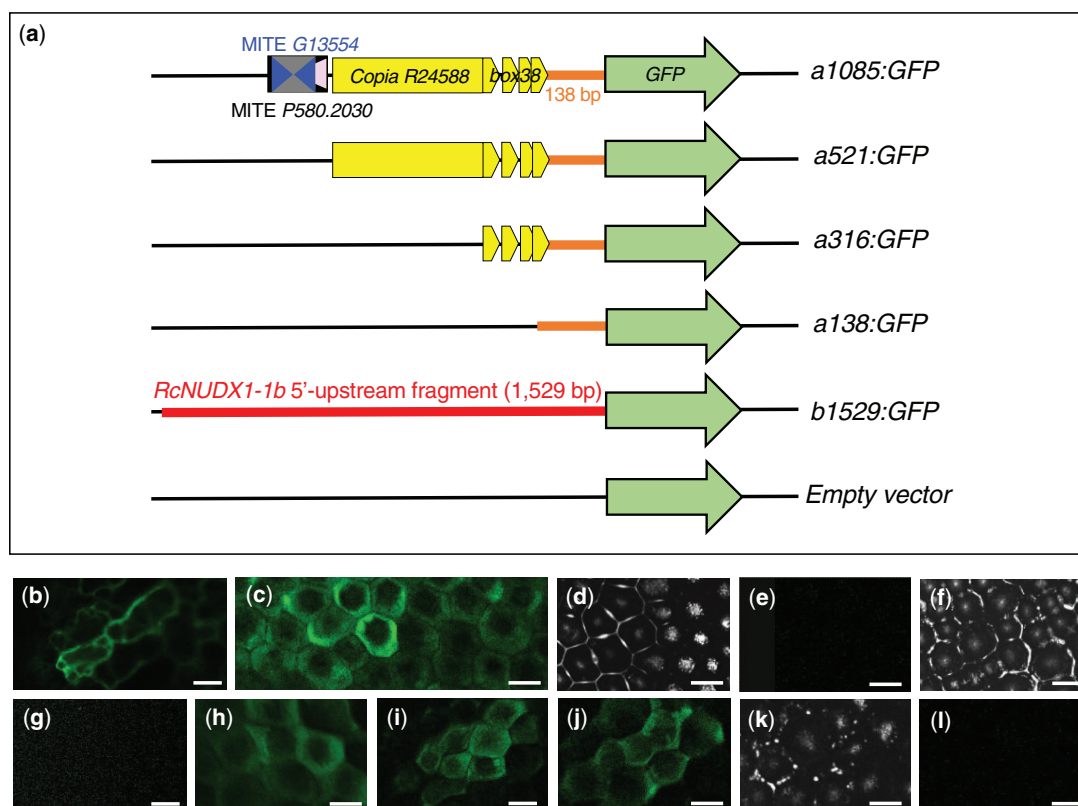


FIG. 9. Confocal laser scanning microscopy of transient expression of GFP constructs in agroinfiltrated petals of Old Blush. (a) Schematic maps of constructs including, respectively, 1085, 521, 316, and 138 bp upstream *RcNUDX1-1a*, 1529 bp upstream *RcNUDX1-1b*, and GFP alone (empty vector). (b)–(i) Confocal images except for (d), (f), and (k) taken by reflection of light on the preparation. Petals were infiltrated with the following constructs: 35S:GFP (b, c), empty vector (d, e), b1529:GFP (f, g), a1085:GFP (h), a521:GFP (i), a316:GFP (j), and a138:GFP (k, l). Cloning was made with FP11-RP11 to FP15-RP11 primers (supplementary table S7, Supplementary Material online). Scale bars, 20 μ m.

beginning or at the end (fig. 6; supplementary fig. S4, Supplementary Material online). In addition, MITE P580.2030, *Copia* R24588, and *NUDX1* homologs were found on one homologous chromosome 4 in a different configuration (*RcNUDX1-1b*/MITE P580.2030/MITE P580.2030/.../*Copia* R24588/^ψ*RcNUDX1-2a*) where MITE P580.2030 does not include MITE G13554, but is *cis*-duplicated in tandem. This suggests that the copies of *NUDX1* on chromosome 4, including uninterrupted MITE P580.2030, are ancestral to those on the chromosome 2 and have been rearranged upon duplication (fig. 6). The parental status of the sequences on chromosome 4 is also supported by the fact that the microsynteny was not shared between *Rosa* species on chromosome 2 (five interspersed copies of *RcNUDX1-1a* in Old Blush, two in *Moschata*, and none in *Laevigata*), but was conserved on chromosome 4. Finally, high expression of *NUDX1-1a*, but not *NUDX1-1b*, in petals of fully opened flowers (table 1; supplementary fig. S1 and table S10, Supplementary Material online), further indicates that the cluster on chromosome 2 acquired petal-specific expression following duplication from chromosome 4 and subsequent

duplication in tandem of the rearranged block. Such *cis*-duplications can occur by nonallelic homologous recombination between two identical sequences that may create an unequal crossing-over, or by microhomology-mediated break-induced replication mechanisms (Żmieńko et al. 2014; Lye and Purugganan 2019), even in synergy with TE mechanisms of translocation (Krasileva 2019). In *M. x domestica*, clusters of *O-METHYLTRANSFERASE* genes are associated with hairpins structures from palindromic TEs provoked by DNA slippage during replication (Han et al. 2007). In our work, MITES P580.2030 and G13554 are also forming ~300–400 bp palindromes associated with each replicated *RcNUDX1-1a* block on chromosome 2.

We also discovered that repeats of a 38-bp fragment derived from the LTR region of *Copia* R24588, and named *box38*, was necessary and sufficient to drive previously discovered petal-specific *NUDX1-1a* expression in petals of fully opened flowers (Magnard et al. 2015) (figs. 7 and 9; supplementary fig. S1, Supplementary Material online). The *Copia* R24588/*box38* location in the 5'-upstream regions of the pseudogenes ^ψ*NUDX1-2a* suggests that this gene may have been expressed

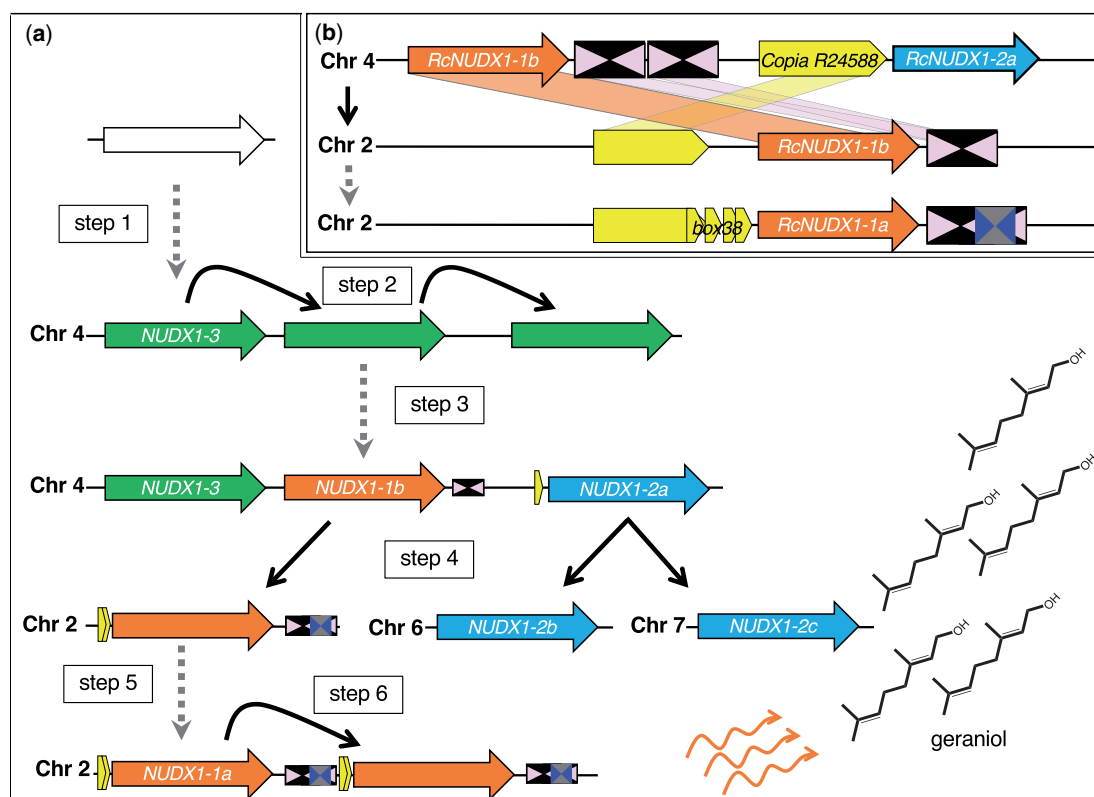


FIG. 10. Scenario of evolution of *NUDX1* in botanical roses. (a) Global scenario of duplications and specializations. Step 1, specialization of an unknown ancestral *NUDX1* into *NUDX1-3*; Step 2, *cis*-duplication of *NUDX1-3*; Step 3, specialization of *NUDX1-3* into *NUDX1-1b* and *NUDX1-2a* (during this step some TEs were probably inserted near *NUDX1-2a*); Step 4, *trans*-duplications of *NUDX1-1b* and *NUDX1-2a* (after this step, *NUDX1-2a* could have pseudogenized); Step 5, functionalization of expression in petals (during this step *box28* could have duplicate); Step 6, *cis*-duplications of *NUDX1-1a* and increase of the level of geraniol emission. (b) Example of possible *RcNUDX1-1b* to *RcNUDX1-1a* transposition. Large white arrow, putative ancestral *NUDX1* gene; large orange arrows, genes from *Nudx1-1* clade; large blue arrows, genes from *Nudx1-2* clade; large green arrows, genes from *Nudx1-3* clade; pink drawings, MITE P580.2030; dark blue drawings, MITE G13554; yellow arrow, *Copia* R24588; dashed gray arrows, specialization steps; black arrows, duplication steps; orange curly arrows, volatile emission; Chr, chromosome.

originally. Thus, even if it really looks like a neofunctionalization process, one cannot exclude subneofunctionalization as well (see review in Baudino et al. 2020). However, during *trans*-duplication from chromosome 4 to chromosome 2, the *box38* repeats were shuffled and ended up in front of *NUDX1-1a* making its expression petal-specific (fig. 10). To date, there is increasing evidence that TEs are a source of diversification of species and can modify gene expression, particularly in the *Rosaceae* (Gu et al. 2016; Wang et al. 2016; Zhao et al. 2016; Daccord et al. 2017; Jiang et al. 2019). Examples include recurrent blooming of roses and strawberries due to an insertion of another *Copia* element in the intron 2 of the antiflorigen homolog *KSN* (Iwata et al. 2012), and formation of more than five petals in roses due to insertion of an uncharacterized TE in the intron 8 of *APETALA2/TOE*, which deregulated its expression (Hibrand Saint-Oyant et al. 2018). Several TE insertions in promoters have also been described in *Rosaceae*, which modified transcription levels as a result of new binding sites for

transcription factors or disruption of existing ones, new methylation/acetylation patterns, or hairpin structure formation (Han and Korban 2007; Wang et al. 2009; Gu et al. 2016; Morata et al. 2018; Ono et al. 2018; Zhang et al. 2019).

Our results show that *box38* is part of the LTR region of *Copia* R24588. LTRs flank the internal coding region of LTR retrotransposons and act as promoter for the selfish transcription of the canonical elements of the retrotransposon. LTR regions contain regulatory sequences that can modify gene expression occurring in *cis* and can contribute to neofunctionalization in plants and eukaryotes (Kobayashi et al. 2004; Grandbastien 2015; Galindo-González et al. 2017). As Old Blush is rich in TEs, which constitute 63.2% of the genome including 35.2% of class I LTR retrotransposons (Hibrand Saint-Oyant et al. 2018), further investigations are necessary to understand the underlying mechanisms of petal-specific expression.

We also found that the number of *NUDX1-1a* copies impacts the level of geraniol emission in wild roses, in a

nonlinear gene dosage effect (fig. 7; supplementary fig. S3, Supplementary Material online). A similar situation was described in mammals, where the copy number of genes encoding amylase was higher in populations with high-starch diets, but not strictly linearly correlated to the amylase concentration in saliva (Perry et al. 2007; Axelsson et al. 2013). In an evolution perspective, if the number of copies increases fitness, these copies can be fixed by adaptive natural selection rather than diverge by genetic drift (Hahn 2009). As *RcNUDX1-1a* copies are very similar to each other (96.8–99.0% of DNA identity resulting in 98.7% of protein identity), it is possible that this gene, and thus geraniol concentration, were important in the adaptation and evolution of *Rosa* species. Interestingly, the blocks on chromosome 2 in *Rosa* look similar to the repetitions of *MATE1* in *Zea mays*, which include copies of *Copia*, *Gypsy*, and *Mutator* in their intergenic regions and for which the total number of gene copies is associated with aluminum tolerance (Maron et al. 2013). This polymorphism is referred as copy number variations (CNVs), that is, variation of number of gene copies between individuals (Lye and Purugganan 2019), or between inbred lines (Maron et al. 2013). It has been demonstrated that such CNVs could be a very strong driving force leading to adaptations (DeBolt 2010) even via secondary metabolism (Prunier et al. 2017; Shirai and Hanada 2019). The differences of copy number between Old Blush, Moschata, and Laevigata (figs. 6 and 7; supplementary fig. S5, Supplementary Material online) could well correspond to ancestral CNVs, because of adaptations of different populations in an ancestral species. It could even have participated in the speciation of these species similar to the situation in *Picea* spp. (Prunier et al. 2017).

Our results also showed the existence of correlation of *NUDX1-1a* activity not only with geraniol levels but also with some other volatiles (supplementary fig. S3, Supplementary Material online). This could be due 1) to an indirect effect (selection pressure on a transcription factor that regulates several biosynthesis genes, or pleiotropic effects), for example, as it was observed for terpenes and phenylpropanoids in an overexpression experiment of *PAP1* in *R. x hybrida* “Pariser Charme” (Ben Zvi et al. 2012); 2) to diffuse selection pressure of pollinators, florivores, or parasites on several volatile compounds (e.g., acyclic terpenoids and 2-phenylethanol are known to be very attractive for insects; Raguso 2004; Trhlin and Rajchard 2011); 3) to common biosynthetic pathway for acyclic terpenoids, as it is the case in other species for geraniol, nerol, β -citronellol and their aldehydes and acetates (e.g., see review in Sun et al. 2016), or 4) other unknown effects, like, for example, modifications or redirections of different fluxes through pathways of precursors or related to precursors.

In conclusion, *NUDX1* genes duplicated several times in *Rosaceae* species and probably acquired different functions. In the *Rosoidae* subfamily, three distinct clades were formed (fig. 10). The *Nudx1-1* clade has evolved forming two subclades by duplication. In the genus *Rosa*, the more ancient *NUDX1-1b* gene was transposed from chromosome 4 and the surrounding TEs rearranged, such as the *Copia R24588* element, providing the building blocks for *box38*. This raises the question of

how its promoter is specifically activated in the petals and by which transcription factors. The resulting *NUDX1-1a* on chromosome 2 was then able to produce geraniol in rose petals, which could be a high driving force of selection. This driving force was amplified in some rose species by several *cis*-duplications of *NUDX1-1a*. It is thus relevant to ask how the nonlinear effect of the gene copy number works in detail. Finally, use of the *box38* sequences for marker-assisted selection of scented roses could be a relevant application.

Materials and Methods

Plant Materials and Sampling

Samples (fig. 1c; supplementary table S1, Supplementary Material online) were collected in France in several botanical gardens (Roseraie de Saint-Clair, Caluire, France; Roseraie de Loubert, Les Brettes, France; Parc de la Tête d’Or, Lyon, France), in the wild (Mornant, France), or in the BVpam laboratory garden (Saint-Etienne, France). The same species or variety in two different collections or different geographic area received two different names of accession. Descriptive data (ploidy, geography, phylogeny, and families) were reported according to the literature (Cairns 2003; Wissemann 2003; Schorr and Young 2007; Masure 2013; Fougère-Danezan et al. 2015; Zhu et al. 2015; Zhang et al. 2017; Debray et al. 2019). Each sampling was repeated at least three times between 2014 and 2019, depending on the location, the flowering period, and the weather forecast. This last point was important because wild roses often bloom during a fortnight. Buds for DNA extraction, and petals for mRNA extraction, were frozen in liquid nitrogen for transport and conserved at -80°C before further experiments. Petals for volatile analysis were directly immersed in hexane containing (+/–)-camphor (#148075, Merck) at 5, 10, or 20 mg/l as an internal standard. Each vial contained 1 g of petals of individual flowers and 2 ml of hexane and (+/–)-camphor mix. Vials were transported to the laboratory in ice.

GC–MS Analyses

The hexane extracts were recovered from the vial after 24 h at $+4^{\circ}\text{C}$ and processed according to Sun et al. (2020): Agilent 6850 gas chromatograph, DB5 apolar capillary column (30 m x 0.25 mm), 7683B series injector, and 5973 Network mass selective detector (Agilent Technologies). Helium at a flow rate of 1.0 ml/min was used as a carrier gas with the following program: 40°C for 3 min, gradient of 3°C min from 40°C to 245°C , and 10 min at 245°C . Injection volume was 2 μl with a split mode (split ratio 1:2) and the injector and detector temperatures were 250°C . The parameters for mass spectrometer detector were set as follows: mass scan range 35–450 m/z , and ionization voltage 70 eV. Kovatz indexes (AI) were calculated according to Adams (2007) and to the Nist Web Book. Names and families of compounds (supplementary table S8, Supplementary Material online) were given by screening Wiley 275 and Nist 08 databases, and by names given by (Knudsen et al. 2006). Spearman’s correlation coefficients (supplementary table S11, Supplementary Material online) and heatmap (supplementary fig. S3, Supplementary

Material online) were calculated with the R language and environment (R Core Team 2015) using Hmisc (Harrell and Dupont 2020) and corr (Kuhn et al. 2020) packages.

DNA and RNA Extractions

For HMW-gDNA extraction, 100 mg of fresh buds were grinded with pestle and mortar in 2 ml of CTAB buffer (100 mM Tris-HCl pH 8.0, 3 M NaCl, 3% CTAB, 20 mM EDTA, and 2% w/v PVP-40). 90 µg of Ribonuclease A (Sigma-Aldrich) was added before heating for 45 min in water bath at 65 °C. Cellular debris were pelleted (13,000 × g, 5 min, 4 °C) and the supernatant was mixed with equal volume of chloroform:isoamyl alcohol (24:1, v/v) and shaken slowly for 1 min. Aqueous phase was separated by centrifugation (12,000 × g, 5 min, 4 °C). The upper phase was carefully recovered and washed three times more. Nucleic acids were precipitated by addition of 0.1 vol of 3 M sodium acetate pH 5.2 and 0.66 volume of cold ethanol 100% (−20 °C). Tubes were mixed by inversion and kept at −20 °C for 1 h. DNA was pelleted by centrifugation at 5,000 × g for 10 min at 4 °C. DNA was washed three times with ethanol 70% and the pellet was dried for 10 min at room temperature and resuspended in 40 µl of TE (10 mM Tris-HCl pH 8, 1 mM EDTA). All centrifugations were performed with slow acceleration and deceleration. Alternatively, the NucleoSpin Plant II Kit was used (Macherey Nagel) for other experiment needing gDNA (cloning and qPCR).

For RNA extraction, petals of opened flowers (anthesis stage) were crushed in liquid nitrogen and extracted with the NucleoSpin RNA Plant kit (Macherey-Nagel) with on-column DNase for gDNA removal with the NucleoSpin rDNase Set (Macherey-Nagel). Absence of gDNA was checked by PCR. cDNA was obtained with the iScript Ready-to-use cDNA Supermix kit (Biorad) at 42 °C for 1 h with 1 µg of RNA. All kits were used according to the manufacturer's instructions.

qPCR, qRT-PCR, and DNA Cloning for Sequencing

Primers used for cloning are given in [supplementary table S7, Supplementary Material](#) online. Cloning of gDNAs and cDNAs ([Clones_gDNAs_cDNAs.fasta](#), [Supplementary Material](#) online) were done after PCR amplification with Phusion High Fidelity polymerase (Thermo Fisher Scientific). The PCR parameters with RP7-FP7 primers were as followed: 98 °C for 1 min, 28 cycles of (98 °C for 10 s, 58 °C for 30 s, and 72 °C for 20 s), and 72 °C for 5 min. After purification of PCR product with the NucoSpin Gel and PCR clean up kit (Macherey-Nagel), ligation was done into pCRBlunt (Invitrogen), and transformed into *Escherichia coli* TOP10 (Invitrogen). Plasmid were purified with the NucleoSpin Plasmid Kit (Macherey-Nagel). *NUDX1-1* gDNA and cDNA inserted into plasmids were sent to MWG Eurofins for sequencing using universal M13uni-21 primer.

Copy number determination of *NUDX1-1* genes by qPCR were performed with FP8-RP8 primers. The qPCR reaction consisted of 10 µl of SsoAdvancedTM SYBR Green Supermix (Bio-Rad), 500 nM R and F primers, 20 ng of diluted gDNA in 20 µl volume reaction. The parameters were as followed:

98 °C for 5 min, 40 cycles of 98 °C for 10 s and 58 °C for 30 s. At the end of each run, the melting curve was set to 0.5 °C every 2 s from 65 °C to 95 °C. The number of copies was calculated by comparison with copies of *RcNUDX1-1* assuming that there were seven copies in Old Blush (five *RcNUDX1-1a* copies and two *RcNUDX1-1b* alleles; [fig. 3; supplementary fig. S5, Supplementary Material](#) online). Three biological replicates were performed with gDNA from three different plants.

Amplifications for qRT-PCR were done according to [Sun et al. \(2020\)](#) with housekeeping gene primers FP5-RP5 and FP6-RP6 designed on *RcEF1* and *RcTUB* sequences, respectively (GenBank accession numbers BI978089 and AF394915) ([Dubois et al. 2012](#)). To determine the expression of the different *RcNUDX1-1* homologs of Old Blush, FP1-RP1 to FP4-RP4 primers were used. For *NUDX1-1* expression measurement in the different *Rosa* species, FP8-RP8 primers were used ([fig. 7; supplementary fig. S5 and table S10, Supplementary Material](#) online). Diluted (1/25) cDNAs were used in 20 µl reaction with SsoAdvancedTM SYBR Green Supermix (Bio-Rad). The PCR parameters were as followed: 95 °C for 30 s, and 30 cycles of (95 °C for 5 s, and 64 °C for *RcEF1* amplification [GenBank accession number BI978089], or 58 °C for *RcTUB* [GenBank accession number AF394915] and *NUDX1-1* amplification for 30 s). At the end of each run, the melting curve was set to 0.5 °C every 2 s from 65 °C to 95 °C. Cq values were automatically determined by the CFX96 Real-Time system with default settings. ΔC_t method ([Pfaffl 2001](#)) was used for quantification by comparison with reference genes. For each species, several independent qRT-PCR on different biological samples were performed.

Long-Read Sequencing

Sequencing library was prepared from 1 µg fresh HMW-gDNA for each species using the genomic DNA ligation sequencing kit (SQK-LSK109, version 14aug2019, Oxford Nanopore Technologies) following the manufacturer's recommendations. Library was then sequenced on a FLO-MIN106 flow cell using a MinION device (Oxford Nanopore Technologies). Obtained reads were subsequently basecalled using guppy software in high accuracy mode with parameters adapted to the sequencing kit and the flowcell (dna_r9.4.1_450bps_hac.cfg) using guppy in GPU mode. Basecalled fastq files were converted in fasta using the fastq_to_fasta program from the FASTX Toolkit v0.0.14. Blast databases were obtained for each species from the fasta files then the BlastN program ([Camacho et al. 2009](#)) was used to search for reads containing *NUDX* genes using either *RcNUDX1-1a*, *1-1b*, *1-2a*, *1-2b*, *1-2c*, and *1-3* sequences as query ([supplementary tables S5 and S13, Supplementary Material](#) online). Hits on identified reads were then manually analyzed to determine the organization of *NUDX* clusters.

Sequence Annotations, Phylogenies, and Synteny Maps

Genes and transposons were named according to the GDR ([Jung et al. 2019](#)). The sequence of *R. x hybrida* cv. 'Papa Meilland' (*RhNUDX1*, GenBank accession number

JQ820249) was used to clone the corresponding gene including the intron. It was named *RhNUDX1-rs* for reference sequence and was used to search sequences in “*Rosa chinensis* Genome v1.0 chromosomes” (Hibrand Saint-Oyant et al. 2018), “*Rosa chinensis* Old Blush Illumina Genome v1.0 chromosomes,” “*Rosa chinensis* Old Blush homozygous Genome v2.0 chromosomes” (Raymond et al. 2018), “*Rosa multiflora* draft Genome v1.0” (Nakamura et al. 2018), “*Fragaria vesca* Genome v4.0” (Edger et al. 2018), “*Malus x domestica* Genome (GDDH13 v1-1)” (Daccord et al. 2017), and “*Prunus persica* Genome v2.0.a1” (Verde et al. 2013, 2017), all published in the GDR. They were searched directly using the blast tool online in the GDR, and/or by downloading the fasta files in Geneious Prime software (Biomatters Limited) for alignments, BlastN, and calculation of identity. The nonassembled genome of *P. micrantha* “*Potentilla micrantha* v1.0” (Buti et al. 2018) of the GDR was also used because of the phylogeny proximity with the genus *Rosa*. Sequences were directly searched in its scaffolds by BlastN in the Geneious Prime software. The ML tree in figure 2 was calculated and drawn in the Geneious Prime software with the plugin PhyML (Guindon et al. 2010) using complete DNA sequences, and non full-identical sequences. The following sequences published in Sun et al. (2020) were used as references to name clades: *RcNUDX1-1a* (*RcHm_v2.0_Chr2g0142071*, *0142081*, *0142111*, and *0142121*), *RcNUDX1-1b* (*RcHm_v2.0_Chr4g0436181*), *RcNUDX1-2a* (*RcHm_v2.0_Chr4g0436151*), *RcNUDX1-2b* (*RcHm_v2.0_Chr6g0244161*), *RcNUDX1-2c* (*RcHt_S2031.3*), *RcNUDX1-3* (*RcHm_v2.0_Chr4g0436191*), and *RwNUDX1-1*, *RwNUDX1-2a*, *RwNUDX1-2b*, *RwNUDX1-2c*, *RwNUDX1-2c'*, *RwNUDX1-3* (Genbank accession numbers, respectively, MT362556 to MT362561). The gene sequences included the intron for increasing bootstraps (*Align_Rosaceae_MLtree.fasta* and supplementary table S2, Supplementary Material online). *AtNUDX1* gene of *A. thaliana* was used as an outgroup (GenBank accession number AT1G68760). The dot-plot of similarity (supplementary fig. S4a, Supplementary Material online) was made with the plugin LASTZ (Harris 2007). For microsynteny (figs. 3, 4, and 6; supplementary table S2, Supplementary Material online), marker genes around the *NUDX1* genes were used to verify correspondences between homologous regions in the GDR and in MinION reads. They were arbitrarily named A to S (full list in supplementary table S14, Supplementary Material online).

The *NUDX1* gene phylogeny (fig. 5; supplementary fig. S2 and *Clones_IntronExonStructure.fasta*, Supplementary Material online) was reconstructed using the entire 660 bp, thus including the intron, with *F. vesca* *NUDX1* gene as outgroup (GenBank accession number XM_004297107.2). Indeed, as the coding parts of the *NUDX1* gene are strongly conserved between species, too little phylogenetic information is contained in the exonic sequences, whereas the intronic sequence is more variable and makes the phylogenetic reconstruction possible. *NUDX1* genes were aligned using Clustalw (Thompson et al. 2003), and sites ambiguously aligned were removed with Gblocks (Castresana 2000), resulting in a 608 bp alignment. ML phylogenetic reconstruction was conducted

using PhyML (Guindon et al. 2010) under a GTR + G + I model (*Align_OldBlush_MLtree.fasta*, Supplementary Material online). Tree was rooted with the *FvNUDX1-1* gene (GenBank accession number XM_004297107.2). In order to understand the history of duplication, we need to know which sequences belongs to the chromosome 2 (*NUDX1-1a* paralog) and which ones belong to the chromosome 4 (*NUDX1-1b* paralog). To achieve that, all sequences were aligned by BlastN against Old Blush *RcNUDX1-1a* (GenBank accession number, CM009583.1, from position 59,567,055 to 59,567,676 bp) and *RcNUDX1-1b* (GenBank accession number CM009585.1, from position 59,520,245 to 59,520,862 bp). Identities of the DNA sequences and the putative proteins were also calculated (supplementary tables S3 and S4, *Align_OldBlush_DNAsequences.fasta*, and *Align_OldBlush_Proteins.fasta*, Supplementary Material online) to draw the comprehensive map (fig. 3), gDNAs displaying identity more than 1% higher with *RcNUDX1-1a* than with *RcNUDX1-1b* were assigned to the *Nudx1-1a* subclade and vice versa (supplementary fig. S2 and table S9, Supplementary Material online). As these two paralogs are very similar, some sequences aligned similarly with BlastN (<1% with both references), and thus were not assigned to one of the subclades.

Promoter Analysis, Cloning, and Transient Expression

For promoter analyses of *Copia* R24588 and *box38* hits and homology, we used BlastN (Camacho et al. 2009) with the minimum seed size (word_size = 7) allowing to recover hits from short query sequences. Multiple alignments were performed with MAFFT (Katoh et al. 2019) using the following parameters (parameters -thread 2 -reorder -adjustdirection -accurately -anysymbol -maxiterate 2 -retree 1 -genafpair). Alignments are given in *Align_CopiaBox38_Chr2.fasta* and *Align_CopiaLTR_Chr2and4.fasta* (Supplementary Materials online). Quality control of the alignment and minor extensions of the BlastN hits (up to 2 bp) within the *box38* consensus were performed manually. A consensus sequence logo for *box38* was created using WebLogo v2.8.2 (Crooks et al. 2004). We also mapped the consensus sequence of *Copia* R24588 of the GDR by using RepeatClassifier, a tool included with RepeatModeler2 (Flynn et al. 2020) and TE-Aid (https://github.com/clemgoub/TE-Aid).

Primers used for cloning are given in supplementary table S7, Supplementary Material online. For promoter cloning, FP9-RP9 (upstream region of *NUDX1-1b*) and FP10-RP10 (upstream region of *NUDX1-1a*) were used and cloned into pCRBlunt (Invitrogen) as mentioned above and sequenced with the same procedure using the M13uni-21 primer for sequencing. Amplification of the different promoter regions was done with Phusion U Hot Start DNA Polymerase (ThermoFisher Scientific) with combinations of USER extended primer FP11 to FP15 and RP11 with RcoB gDNA as template (fig. 9). The PCR parameters' primers were as followed: 98 °C for 1 min, 25 cycles of (98 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s), and 72 °C for 5 min. PCR products were cloned into a pCambia2300 binary base vector with linearized *PacI*-USER cassette upstream the GFP and NOS-terminator using USER enzyme (New England Biolabs). The

control construct based on double CaMV 35S promoter was cloned into the same vector with the same method using the binary vector pMDC32 containing this promoter as matrix with FP16-RP16. All USER reaction was transformed into *E. coli* TOP10 (Invitrogen). Plasmids were purified with the NucleoSpin plasmid kit (Macherey Nagel). Sequence of constructs was verified before use.

These constructs were transformed into the *Agrobacterium* strain LBA4404. *Agrobacteria* were grown on LB agar with rifampicin (50 µg/ml), gentamicin (20 µg/ml), and kanamycin (50 µg/ml), and then screened by PCR for the presence of the construct. *Agrobacteria* were grown in 25 ml of liquid LB with antibiotics and collected by centrifugation at room temperature for 8 min at $4,000 \times g$ and washed in 10 mM MgCl₂ and 10 mM MES pH 5.7 buffer three times. They were diluted to OD_{600nm} = 1.0 with wash buffer and infiltrated on the abaxial side of Old Blush petals with a syringe. After 3 days, infiltrated petals were observed with a TCS-SP2 inverted confocal scanning laser microscope (Leica) with a $\times 40/0.80W$ lens. The argon laser was set at 488 nm for GFP excitation and the fluorescent signal was captured at 500–550 nm.

Enzyme Assay

RcNUDX1-1a and *RcNUDX1-1b* cDNA sequences corresponding to Old Blush gDNA1 and 2, respectively ([Clones_gDNAS_cDNAs.fasta](#), [Supplementary Material](#) online), were amplified by PCR (primers FP17-RP17; [supplementary table S7](#), [Supplementary Material](#) online) and cloned in pET-30a(+) between the *KpnI* and *Sall* restriction sites. *RmNUDX1-1a* and *RmNUDX1-1b* cDNAs corresponding to Moschata gDNA10 and gDNA2, respectively ([Clones_gDNAS_cDNAs.fasta](#), [Supplementary Material](#) online), were synthesized (GenScript) and cloned in pET-30a(+) between the *KpnI* and *Sall* restriction sites. Sequences and vectors were verified by sequencing and transformed into *E. coli* BL21(DE3)pLysS.

Transformants were grown at 37 °C in LB medium until OD_{600nm} = 0.4. Proteins were produced by overnight induction at 16 °C with 1 mM IPTG. After centrifugation, bacteria pellet was resuspended in buffer (50 mM Tris-HCl pH 8.5, 500 mM NaCl, 2 mM DTT, 8% glycerol v/v, 10 mM imidazole, 0.25 mg/ml lysozyme) and lysed by sonication. Supernatant was mixed with Ni-NTA agarose resin (Qiagen) for 1 h. Resin was rinsed five times with 50 mM Tris-HCl pH 8.5, 500 mM NaCl, 2 mM DTT, 8% v/v glycerol, and 50 mM imidazole, and finally eluted in the same buffer but containing 250 mM imidazole. Proteins were desalted by passing through a PD10 desalting column (GE Healthcare) equilibrated with the assay buffer (50 mM HEPES pH 8, 5 mM MgCl₂, 5% v/v glycerol) and quantified with the Bradford method. All steps of purification were conducted on ice.

Enzymatic reactions were performed in assay buffer containing different concentrations of GPP (0.5, 1, 2, 5, 10, 30, or 50 µM) in 100 µl reaction volume at 30 °C for 4 min, and using 20 ng of proteins. Reactions were stopped by adding 100 µl MeOH:H₂O (10 mM NH₄OH) 7:3 and mixed for 30 s.

Product analyses were performed on an Agilent 1260 infinity II LC system coupled to an Agilent Ultivo triple

quadrupole mass spectrometer (Agilent Technologies, Santa Clara) using a Poroshell 120 HPH-C18 column (50 mm \times 2.1 mm, particle size 1.9 µm, Agilent) heated at 35 °C. The mobile phases consisted of 10 mM ammonium bicarbonate pH 10.2 with 0.15% v/v ammonia, as solvent A, and acetonitrile with 0.15% v/v ammonia, as solvent B, with a 0.6 ml min flow rate. Two microliters of reaction mixture were injected for each sample. Separation was achieved with a gradient starting with 2% B reaching 98% B in 2, 1 min isocratic at 98% B and return at 2% B at 3.10 min with equilibration until 6.5 min. Mass spectrometer tunings were as follow: capillary voltage 5000 V, gas temperature 350 °C, gas flow 12 l/min, and nebulizer 55 psi. Products detection was achieved in negative and MRM modes with the following MS/MS transitions and tunings: 312.2–78.9 *m/z* for GPP with Fragmentor at 70 V and Collision Energy at 92 V and 233.1–78.9 *m/z* for GP with Fragmentor at 75 V, and collision energy at 60 V. Data analysis was performed with MassHunter quantitative software (Agilent Technologies). Enzyme Kinetic parameters were determined using the Lineweaver-Burk plot model.

Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by Agence Nationale de la Recherche (grant number ANR-16-CE20-0024-01 to S.B.), by Centre National de la Recherche Scientifique (grant reference MITI-ExoMod to J.C.C.), by Fondation de l'Université Jean Monnet (grant to J.C.C.), by National Science Foundation IOS (grant number 1655438 to N.D.), and by USDA National Institute of Food and Agriculture Hatch Project (grant number 177845 to N.D.). The authors wish to thank Thérèse Loubert and the town halls of Caluire and of Lyon, for sampling authorization in "Roseraie de Loubert," "Roseraie de Saint-Clair," and "Parc de la Tête d'Or." They also thank "Groupe de Recherche MédiatEC" and "Réseau MétaSP" for the discussions on specialized metabolism, and Laurent Duret and Tristan Lefébure (laboratoire de Biométrie et Biologie Evolutive, Lyon, France) for the discussions on gene evolution and functionalization.

Author Contributions

J.C.C. and S.B. conceived and managed the project. A.B., B.N., J.C.C., L.H.S.O., S.B., S.M., and T.T. sampled the botanical roses in the different collections. A.B., C.C., and P.S. cloned the cDNAs and gDNAs. P.S. and R.S. defined the Nudix clades in *Rosa*, and cloned the different homologs. C.D. and N.S. performed the Bayesian phylogeny. J.C.C., S.B., S.M., and S.N.P. performed the GC-MS analyses. S.M. and S.N.P. performed the statistical analyses. C.C. and C.G. performed the bioinformatic work on TEs. C.C., J.C.C., and F.F. studied the synteny. C.C., D.S.M., J.J. and M.B. set up, used the MinION technology, and performed the bioinformatic work with the reads. C.C., C.D., J.L.M., and N.S. cloned the promoter and made the constructs. C.C. and J.C.C. analyzed transient

expression. J.C.C., B.B., C.C., S.B., and N.D. analyzed data, designed figures, and wrote the manuscript, with input from all authors.

Data Availability

Raw data are given in [Supplementary Material](#) online, including fasta sequences of cDNAs and gDNAs cloned in this paper. *NUDX1*-rs sequence is deposited in the GenBank with the accession number MW762674. Reads from MinION sequencing are available in the SRA database in FASTQ format under the bioproject accession number PRJNA706580.

References

- Adams RP. 2007. Identification of essential oil components by gas chromatography/mass spectrometry. 4th ed. Carol Stream (IL): Allure.
- Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
- Baudino S, Huguency P, Caissard JC. 2020. Evolution of scent genes. In: Pichersky E, Dudareva N, editors. *Biology of plant volatiles*. Boca Raton (FL): CRC Press. p. 217–234.
- Ben Zvi MM, Shklarman E, Masci T, Kalev H, Debener T, Shafir S, Ovadis M, Vainstein A. 2012. PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. *New Phytol.* 195(2):335–345.
- Boutanaev AM, Osbourn AE. 2018. Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc Natl Acad Sci U S A.* 115(28):E6650–E6658.
- Buti M, Moretto M, Barghini E, Mascagni F, Natali L, Brilli M, Lomsadze A, Sonego P, Giongo L, Alonge M, et al. 2018. The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry). *Gigascience* 7(4):1–14.
- Cairns T. 2003. Classification. In: Roberts A, Debener T, Gudin S, editors. *Encyclopedia of Rose Science*. Amsterdam (The Netherlands): Elsevier. p. 117–123.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17(4):540–552.
- Cerbin S, Jiang N. 2018. Duplication of host genes by transposable elements. *Curr Opin Genet Dev.* 49:63–69.
- Chen F, Tholl D, D'Auria JC, Farooq A, Pichersky E, Gershenzon J. 2003. Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell* 15(2):481–494.
- Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14(6):1188–1190.
- Daccord N, Celton J-M, Linsmith G, Becker C, Choisine N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, et al. 2017. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet.* 49(7):1099–1106.
- DeBolt S. 2010. Copy Number Variation shapes genome diversity in *Arabidopsis* over immediate family genomic scales. *Genome Biol Evol.* 2:441–453.
- Debray K, Marie-Magdelaine J, Ruttink T, Clotault J, Foucher F, Malécot V. 2019. Identification and assessment of variable single-copy orthologous (SCO) nuclear loci for low-level phylogenomics: a case study in the genus *Rosa* (Rosaceae). *BMC Evol Biol.* 19(1):152.
- Dubois A, Carrere S, Raymond O, Pouvreau B, Cottret L, Rocca A, Onesto JP, Sakr S, Atanassova R, Baudino S, et al. 2012. Transcriptome database resource and gene expression atlas for the rose. *BMC Genomics* 13:638–648.
- Edger PP, VanBuren R, Colle M, Poorten TJ, Wai CM, Niederhuth CE, Alger EI, Ou S, Acharya CB, Wang J, et al. 2018. Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (*Fragaria vesca*) with chromosome-scale contiguity. *Gigascience* 7(2):1–7.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable elements families. *Proc Natl Acad Sci U S A.* 117(17):9451–9457.
- Fougère-Danezan M, Joly S, Bruneau A, Gao XF, Zhang LB. 2015. Phylogeny and biogeography of wild roses with specific attention to polyploids. *Ann Bot.* 115(2):275–291.
- Galindo-González L, Mhiri C, Deyholos MK, Grandbastien MA. 2017. LTR-retrotransposons in plants: engines of evolution. *Gene* 626:14–25.
- Grandbastien M-A. 2015. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochim Biophys Acta* 1849(4):403–416.
- Gu T, Han Y, Huang R, McAvoy RJ, Li Y. 2016. Identification and characterization of histone lysine methylation modifiers in *Fragaria vesca*. *Sci Rep.* 6:23581.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321.
- Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered.* 100(5):605–617.
- Han Y, Gasic K, Korban SS. 2007. Multiple-copy cluster-type organization and evolution of genes encoding O-methyltransferases in the apple. *Genetics* 176(4):2625–2635.
- Han Y, Korban SS. 2007. *Spring*: a novel family of miniature inverted-repeat transposable elements is associated with genes in apple. *Genomics* 90(2):195–200.
- Harrell F, Dupont C. 2020. Hmisc: Harrell miscellaneous R package version 4.4.2. Available from: <https://hbiostat.org/R/Hmisc>.
- Harris RS. 2007. Improved pairwise alignment of genomic DNA [PhD thesis]. [State College (PA)]: Pennsylvania State University.
- Henry LK, Gutensohn M, Thomas ST, Noel JP, Dudareva N. 2015. Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. *Proc Natl Acad Sci U S A.* 112(32):10050–10055.
- Henry LK, Thomas ST, Widhalm JR, Lynch JH, Davis TC, Kessler SA, Bohlmann J, Noel JP, Dudareva N. 2018. Contribution of isopentenyl phosphate to plant terpenoid metabolism. *Nat Plants* 4(9):721–729.
- Hibrand Saint-Oyant L, Ruttink T, Hamama L, Kirov I, Lakhwani D, Zhou NN, Bourke PM, Daccord N, Leus L, Schulz D, et al. 2018. A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nat Plants* 4(7):473–484.
- Iwata H, Gaston A, Remay A, Thouroude T, Jeauffre J, Kawamura K, Oyant LH-S, Araki T, Denoyes B, Foucher F. 2012. The *TFL1* homologue *KSN* is a regulator of continuous flowering in rose and strawberry. *Plant J.* 69(1):116–125.
- Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR. 2004. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431(7008):569–573.
- Jiang S, Wang X, Shi C, Luo J. 2019. Genome-wide identification and analysis of high-copy-number LTR retrotransposons in Asian pears. *Genes* 10(2):156.
- Jung S, Lee T, Cheng C-H, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, et al. 2019. 15 years of GDR: new data and functionality in the genome database for Rosaceae. *Nucleic Acids Res.* 47(D1):D1137–D1145.
- Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinformatics* 20(4):1160–1166.
- Knudsen JT, Eriksson R, Gershenzon J, Ståhl B. 2006. Diversity and distribution of floral scent. *Bot Rev.* 72(1):1–120.
- Kobayashi S, Goto-Yamamoto N, Hirochika H. 2004. Retrotransposon-induced mutations in grape skin color. *Science* 304(5673):982.

- Krasileva KV. 2019. The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. *Curr Opin Plant Biol.* 48:18–25.
- Kuhn M, Jackson S, Cimentada J. 2020. Corrr: correlations in R. R package version 0.4.3. Available from: <https://github.com/tidymodels/corrr>, <https://corrr.tidymodels.org>.
- Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30(1):325–327.
- Li W, Lybrand DB, Xu H, Zhou F, Last RL, Pichersky E. 2020. A trichome-specific, plastid-localized *Tanacetum cinerariifolium* Nudix protein hydrolyzes the natural pyrethrin pesticide biosynthetic intermediate trans-Chrysanthemyl diphosphate. *Front Plant Sci.* 11:482.
- Lu H, Giordano F, Ning Z. 2016. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14(5):265–279.
- Lye ZN, Purugganan MD. 2019. Copy number variation in domestication. *Trends Plant Sci.* 24(4):352–365.
- Magnard J-L, Roccia A, Caissard J-C, Vergne P, Sun P, Hecquet R, Dubois A, Hibrand-Saint Oyant L, Jullien F, Nicolè F, et al. 2015. Biosynthesis of monoterpene scent compounds in roses. *Science* 349(6243):81–83.
- Maron LG, Guimarães CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, et al. 2013. Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. *Proc Natl Acad Sci U S A.* 110(13):5241–5246.
- Masure P. 2013. Guide des rosiers sauvages. Paris (France): Delachaux et Niestlé.
- McLennan A. 2013. Substrate ambiguity among the Nudix hydrolases: biologically significant, evolutionary remnant, or both? *Cell Mol Life Sci.* 70(3):373–385.
- Morata J, Marín F, Payet J, Casacuberta JM. 2018. Plant lineage-specific amplification of transcription factor binding motifs by miniature inverted-repeat transposable elements (MITEs). *Genome Biol Evol.* 10(5):1210–1220.
- Nakamura N, Hirakawa H, Sato S, Otagaki S, Matsumoto S, Tabata S, Tanaka Y. 2018. Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. *DNA Res.* 25(2):113–121.
- Ono K, Akagi T, Morimoto T, Wünsch A, Tao R. 2018. Genome resequencing of diverse sweet cherry (*Prunus avium*) individuals reveals a modifier gene mutation conferring pollen-part self-compatibility. *Plant Cell Physiol.* 59(6):1265–1275.
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet.* 39(10):1256–1260.
- Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:e45.
- Prunier J, Caron S, MacKay J. 2017. CNVs into the wild: screening the genomes of conifer trees (*Picea spp.*) reveals fewer gene copy number variations in hybrids and links to adaptation. *BMC Genomics* 18(1):97.
- R Core Team. 2015. R: A language and environment for statistical computing. Vienna (Austria): R Core Team.
- Raguso RA. 2004. Why do flowers smell? The chemical ecology of fragrance-driven pollination. In: Millar JG, Cardé RT, editors. *Advances in insect chemical ecology*. Cambridge (United Kingdom): Cambridge University Press. p. 151–178.
- Raymond O, Gouzy J, Just J, Badouin H, Verdenaud M, Lemainque A, Vergne P, Moja S, Choisine N, Pont C, et al. 2018. The *Rosa* genome provides new insights into the domestication of modern roses. *Nat Genet.* 50(6):772–777.
- Roccia A, Hibrand-Saint Oyant L, Cavel E, Caissard J-C, Machenaud J, Thouroude T, Jeuffre J, Bony A, Dubois A, Vergne P, et al. 2019. Biosynthesis of 2-phenylethanol in rose petals is linked to the expression of one allele of *RhPAAS*. *Plant Physiol.* 179(3):1064–1079.
- Schorr P, Young MA. 2007. Modern roses 12: the comprehensive list of roses in cultivation or of historical or botanical importance. Shreveport (LA): American Rose Society.
- Shirai K, Hanada K. 2019. Contribution of functional divergence through copy number variations to the inter-species and intra-species diversity in specialized metabolites. *Front Plant Sci.* 10:1567.
- Srouji JR, Xu A, Park A, Kirsch JF, Brenner SE. 2017. The evolution of function within the Nudix homology clan. *Proteins* 85(5):775–811.
- Sun P, Dégut C, Réty S, Caissard J-C, Hibrand-Saint Oyant L, Bony A, Paramita SN, Conart C, Magnard J-L, Jeuffre J, et al. 2020. Functional diversification in the Nudix hydrolase gene family drives sesquiterpene biosynthesis in *Rosa x wichurana*. *Plant J.* 104(1):185–199.
- Sun P, Schuurink RC, Caissard JC, Huguene P, Baudino S. 2016. My way: noncanonical biosynthesis pathways for plant volatiles. *Trends Plant Sci.* 21(10):884–894.
- Thompson JD, Gibson TJ, Higgins DG. 2003. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* 2(2.3):1–22.
- Trhlin M, Rajchard J. 2011. Chemical communication in the honeybee (*Apis mellifera* L.). *Vet Med.* 56(6):265–273.
- Verde I, Abbott AG, Scalabrini S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al.; International Peach Genome Initiative. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet.* 45(5):487–494.
- Verde I, Jenkins J, Dondini L, Micali S, Pagliarini G, Vendramin E, Paris R, Aramini V, Gazza L, Rossini L, et al. 2017. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics* 18(1):225.
- Wang A, Yamakake J, Kudo H, Wakasa Y, Hatsuyama Y, Igarashi M, Kasai A, Li T, Harada T. 2009. Null mutation of the *MdACS3* gene, coding for a ripening-specific 1-aminocyclopropane-1-carboxylate synthase, leads to long shelf life in apple fruit. *Plant Physiol.* 151(1):391–399.
- Wang L, Peng Q, Zhao J, Ren F, Zhou H, Wang W, Liao L, Owiti A, Jiang Q, Han Y. 2016. Evolutionary origin of Rosaceae-specific active non-autonomous *hAT* elements and their contribution to gene regulation and genomic structural variation. *Plant Mol Biol.* 91(1–2):179–191.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 8(12):973–982.
- Wissemann V. 2003. Conventional taxonomy of wild roses. In: Roberts A, Debener T, Gudini S, editors. *Encyclopedia of rose science*. London (United Kingdom): Elsevier. p. 111–117.
- Xiang Y, Huang CH, Hu Y, Wen J, Li S, Yi T, Chen H, Xiang J, Ma H. 2017. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol Biol Evol.* 34(2):262–281.
- Yoshimura K, Shigeoka S. 2015. Versatile physiological functions of the Nudix hydrolase family in *Arabidopsis*. *Biosci Biotechnol Biochem.* 79(3):354–366.
- Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, et al. 2019. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun.* 10(1):1494.
- Zhang S-D, Jin J-J, Chen S-Y, Chase MW, Soltis DE, Li H-T, Yang J-B, Li D-Z, Yi T-S. 2017. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytol.* 214(3):1355–1367.
- Zhao D, Ferguson AA, Jiang N. 2016. What makes up plant genomes: the vanishing line between transposable elements and genes. *Biochim Biophys Acta* 1859(2):366–380.
- Zhu ZM, Gao XF, Fougere-Danezan M. 2015. Phylogeny of *Rosa* sections *Chinenses* and *Synstylae* (Rosaceae) based on chloroplast and nuclear markers. *Mol Phylogenet Evol.* 87:50–64.
- Żmienko A, Samelak A, Kozłowski P, Figlerowicz M. 2014. Copy number polymorphism in plant genomes. *Theor Appl Genet.* 127(1):1–18.

General outline chapter 3:

Chapter 3 “A conserved bifunctional geranyl/farnesyl diphosphate synthase in *Rosaceae* provides cytosolic GPP precursor for NUDX1-1a-dependent geraniol biosynthesis in rose flowers”

Corentin CONART, Dikki PEDENLA BOMZAN, Xing-Qi HUANG, Jean-Etienne BASSARD, Saretta N. PARAMITA, Denis SAINT-MARCOUX, Aurélie RIUS-BONY, Gal HIVERT, Anthony ANCHISI, Hubert SCHALLER, Latifa HAMAMA, Jean-Louis MAGNARD, Agata LIPKO, Ewa SWIEZEWSKA, Jame PATRICK, Laurence HIBRAND SAINT-OYANT, Michel ROHMER, Efraim LEWINSOHN, Natalia DUDAREVA, Sylvie BAUDINO, Jean-Claude CAISSARD and Benoît BOACHON.

This chapter is under preparation for submission

In the two previous articles, we described evolution and specialization of *NUDX1-1a* in *Rosaceae*, *Rosidae*, and *Rosa*. However, we did not answer the question about the origin of GPP, the substrate of NUDX1-1a enzyme. Indeed, NUDX1-1a is clearly cytosolic but, in general, GPP is biosynthesized in plastids. In this article we asked the question : where does the cytosolic GPP come from? Thanks to physiological experiments with stable isotopic labelled precursors, and specific inhibitors of the terpene pathways, we demonstrated that the GPP was biosynthesized in the cytosol by the MVA pathway. We also cloned all IDSs expressed in rose petals to find the enzyme responsible of the GPP biosynthesis. We characterized RcG/FPPS1, a mutated FPPS, that could synthesized both GPP and FPP in the cytosol. We then focused on the evolution of this bi-functional activity in *Rosaceae* and *Rosids*, and found that this novel function appeared before *Rosaceae* diversification. Furthermore, we found the mutation responsible of this novel function in the active site, we modeled the enzyme, and we built site-directed mutagenesis enzymes, and finally explained this bi-functional activity.

In this work, I did all the molecular biology work (DNA constructs, qRT-PCR, site-directed mutagenesis...), and the *in vitro* characterization of IDSs after heterologous expression in *E.coli* and purification of proteins. I also took part in the identification of amino acids that can be responsible of the bi-functional activity of RcG/FPPS1. I optimized the culture condition of roses *in vitro*, in greenhouse, the transient experiments of petals, and I optimized all physiological experimentations such as the use of inhibitors and precursors of the MVA and MEP pathways.

I- Introduction

Roses are known for centuries for their pleasant characteristic fragrance and esthetic morphological traits appealing to humans (Krüssmann, 1981). Despite the wide diversity of volatiles emitted by thousands of rose hybrids created so far, geraniol and its derivatives such as citronellol, for example, are essential compounds contributing to unique and well-known rose scents.

Unlike other monoterpenes, the carbon skeleton of geraniol is identical to that of its precursor GPP. This structural feature initially allowed to hypothesize that geraniol could be formed by the action of phosphatase. However, the characterization of the first GES from sweet basil (Iijima *et al.* 2004) revealed that geraniol biosynthesis does not rely on phosphatase enzyme and involves, like in the case of other monoterpene synthases, the formation of a carbocation intermediate from GPP substrate in plastids. Since then, this canonical geraniol biosynthetic pathway was described in most geraniol-producing plants, except for roses. Instead of plastidic GES, a cytosolic RhNUDX1 hydrolase was shown to be responsible for the production of geraniol in flowers of rose hybrids (Magnard *et al.* 2015). NUDX hydrolases are conserved enzymes found in all types of organisms and considered as “housecleaning” proteins associated with cell detoxification such as the dephosphorylation of organic pyrophosphates (Srouji *et al.* 2017). However, in rose hybrids and wild species producing geraniol, *NUDX1-1a* is highly and specifically expressed in petals and its encoding protein dephosphorylates GPP to GP in the cytosol (Conart *et al.* 2022). A yet unknown phosphatase is assumed to catalyze the final step of geraniol biosynthesis. The fact that the rose *NUDX1-1a* is localized exclusively in the cytosol, and not in plastids where GPP is generally assumed to be synthesized from five-carbon building blocks, IPP and its isomeric form DMAPP derived from the MEP pathway, raised yet the unanswered question about the origin of GPP metabolized by cytosolic *NUDX1-1a* in roses.

In plants, IPP and DMAPP are synthesized by two alternative and compartmentally separated pathways, the plastidic MEP pathway and the MVA pathway distributed between the cytosol, endoplasmic reticulum, and peroxisomes (Hemmerlin *et al.* 2012; Tholl, 2015). These two pathways are connected via a metabolic “cross-talk”, which is species- and organ-specific (Vranová *et al.* 2012). In general, the MVA-derived IPP and DMAPP are used by FPPS to produce FPP for cytosolic biosynthesis of sesquiterpenes and triterpenes and mitochondrial biosynthesis of ubiquinones and polyprenols. The MEP-derived plastidic IPP and DMAPP serve as precursors for GPP and GGPP, and ultimately of monoterpenes, diterpenes and tetraterpenes such as carotenoids. All plant short-chain *trans*-IDS using IPP and DMAPP substrates are homodimeric enzymes except for GPPSs (Chang *et al.* 2010), which have both homodimeric and heterodimeric architecture depending on plant species. The homomeric GPPSs have been described in gymnosperm and some angiosperm species (Bouvier *et al.* 2000; van Schie *et al.* 2007; Schmidt and Gershenzon, 2008; Hsiao *et al.* 2008; Rai *et al.* 2013), while the heteromeric GPPSs have been reported only in angiosperms so far including *A. thaliana*, *M. piperita*, *S. lycopersicum*, *A. majus*, *C. roseus* and *H. lupulus* (Burke and Croteau, 2002; Tholl *et al.* 2004; Orlova *et al.* 2009; Wang and Dixon, 2009; Rai *et al.* 2013; Hivert *et al.* 2020;). The heterodimeric GPPSs consist of a LSU, usually exhibiting GGPPS activity alone, and a SSU, which is itself generally catalytically

inactive but upon interaction with the large subunit, favors GPP formation. Thus, due to the existence of cross-talk between the two terpenoid biosynthetic pathways via the exchange of IPP and GPP (Bick and Lange, 2003), the cytosolic GPP in roses could derive from products of the MVA, the MEP pathway, or both, and rely either on homodimeric or heterodimeric GPPs or on FPPS-like enzyme producing GPP as was reported in insects and *Lithospermum erythrorhizon* roots (Vandermoten *et al.* 2008; Ueoka *et al.* 2020; Suttiyut *et al.* 2022).

In this study we investigated the origin of cytosolic GPP metabolized by NUDX1-1a in rose flowers to produce geraniol. Using OB flowers, we demonstrated that GPP is synthesized in the cytosol through the MVA pathway. We biochemically characterized the five IDS candidates retrieved from OB genome of which one cytosolic FPPS-like enzyme, named RcG/FPPS1, exhibited bifunctional G/FPPS activity *in vitro*. When the five IDS candidates were co-expressed in *N. benthamiana* leaves with *RcNUDX1-1a*, only *RcG/FPPS1* co-expression enabled increased geraniol production, while its expression alone increased FPP-derived capsidiol accumulation. In addition, *RcG/FPPS1* expression profile in rose flowers was found to be rhythmical and to be associated with both GPP and FPP rhythmical accumulations before geraniol emission. Finally, transient down- and upregulation of *RcG/FPPS1* in OB flowers resulted in increased and decreased geraniol emission, respectively, and had similar effect on emissions of FPP-derived germacrene D and carotenoid derived dihydro- β -ionol (dh β -ionol), thus providing genetic evidence for *RcG/FPPS1* endogenous bifunctional activity. Reconstitution of RcG/FPPS evolution in addition to biochemical characterization of *Rosaceae* G/FPPSs and RcG/FPPS1 point mutations we identified two amino acid conserved only in *Rosaceae* species and involved in the apparition of bifunctional activity of this enzyme. This enabled to solve the unanswered question about the cytosolic biosynthesis of monoterpenes in *Rosaceae* species such as strawberries and the origin of cytosolic GPP necessary for NUDX1-1a-dependent biosynthesis of geraniol in rose flowers.

II- Results

Geraniol is synthesized via the MVA pathway in rose flowers.

To investigate the biosynthetic origin of cytosolic GPP metabolized by RcNUDX1-1a for geraniol production in rose flowers several independent approaches were used including (i) feeding of rose flowers with stable isotope-labeled pathway-specific precursors, (ii) inhibitory experiments with pathway-specific inhibitors, (iii) analysis of natural isotopic ratios and (iv) analysis of subcellular GPPS activity. [2-¹³C]-mevalonolactone (¹³C-MEV), a specific precursor of the MVA pathway, rapidly incorporated in germacrene D, known to be produced from the MVA pathway-derived precursors by the previously characterized cytosolic sesquiterpene synthase (Guterman *et al.* 2002), and in geraniol reaching 64 % and 42 % of labeling after 54 h of feeding, respectively. Only relatively low incorporation of ²H₂-DOX, a specific precursor of the MEP pathway, was observed in these compounds (Figure 3.1A). In comparison, both precursors were slowly incorporated in dh β -ionol, an apocarotenoid volatile synthesized from plastidic GGPP (Huang *et al.* 2009a; b), resulting in only 13 % for ¹³C-MEV and 21 % for ²H₂-DOX of labeling after 54 h of feeding (Fig. 1A). Consistently, treatment of rose flowers with mevinolin, a specific inhibitor of the MVA pathway, decreased geraniol emission by 68 %, which was similar to a reduction in

emission of germacrene D (73 %; Figure 3.1B). On the other hand, fosmidomycin, a specific inhibitor of the MEP pathway, did not affect the emission of these compounds, suggesting that the cytosolic MVA pathway makes a major contribution to geraniol biosynthesis in rose flowers. In contrast, the emission of $\delta\beta$ -ionol was reduced by both inhibitors suggesting that both the MVA and MEP pathways contribute to its formation and that a substantial metabolic exchange from the cytosolic MVA pathway to plastids occurs in rose flowers. The emission of TMB, a non-terpene volatile produced from acetyl-CoA independently of the MEP or MVA pathways (Scalliet *et al.* 2002) was used as a control and remained unchanged in these experiments.

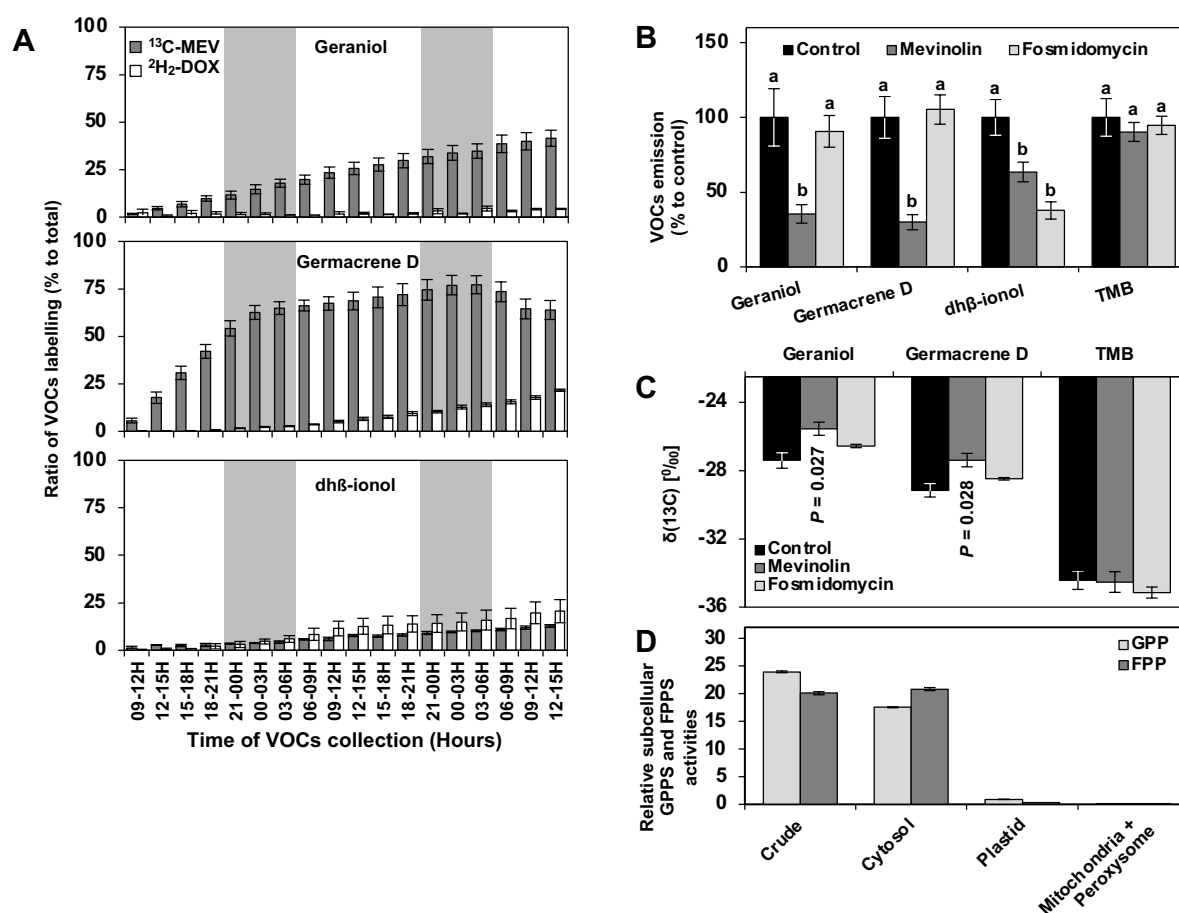


Figure 3.1 | Geraniol is synthesized through the MVA pathway in cytosol of rose flowers. (A) Kinetic of stable isotopes ^{13}C -MEV or $^2\text{H}_2$ -DOX incorporation in VOCs emitted from OB flowers analyzed by GC-MS and expressed as % relative to total, mean \pm SEM n = 4. (B) Effect of inhibitor treatments mevinolin or fosmidomycin on VOCs emitted from OB flowers analyzed by GC-MS and expressed as % relative to untreated control flowers. Data are means \pm SEM, n = 18. Letters indicate statistically significant differences between samples analyzed by ANOVA and Tukey post-hoc test. (C) Natural $^{12}\text{C}/^{13}\text{C}$ isotopic ratios $\delta(13\text{C})$ of the VOCs emitted from rose petals treated or not with mevinolin or fosmidomycin analyzed by GC-IRMS. Data are means \pm SEM, n = 5. *p* values indicate statistically significant differences analyzed by two-tailed Student's *t*-test. (D) GPPs and FPPs activities in crude protein extracts from OB flowers and enriched sub-cellular cytosolic, plastidic and mitochondrial/peroxysomal fractions. 20 μg of proteins per fraction were incubated with IPP and DMAPP and MeOH extracts were analyzed by LC-MS/MS. Activities are expressed in pKat normalized by the ratio of proteins recovered from each compartment compared to crude. Data are means \pm SEM, n = 3.

It was previously shown that the natural $^{12}\text{C}/^{13}\text{C}$ isotope ratios $\delta(^{13}\text{C})$ of terpenes differs depending on their origin due to preference of the pyruvate dehydrogenase from the MVA pathway for lighter substrates and that the use of pathway-specific inhibitors can modify this $\delta(^{13}\text{C})$ as a result of reduced contribution from the inhibited pathway (Jux *et al.* 2001). VOCs emitted from rose petals treated or not with inhibitors were thus analyzed by gas chromatography-isotope ratio mass spectrometry (GC-IRMS) (Figure 3.1C). Consistently with results obtained with inhibitory and labelling experiments the $\delta(^{13}\text{C})$ values measured for germacrene D and geraniol were similarly and significantly increased in roses treated with mevinolin compared to control flowers as a result of decreased contribution of the MVA pathway. $\delta(^{13}\text{C})$ values for these compounds were unaffected by fosmidomycin treatment confirming their essentially MVA pathway origin. Finally, $\delta(^{13}\text{C})$ values were not affected by any treatment for the control TMT and could not be determined for dh β -ionol due to low signal and potential impurity in the pics.

Since both inhibitory and the precursor feeding experiments suggested that the MVA pathway predominantly contributes to geraniol formation in rose flowers, but GPPSs in most angiosperm species are localized in plastids, GPPS and FPPS activities were analyzed in different subcellular fractions. Both GPPS and FPPS activities were found in the cytosolic fraction, suggesting that GPP formation occurs mainly in the cytosol in rose flowers (Figure 3.1D). In addition, plastidic and mitochondrial fractions contained negligible GPPS and FPPS activities relative to the cytosol. GGPPS activity was not detected in these experiments because of being below the detection limit. Analysis of activities of marker enzymes and chlorophyll content in isolated fractions confirmed that the cytosolic fraction was barely contaminated by the other compartments (Figure S3.1). Taken together, these results indicate that in rose flowers, GPP is formed in the cytosol by unknown GPPS synthase from precursors derived from the MVA pathway.

The cytosolic RcG/FPPS1 exhibits both GPP and FPP synthase activities.

To identify the enzyme responsible for cytosolic GPP production in rose we blast searched the recently published genome of OB (Hibrand Saint-Oyant *et al.* 2018; Raymond *et al.* 2018) for *trans*-short-chain IDSs. This analysis yielded six IDS candidates including two putative FPPSs, RcG/FPPS1 (see below) and RcFPPS2, two putative large subunits of G(G)PPS designated RcGGPPS.LSU1 and RcGGPPS.LSU2, one small subunit of GPPS designated RcGPPS.SSU and one putative homodimeric GPPS designated RcGPPS.HOMO (Figure 3.2A) which all contained the expected conserved domains (Figure S3.2). RNAseq analysis of flowers from OB and nine rose hybrids producing different levels of geraniol revealed that all IDSs were expressed except for *RcGGPPS.LSU2* which was not included in further experiments (Figure 3.2B). The expression of the five IDSs was found relatively low compared to *NUDX1-1a* and none of them was correlated to geraniol content contrarily to *NUDX1-1a* expression. A protein targeting prediction program, TargetP, predicted that only FPPSs have no transit peptide and are likely localized in the cytosol while the small and two large subunits of GPPS contain putative plastidic transit peptides and RcGPPS.HOMO contains a putative mitochondrial transit peptide (Figure S3.2A). To verify the program predictions, the CDS of each IDS candidate was fused to the N-terminus

of the coding sequence of the green fluorescent protein (eGFP) reporter for transient expression in OB petals. The subcellular localization observed in conical epidermal cells of OB petals was consistent with the program predictions (Figure 3.2C and Figure S3.2A). RcGGPPS.SSU and RcGGPPS.LSU1 were both colocalized with the plastidic marker and also co-localized with each other. RcGGPPS.HOMO was colocalized with the mitochondrial marker while both RcG/FPPS1 and RcFPPS2 were localized in the cytosol like RcNUDX1-1a.

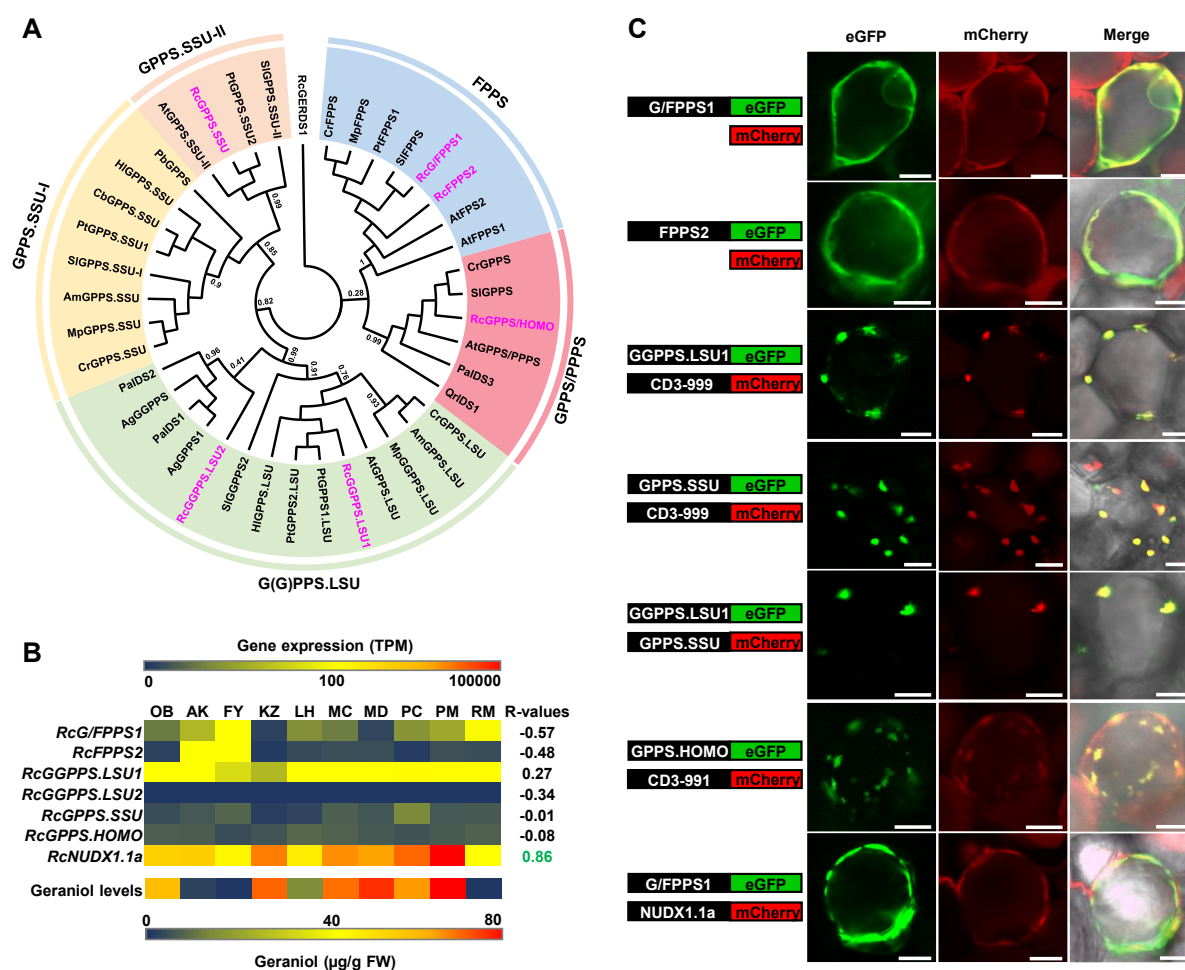


Figure 3.2 | The OB genome contains six IDS candidates of which two FPPS-like synthases localized in the cytosol. (A) Maximum likelihood tree of protein sequences of the six OB *trans*-short-chain IDSs (highlighted in pink) with characterized IDSs from *Abies grandis* (Ag), *Antirrhinum majus* (Am), *Arabidopsis thaliana* (At), *Clarkia breweri* (Cb), *Catharanthus roseus* (Cr), *Humulus lupulus* (Hl), *Mentha piperita* (Mp), *Phalaenopsis bellina* (Pb), *Picea abies* (Pa), *Populus trichocarpa* (Pt), *Quercus robur* (Qr), *Rosa chinensis* ‘Old Blush’ (Rc) and *Solanum lycopersicum* (Sl). (B) Transcriptomic analysis of the six OB IDSs and NUDX1.1a compared to geraniol levels in 10 rose hybrids including *R. chinensis* ‘Old Blush’ (OB), *R. x hybrida* ‘AkitoR’ (AK), *R. x hybrida* ‘The Fairy’ (FY), *R. x damascena* ‘Kazanlik’ (KZ), *R. x odorata* ‘Lady Hillingdon’ (LH), *R. x hybrida* ‘Mc Cartney’ (MC), *R. x hybrida* ‘Marius Ducher’ (MD), *R. x hybrida* ‘Pariser Charmehy’ (PC), *R. x hybrida* ‘Papa Meiland’ (PM) and *R. x hybrida* ‘Rouge Meiland’ (RM). Heatmap on top panel shows the expression levels of the six IDS candidates and NUDX1.1a from RNAseq analysis on open flowers from the 10 rose cultivars and expressed as transcripts per million (TPM). Data are means from 3 biological replicates. Heatmap on bottom panel shows the geraniol levels analyzed in the opened flowers of these 10 cultivars. Data are means from 3 biological replicates. Pearson correlation coefficients (R-values) comparing the transcript levels of each

gene to geraniol content in the 10 rose cultivars is shown on the right of top panel. R-values with significant p values (≤ 0.001) are highlighted in green. (C) Subcellular localization of OB IDSs in epidermal cells of OB petals. Schematic diagrams of the constructs used are shown on the left with corresponding transient expression in OB cells on the right. CDS of OB IDS candidates and *RcNUDX1.1a* were fused with eGFP or mCherry at their C-terminus as indicated. mCherry was fused to different subcellular markers for cytosol (untargeted mCherry), plastids (CD3-999) and mitochondria (CD3-991). Merge channel shows both eGFP and mCherry signals with bright field. Scale bar, 10 μm .

To determine whether identified cytosol-localized candidates can produce GPP, recombinant mature RcG/FPPS1 and RcFPPS2 proteins were heterologously produced in *E. coli*. Copurification of 6-His-tagged and untagged RcG/FPPS1 confirmed that it can form a homodimer (Figure S3.3A). Purified proteins were incubated with IPP and DMAPP followed by product analysis by LCMS (Figure 3.3A and Figure S3.4A). Surprisingly, RcG/FPPS1 efficiently converted IPP and DMAPP substrates into both GPP and FPP, while RcFPPS2 almost exclusively produced FPP. When both enzymes were incubated with IPP and GPP, FPP was formed (Figure S3.4B). In addition, the GPP/FPP product ratio of RcG/FPPS1 was modified when changing the ratio of IPP and DMAPP concentrations enabling to produce more GPP when DMAPP was provided in excess and more FPP when IPP was provided in excess (Figure 3.3B). Analysis of the kinetic parameters revealed that RcG/FPPS1 had a high affinity toward IPP in the presence of saturating DMAPP concentration with an apparent K_m of 0.44 μM (Table 3.1). The kinetic parameters of RcFPPS2 for the GPP production could not be determined as the enzyme always produced FPP. However, when kinetic parameters were determined for FPP production and for IPP in the presence of the GPP as allylic cosubstrate, RcG/FPPS1 had a four-fold lower affinity for IPP than RcFPPS2. Moreover, the RcG/FPPS1 affinity was 120-fold lower for IPP when GPP was supplied instead of DMAPP as cosubstrate and RcG/FPPS1 produced FPP 15-fold less efficiently (k_{cat}/K_m ratio) than GPP. Thus, RcG/FPPS1 gained the ability to produce GPP while keeping similar catalytic efficiency for FPP production as RcFPPS2 suggesting that RcG/FPPS1 is a bifunctional geranyl/farnesyl diphosphate synthase whose product specificity could be modified by substrate availability.

Table 3.1 | Kinetic parameters of RcOB IDS candidates and RcG/FPPS1 mutants^a.

| Substrates | Product | Protein | K_m (μM) | K_{cat} (sec^{-1}) | k_{cat}/K_m ($\mu\text{M}^{-1}.\text{sec}^{-1}$) | |
|------------|--------------|---------|-------------------------|---------------------------------|--|----------------------|
| DMAPP | Δ IPP | GPP | G/FPPS1 | 0.44 \pm 0.03 | 0.321 \pm 0.009 | 0.74 \pm 0.03 |
| | | | GGPPS.LSU1 + GPPS.SSU | 2.50 \pm 0.15 | 0.023 \pm 0.0008 | 0.0091 \pm 0.0003 |
| | | | GGPPS.LSU1 | 6.74 \pm 0.84 | 0.021 \pm 0.002 | 0.0032 \pm 0.0001 |
| | | | GPPS.HOMO | 5.00 \pm 1.29 | 0.001 \pm 0.0001 | 0.0001 \pm 0.00001 |
| | | | FPPS2 | nd | nd | nd |
| GPP | Δ IPP | FPP | G/FPPS1 | 53.1 \pm 5.1 | 2.87 \pm 0.26 | 0.05 \pm 0.000 |
| | | | FPPS2 | 13.1 \pm 3.4 | 0.91 \pm 0.17 | 0.07 \pm 0.006 |
| | | | G/FPPS1-F88Y | 9.08 \pm 0.74 | 0.54 \pm 0.03 | 0.06 \pm 0.002 |
| | | | G/FPPS1-V123I | 14.74 \pm 1.94 | 0.93 \pm 0.08 | 0.06 \pm 0.003 |
| | | | G/FPPS1-F88Y/V123I | 12.78 \pm 3.07 | 1.78 \pm 0.33 | 0.14 \pm 0.011 |

^aAll values represent mean \pm SE, n = 3. nd, not determined (below detection limit). Δ indicates the variable substrate.

To test whether the other isolated candidates, RcGGPPS.LSU1, RcGPPS.SSU and RcGPPS.HOMO could produce GPP, their corresponding mature recombinant proteins were analyzed. The identification of RcGPPS.SSU suggested that rose petals might contain a heterodimeric GPPS, which was confirmed by copurification of 6-His-tagged RcGGPPS.LSU1 with untagged RcGPPS.SSU (Figure S3.3B). Thus, product specificities of the small and large subunits were characterized by themselves and after copurification (Figure S3.4, Table 3.1). In the presence of IPP and DMAPP, GPP production was found for all of the other IDS candidates but was never found as the end product. RcGPPS.HOMO produced GPP and traces of FPP as well as to RcGPPS.SSU, surprisingly, an unexpected result which could be due to its alternative SARM motif (Figure S3.2). RcGGPPS.LSU1 also produced GPP in addition to GGPP and the ratio between products only slightly changed when the small subunit was coexpressed with RcGGPPS.LSU1 (Figure S3.4A). Kinetic evaluation of recombinant proteins revealed that their *K_m* for IPP ranged from 2.5 to 6.74 μ M. Out of four proteins the heterodimeric RcGGPPS.LSU1/GPPS.SSU (Figure S3.5) had the highest affinity towards IPP and the highest catalytic efficiency, which were respectively 5.7-fold and 81-fold lower than that of RcG/FPPS1 (Table 3.1). To assess the final product formed by each IDS, the enzymes were incubated with GPP or FPP in the presence of IPP (Figure S3.4C and S3.4D). FPP was the final product of RcGPPS.HOMO and RcGPPS.SSU. RcGGPPS.LSU1 produced GGPP as a final product and its coexpression with RcGPPS.SSU made it slightly less efficient in converting IPP and FPP to GGPP supporting that the heterodimeric GPPS in rose is not efficient to produce GPP as sole product.

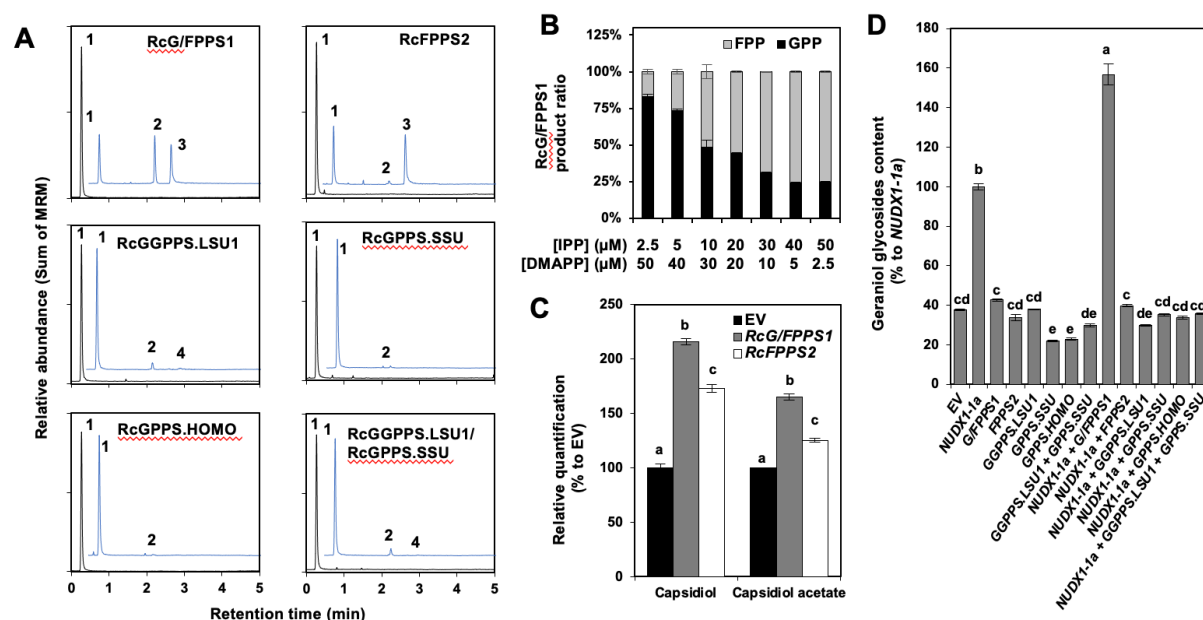


Figure 3.3 | RcG/FPPS1 is a bifunctional enzyme producing GPPS and FPPS in vitro and in planta. (A) LC-MS/MS chromatograms of the reaction products from in vitro incubation of the five OB IDS candidates and the heterodimeric GPPS (RcGGPPS.LSU1/RcGPPS.SSU) with IPP and DMAPP. Chromatograms in blue represent the incubation of 250 ng of each protein for 5 minutes with 10 μ M of the IPP and DMAPP substrates at 30 $^{\circ}$ C. Chromatograms in black represent the negative controls with incubation of the corresponding boiled proteins. 1 = DMAPP + IPP, 2 = GPP, 3 = FPP, 4 = GGPP. (B) LC-MS/MS quantification of the RcG/FPPS1 GPP/FPP product ratio depending on the indicated

concentrations of IPP and DMAPP provided. Incubations were realized as described in (A). Data are means \pm SEM, $n = 3$. (C) GC-MS quantification of the FPP-derived capsidiol and capsidiol acetate accumulating in *N. benthamiana* leaves transiently expressing empty vector (EV), *RcG/FPPS1* or *RcFPPS2* alone. Data are relative quantification to EV set at 100 %, mean \pm SEM, $n = 3$. Letters depict statistically significant differences analyzed by ANOVA and Tukey post-hoc test. (D) LC-MS/MS quantification of geraniol glycosides extracted from *N. benthamiana* leaves transiently expressing the five OB IDS candidates and the heterodimeric GPPS (*RcGGPPS.LSU1/RcGPPS.SSU*) alone or in coexpression with *RcNUDX1-1a*. Data are relative quantification to *NUDX1.1a* set at 100 %, mean \pm SEM, $n = 4$. Letters depict statistically significant differences analyzed by ANOVA and Tukey post-hoc test.

Then, the bifunctional G/FPP biosynthetic capacity of *RcG/FPPS1* was assessed *in planta*. First, *RcG/FPPS1* and *RcFPPS2* were transiently expressed alone in *N. benthamiana* leaves which, compared to control leaves, resulted in increased accumulation of capsidiol and capsidiol acetate (Figure 3.3C), two FPP derived sesquiterpenoids known to accumulate in tobacco leaves (Takahashi *et al.* 2005; Li *et al.* 2015). This result confirmed that *RcG/FPPS1* exhibits FPPS activity *in planta*. Second, all IDS candidates were transiently expressed alone and in combination with *RcNUDX1-1a*. Previously it was reported that *N. benthamiana* leaves produced geraniol glycosides and *RcNUDX1-1a* overexpression increased their levels (Magnard *et al.* 2015; Sun *et al.* 2020). Coexpression of *RcNUDX1.1a* only with *RcG/FPPS1* and not any of the other IDS candidates further increased geraniol glycoside levels when compared to *RcNUDX1.1a* overexpression alone (Figure 3.3D) supporting that *RcG/FPPS1*, in addition to FPPS activity, also exhibit GPPS activity *in planta*. Overall, these results provide biochemical and genetic evidence that the cytosolic *RcG/FPPS1* is a bifunctional enzyme capable of producing *in planta* cytosolic FPP in addition to GPP which results in geraniol production if *NUDX1.1a* is expressed.

RcG/FPPS1 rhythmic expression precedes rhythmic GPP and FPP production.

To further investigate the endogenous role of *RcG/FPPS1* for geraniol production in OB flowers we analyzed RNA-seq datasets generated from petals over six flower development stages starting from closed buds to fully opened flowers (Figure S3.5A) and at two time points during a day/night cycle (12:00 h and 24:00 h) (Figure S3.5B). All identified IDSs exhibited very low expression levels compared to *RcNUDX1.1a* and none of them correlated with geraniol emission which was confirmed by RT-qPCR analysis (Figure S3.5C and S3.5D). In contrast, *RcNUDX1-1a*, *HMGRs* and *HDR*, genes known to encode proteins catalyzing rate-limiting steps in the geraniol, the MVA and the MEP biosynthetic pathways, respectively, exhibit a significant positive correlation with geraniol emission (Figure S3.5B). However, because RNA-seq analysis revealed that *RcG/FPPS1* is more expressed at night (24H) than during day (12H) (Figure S3.5B), a detailed analysis of IDSs' expression over a daily light/dark cycle by RT-qPCR revealed that only *RcG/FPPS1* out of five identified IDSs exhibits a strong rhythmic expression pattern with a maximum expression at night (Figure 3.4A and Figure S3.5D). Outstandingly, this rhythmic expression was shown to correlate with the peaks of both rhythmic GPP and FPP accumulation inside petal tissues (Figure 3.4B) and ultimately geraniol emission (Figure 3.4C) yet with a 3 h lag time between the peaks of *RcG/FPPS1* expression, GPP and FPP levels and then geraniol emission (Figure

3.4A to 3.C). These results further supported that *RcG/FPPS1* is associated with GPP and FPP formation in rose petals and ultimately with their derived VOCs products.

RcG/FPPS1 down- and up-regulations in rose flowers affect geraniol, germacrene and dhB-ionol emissions

To confirm the endogenous function of *RcG/FPPS1*, OB flowers were agroinfiltrated with *RcG/FPPS1*_RNAi construct or empty vector control. The resulting down-regulation of *RcG/FPPS1* observed in OB petals (Figure 3.4D) resulted in significant decrease of geraniol emission as well as germacrene D and dhB-ionol emissions compared to control while TMB was not affected (Figure 3.4E). Accordingly, rose flowers overexpressing *RcG/FPPS1* (Fig. 4F) emitted significant higher amounts of these three compounds compared to control while TMB emission was unchanged (Fig. 4G). These results were consistent with inhibitory and precursor feeding experiments (Figure 3.1) as well as *RcG/FPPS1* biochemical characterization (Figure 3.3) and confirmed that *RcG/FPPS1* is involved in both GPP and FPP rhythmic production in rose flowers. This leads both to NUDX1.1a-dependent geraniol production as well as cytosolic sesquiterpenes and plastidic apo-carotenoid VOCs, whose emissions, similarly to geraniol, were found to be rhythmic (Figure 3.4C) and follow *RcG/FPPS1* expression.

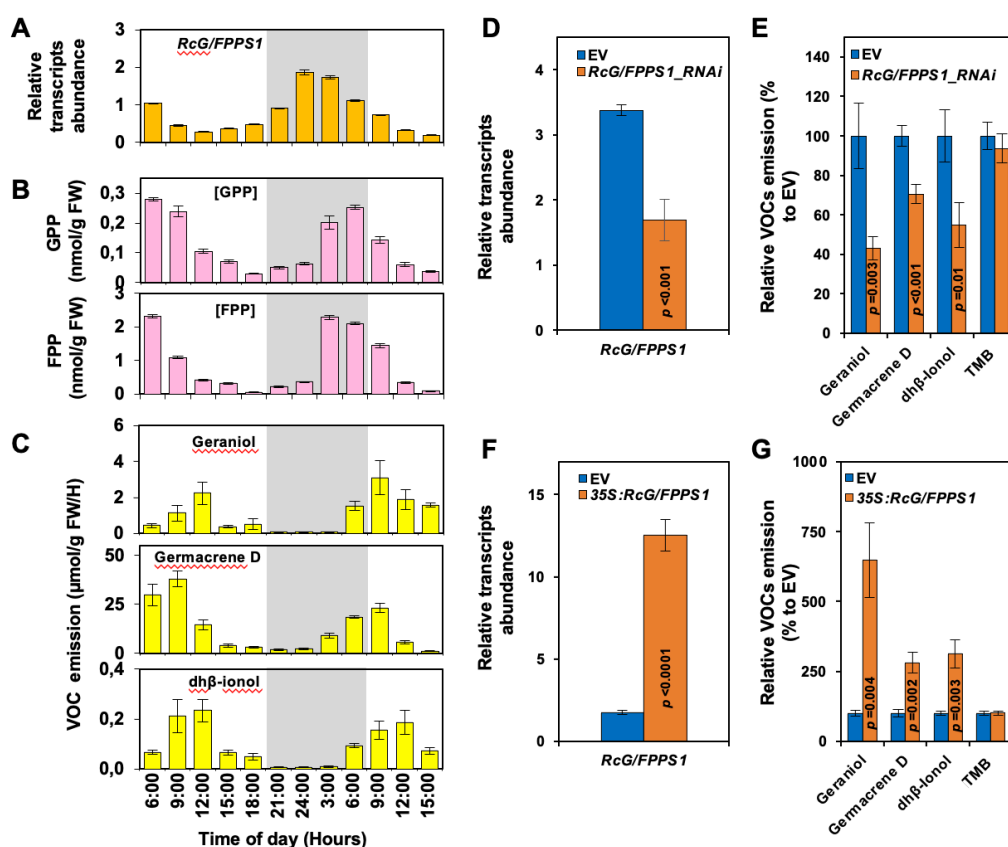


Figure 3.4 | *RcG/FPPS1* rhythmic expression pattern precedes GPP and FPP accumulation and geraniol emission in rose flowers. (A) RT-qPCR quantification of *RcG/FPPS1* transcript levels during a day/night cycle in OB petals, means \pm SEM, $n = 4$. (B) LC-MS/MS quantification of GPP and FPP accumulations during a day/night cycle in OB petals, means \pm SEM, $n = 4$. (C) GC-MS quantification of

VOC emissions during a day/night cycle in OB petals means \pm SEM, $n = 6$ to 13 . Night time is highlighted in gray in (A) to (C). (D) RT-qPCR quantification of *RcG/FPPS1* transcript levels in OB flowers agroinfiltrated with empty vector control or *RcG/FPPS1*_RNAi construct. (E) GC-MS quantification of VOCs emitted from OB flowers agroinfiltrated as described in (D). Data in (D) and (E) are means \pm SEM, $n = 12$. (F) RT-qPCR quantification of *RcG/FPPS1* transcript levels in OB flowers agroinfiltrated with empty vector control or 35S:*RcG/FPPS1* construct. (F) GC-MS quantification of VOCs emitted from OB flowers agroinfiltrated as described in (E). Data in (F) and (G) are means \pm SEM, $n = 6$. p values in (D) to (G) indicate statistically significant differences analyzed by two-tailed Student's t -test.

The G/FPPS activity is conserved in *Rosaceae*

To understand when the G/FPPS activity discovered in rose evolved from ancestral and bona fide FPPSs, we investigated the FPPS family in *Rosid* species. FPPS protein sequences were retrieved from both *Rosoideae* (*R. chinensis*, *F. vesca*, and *Potentilla micrantha*) and *Amygdaloideae* (*Malus domestica*, *Prunus persica*, and *Prunus domestica*) subfamilies within *Rosaceae*, and several *Rosid* species including previously characterized FPPSs such as *A. thaliana* and *P. trichocarpa* FPPSs. Phylogenetic analysis and comparisons of genes surrounding *RcG/FPPS1* and *RcFPPS2* (Figure 3.5A and B) enabled to place each *Rosid* FPPSs into either the FPPS1s or FPPS2s groups (highlighted on Figure 3.5A) as orthologs according to their shared synteny with one or the other rose (G)/FPPSs. This confirmed that *RcG/FPPS1* evolved from a common ancestor in *Rosids* originally exhibiting FPPS activity such as *AtFPPS1* and *PtFPPS1* (Cunillera *et al.* 1996; Keim *et al.* 2012; Lackus *et al.* 2019). To identify when the bifunctional G/FPPS activity emerged in *Rosids* within the FPPS1 group we characterized several of its members including the phylogenetically distant to rose and previously characterized *PtFPPS1* used as control, the uncharacterized *MtFPPS1* belonging to the *Fabaceae* family and the closer FPPS1s from *F. vesca* and *P. persica* from *Rosoideae* and *Amygdaloideae* subfamilies, respectively. Their biochemical activities characterized *in vitro* (Figure 3.5C) and *in planta* (Figure 3.5D) revealed that the G/FPPS activity is well conserved in both *Rosoideae* and *Amygdaloideae*. *FvG/FPPS1* and *PpG/FPPS1*, similarly to *RcG/FPPS1*, produced both GPP and FPP *in vitro* and their coexpression with *RcNUDX1.1a* in *N. benthamiana* leaves resulted in about 2-fold increase in geraniol production compared to leaves expressing *RcNUDX1.1a* only. Interestingly, *PtFPPS1* and *MtFPPS1*, which still produced mostly FPP *in vitro*, also produced small amount of GPP, but 4-fold and 8-fold less importantly than *RcG/FPPS1*, respectively (Figure 3.5C). This was confirmed *in planta* where their co-expression with *RcNUDX1.1a* in *N. benthamiana* leaves also slightly improved geraniol production by about 1.2-fold compared to leaves expressing *RcNUDX1.1* alone (Figure 3.5D). These results suggested that G/FPPS1 was conserved in *Rosacea* species but probably started to evolve earlier in *Rosids*.

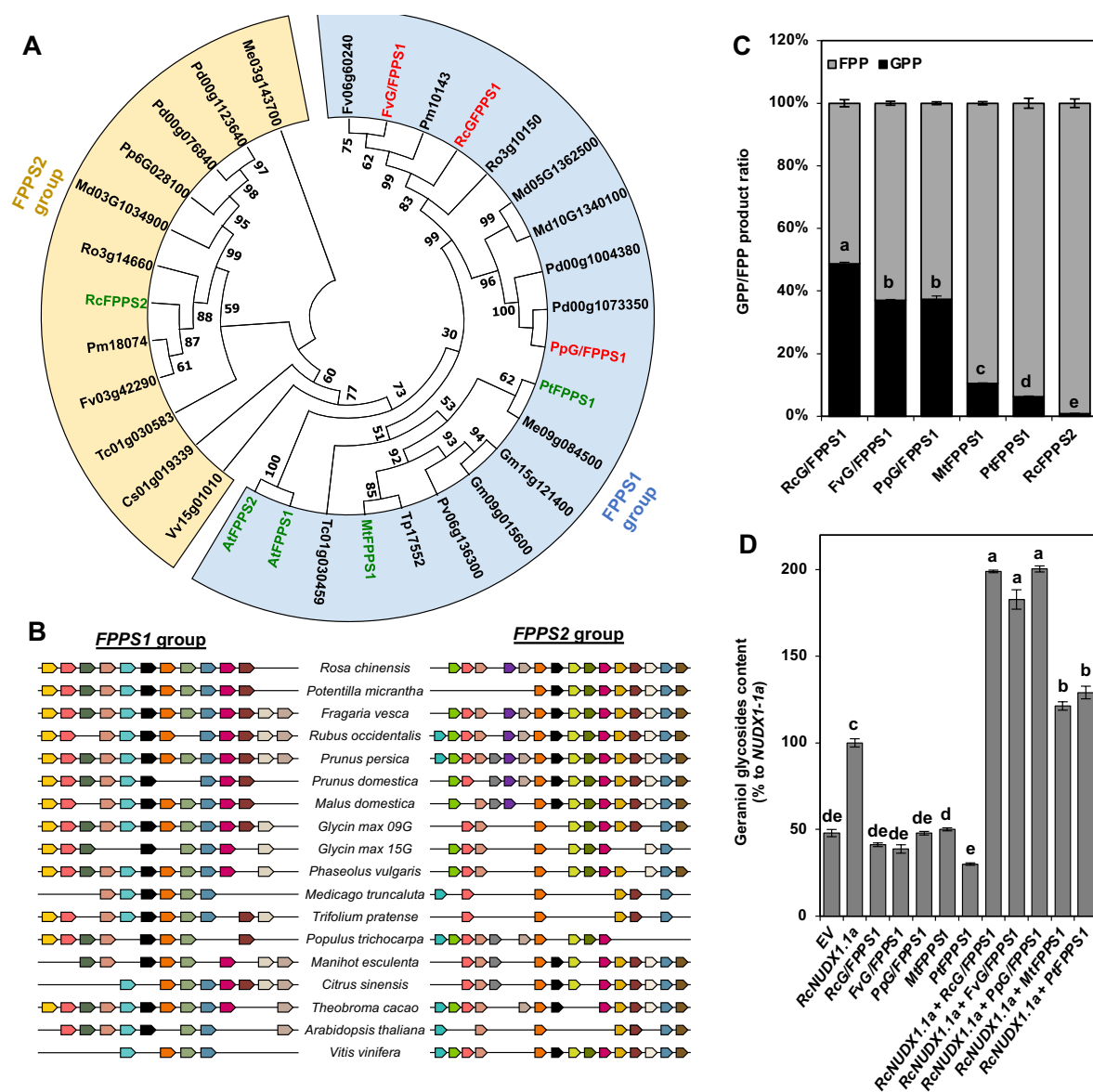


Figure 3.5 | The bifunctional G/FPPS activity is conserved in *Rosaceae*. (A) Maximum likelihood tree of Rosid FPPSs from OB (Rc), *Arabidopsis thaliana* (At), *Citrus sinensis* (Cs), *Fragaria vesca* (Fv), *Glycin max* (Gm), *Malus domestica* (Md), *Manihot esculenta* (Me), *Medicago truncatula* (Mt), *Prunus dulcis* (Pd), *Phaseolus vulgaris* (Pv), *Populus trichocarpa* (Pt), *Potentilla micrantha* (Pm), *Prunus persica* (Pp), *Rubus occidentalis* (Ro), *Theobroma cacao* (Tc), *Trifolium pratense* (Tp) and *Vitis vinifera* (Vv). FPPS1 (blue) and FPPS2 (green) groups are highlighted and were defined according to phylogeny and synteny analysis (See (B)). Characterized classical FPPSs are depicted in green and characterized G/FPPS1s are depicted in red. (B) Microsynteny of genes encoding for *Rosids* FPPSs from FPPS1s and FPPS2s groups as used in (A). (C) LC-MS/MS quantification of the GPP/FPP product ratio formed by 250 ng of the indicated FPPSs incubated for 20 minutes with 10 μ M DMAPP and IPP. Data are means \pm SEM, n = 3. (D) LC-MS/MS quantification of geraniol glycosides accumulating in *N. benthamiana* leaves transiently expressing the indicated FPPSs alone of in coexpression with *RcNUDX1-1a*. Data are relative quantification to *NUDX1.1a* set at 100 %, mean \pm SEM, n = 4., Letters in (C) and (D) indicate statistically significant differences analyzed by ANOVA and Tukey post-hoc test.

At least two amino acid residues evolved in *Rosaceae* G/FPPS are responsible for emergence the GPPS activity

Rosaceae G/FPPSs gain their GPPS activity due to lowering their affinity toward IPP in the presence of GPP as cosubstrate (Table 3.1) likely through point mutations during the evolution. Thus, *Rosaceae* G/FPPS protein sequences were searched for conserved amino acid residues which evolved during the *Rosaceae* G/FPPSs evolution from ancestral orthologous FPPSs. When compared with the bona fide FPPSs from *Rosids* known to produce FPP, sequence alignment showed that a short QLLQ sequence at positions 59-62 (QLLQ59-62), the phenylalanine residue 88 (F88), and valine residue 123 (V123) are found in the *Rosaceae* G/FPPSs but not in the FPPSs (Figure 3.6A). To test whether these identified conserved residues could contribute to lower affinity for GPP resulting in its release from active site, a computational strategy combining homology-based modeling and molecular dynamics (MD) simulations was used. Binding free energies for GPP in presence of IPP and Mg²⁺ ions, which mimics the substrate-binding step for FPPS activity, were calculated for RcG/FPPS1 (wild-type) and its mutated versions harboring ancestral orthologous amino acids for the identified conserved residues. Reversing both wild type F88 to Y88 and V123 to I123 independently but not QLLQ59-62 to KLLK59-62 resulted in significant lower binding free energy for GPP (Figure 3.6B). This suggested that Y88 and I123 in ancestral orthologs favors FPP production and their substitution to F88 and V123 during *Rosaceae* evolution could be responsible for gain of GPPS activity.

To validate the model prediction and also examine the contribution of each conserved residue on the G/FPPS activity, several single and multiple RcG/FPPS1 mutants were generated by site-directed mutagenesis, and the corresponding recombinant proteins were biochemically characterized. Consistent with the model prediction, both F88Y and V123I substitutions independently reduced GPP formation similarly by about 55% compared to WT RcG/FPPS1 with a simultaneous increase in FPP formation (Figure 3.6C). A synergistic effect was detected in the double mutant (F88Y/V123I) with GPP production further reduced by an additional 45%. Kinetic analysis of both single and double mutants revealed that both F88Y and V123I substitutions significantly increase (3.6 - 5.9-fold) the affinity of proteins for IPP in the presence of GPP as cosubstrate (Table 3.1) thus making the mutated RcG/FPPS1 biochemically very similar to the relatively distant RcFPPS2 although the latter produces almost exclusively FPP (Figure 3.6C). Moreover, *N. benthamiana* leaves coexpressing *RcNUDX-1-a* with each single mutant *RcG/FPPS1-F88Y* or *RcG/FPPS1-V123I* produced significantly less geraniol glycosides than leaves co-expressing *RcNUDX1-1a* with the native wild-type *RcG/FPPS1* (Figure 3.6D). Coexpression of *NUDX1.1a* with the double mutant *RcG/FPPS-F88Y-V123I* further decreased geraniol production. When the opposite mutations, Y88F and I123V, were introduced into RcFPPS2, the recombinant mutant enzyme started to release GPP as a product (Figure 3.6C).

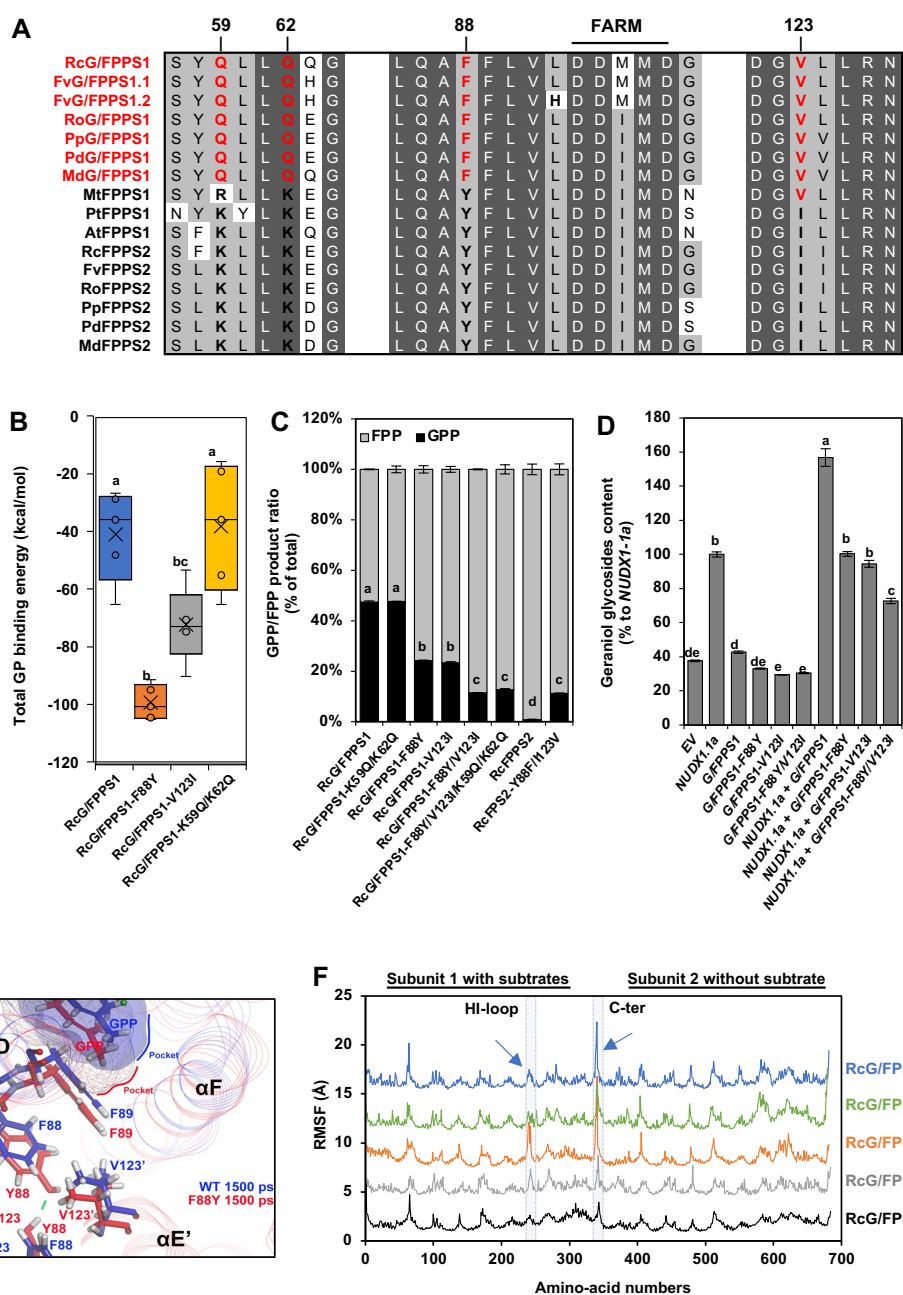


Figure 3.6| Two amino acid are critical for the GPPS activity of RcG/FPPS1. (A) Protein sequence alignment of *Rosids* FPPS1s and FPPS2s including characterized *Rosaceae* G/FPPS1s. Only relevant sequence parts are shown to focus on residues conserved only in *Rosaceae* G/FPPS1s (highlighted in red) compared to bona fide FPPS1s and FPPS2s. (B) Binding free energies of GPP for RcG/FPPS1 and its mutated version F88Y, V123I and K59Q/K62Q after molecular dynamic (MD) simulation of 1500 ps with GPP, IPP and Mg^{2+} placed in the active site. Data are means \pm SEM of the five last recorded frames. (C) LC-MS/MS quantification of the GPP/FPP product ratio formed by the indicated recombinant proteins incubated for 20 minutes with 10 μ M DMAPP and IPP. Data are means \pm SE, n = 3. (D) LC-MS/MS quantification of geraniol glycosides accumulated in *N. benthamiana* leaves transiently expressing the indicated genes alone of in coexpression with *RcNUDX1-1a*. Data are relative quantification to *NUDX1.1a* set at 100 %, mean \pm SEM, n = 4. Letters in (B), (C) and (D) indicate statistically significant differences between samples analyzed by ANOVA and Tukey post-hoc test. (E) Superimposed model of RcG/FPPS1 homodimer and its mutated version RcG/FPPS1-F88Y at 1500 ps of MD simulation as described in (C). The model focuses of the elongation pocket containing GPP (top) and the interaction of both protein subunits around the amino acids Y88-Y'88 and V123-V123'. α helices

are shown as well as the volumes of the elongation pockets for both protein versions and the hydrogen bound in between both Y88-Y88' subunits (green dash). (F) Root mean square fluctuations (RMSF) of each amino acid of RcG/FPPS1 homodimer and its mutated versions during the 1500 ps of MD simulation as described in (C). Two groups of amino acid, the Hi-loop and the C-ter of subunit 1 displaying improved movements in mutants compared to WT RcG/FPPS1 are highlighted.

To validate the model prediction and also examine the contribution of each conserved residue on the G/FPPS activity, several single and multiple RcG/FPPS1 mutants were generated by site-directed mutagenesis, and the corresponding recombinant proteins were biochemically characterized. Consistent with the model prediction, both F88Y and V123I substitutions independently reduced GPP formation similarly by about 55% compared to WT RcG/FPPS1 with a simultaneous increase in FPP formation (Figure 3.6C). A synergistic effect was detected in the double mutant (F88Y/V123I) with GPP production further reduced by an additional 45%. Kinetic analysis of both single and double mutants revealed that both F88Y and V123I substitutions significantly increase (3.6 - 5.9-fold) the affinity of proteins for IPP in the presence of GPP as cosubstrate (Table 3.1) thus making the mutated RcG/FPPS1 biochemically very similar to the relatively distant RcFPPS2 although the latter produces almost exclusively FPP (Figure 3.6C). Moreover, *N. benthamiana* leaves coexpressing *RcNUDX-1-a* with each single mutant *RcG/FPPS1-F88Y* or *RcG/FPPS1-V123I* produced significantly less geraniol glycosides than leaves co-expressing *RcNUDX1-1a* with the native wild-type *RcG/FPPS1* (Figure 3.6D). Coexpression of *NUDX1.1a* with the double mutant *RcG/FPPS-F88Y-V123I* further decreased geraniol production. When the opposite mutations, Y88F and I123V, were introduced into RcFPPS2, the recombinant mutant enzyme started to release GPP as a product (Figure 3.6C).

A closer look at the RcG/FPPS model after the 1500 ps of MD simulations revealed that double F88Y/V123I and single F88Y substitutions leads to the quick formation of a hydrogen bond between introduced tyrosine of each subunit (Y88-Y88') (Figure 3.6E). The inability of F88 to form such a hydrogen bond likely renders the binding site shallow and less favorable for GPP binding preventing its full conversion to FPP. Modeling also revealed that isoleucine substituted to valine from both subunits (I123-123') is located in the direct vicinity of this hydrogen bound and likely participate to better metabolism of GPP potentially through hydrophobic interactions. Moreover, F88 and V123 in WT RcG/FPPS1 preclude movements of the HI-loop (Substrates entrance in active site: Rondeau *et al.* 2006; Kavanagh *et al.* 2006; Zhang *et al.* 2009) and C-terminal of enzyme (Known to refold and interact with substrate in active site: Fisher *et al.* 2011) observed in RcG/FPPS1 mutants within MD simulation (Figure 3.6F). Substitutions of Q to K in the QLLQ59-62 sequence had no effect on the G/FPPS activity of RcG/FPPS1 neither alone nor in combination with F88Y/V123I (Fig. 6C). Taken together, these results suggest that conserved F88 and V123 in *Rosaceae*, which evolved from ancestral Y88 and I123 are critical for bifunctional G/FPPS activity.

III- Discussion

It is generally accepted that in most plants monoterpenes are produced in plastids by monoterpene synthases using GPP derived from the MEP pathway (Tholl, 2015). However, there are known few exceptions showing that in *Rosaceae* and *Boraginaceae* (*L. erythrorhizon*) monoterpenes or their derivatives are produced in the cytosol. Examples include the existence of a cytosolic pinene synthase in wild strawberries (Aharoni *et al.* 2004a), the cytosolic RcNUDX1.1a enzyme responsible for geraniol production in roses (Magnard *et al.* 2015), and the detection of GPPS activity in the cytosolic fraction of *L. erythrorhizon* (Sommer *et al.* 1995). Although cytosolic GPP could originate via crosstalk from the MEP pathway, feeding experiments with labeled precursors in *F. vesca*, *Rubus idaeus*, *Rose hybrida* (Francis and O’connell, 1969; Hampel *et al.* 2006, 2007), and in *L. erythrorhizon* (Inouye *et al.* 1979) suggested that the MVA pathway is the major contributor to the biosynthesis of monoterpenes and GPP-derived compounds in these plant species. Recently *LeGPPS* encoding a cytosolic GPPS involved in shikonin biosynthesis has been isolated and characterized from *L. erythrorhizon* (Ueoka *et al.* 2020; Suttiyut *et al.* 2022), but it still remained unknown how cytosolic GPP production is achieved in roses for RcNUDX1-1a-dependent geraniol biosynthesis and in other *Rosaceae* for MVA-dependent monoterpene productions. Here, we provide biochemical (Figure 3.3, Figure S3.4 and Table 3.1) and genetic evidences (Figure 3.3B, Figure 3.4D to 3.4G) that a cytosolic bifunctional RcG/FPPS1 (Fig. 2B and Fig.3A) is involved in GPP biosynthesis in rose petals. This enzyme, which belongs to FPPS family (Fig. 2A) still keeps its ability to synthesize FPP but produces GPP 15-fold more efficiently (k_{cat}/K_m ratio) than FPP *in vitro* (Table 3.1) and is the only one out of five identified IDS candidates able to provide cytosolic GPP substrate for RcNUDX1.1a *in planta* (Figure 3.3D) while also providing FPP for capsidiol production (Figure 3.3C). Naturally, GPP is an intermediate product of FPPSs, but it does not release efficiently from the active site. The first short-chain prenyltransferase that can generate both GPP and FPP has been identified in aphids (Vandermoten *et al.* 2008), but was not reported in plants so far.

The biochemical characterization of orthologous FPPSs from several *Rosid* species guided by phylogenetic analysis (Figure 3.5A and 3.5B) showed that the bifunctional G/FPPS activity was gained during evolution of *Rosids* from the classical FPPSs and was conserved in *Rosaceae* including OB, *F. vesca* and *P. persica*. Thus, these results also provide an answer to a long-standing question about the cytosolic and MVA-dependent monoterpene production in strawberries and raspberries (Figure 3.5C). This so far unknown G/FPPS activity is likely conserved in all *Rosaceae* species in which monoterpenes are known to greatly contribute to fruit aroma (Zhang *et al.* 2022)

Moreover, FPPSs in FPPS1 groups (MtFPPS1 and PtFPPS1), but not FPPS2, exhibited dual activity although producing GPP at lower levels and less efficiently than *Rosaceae* G/FPPs (Figure 3.5C and 3.5D), a property of enzymes within FPPS1 clade, which was not shown before. This residual GPPS activity of cytosolic FPPSs might contribute to the cytosolic GPP pool used for the synthesis of geranylated plant secondary metabolites as observed for cannabinoids and flavonoids (de Bruijn *et al.* 2020) and potentially yet unknown proteins.

At least two amino acids F88 and V123, which are highly conserved in *Rosaceae* are responsible for efficient GPP production in addition to FPP by G/FPPSs (Figure 3.6). They modulate the affinity of enzymes towards GPP substrate and the product selectivity in favor of GPP production in addition to keeping FPP production (Table 3.1). The reasons why these mutations were selected during evolution for high cytosolic production of monoterpenes in specialized plant organs (Aharoni *et al.* 2004b) are currently unknown. It is possible that G/FPPS were selected during evolution because of the deficient capacity of plastidic heterodimeric GPPSs to produce efficiently GPP as observed in rose (Figure 3.3, Figure S3.4 and Table 3.1).

A comparison of the cytosolic GPP formation in roses and *L. erythrorhizon* reveals that both GPP-producing enzymes evolved independently from conventional FPPSs. In the case of the rose enzyme, it had occurred through mutations leading to a bifunctional protein that favors GPP formation while continuing to produce FPP (Figure 3.3A). In contrast, in *L. erythrorhizon*, the enzyme produced exclusively GPP with the His residue at position 100 adjacent to the first Asp-rich motif, contributing to product specificity (Ueoka *et al.* 2020). Out of the three *Rosaceae* G/FPPS1 characterized in this study, RcG/FPPS1, FvG/FPPS1 and PpG/FPPS1 (Figure 3.5C and 3.D), all contained Leu instead of H100. Interestingly, FvG/FPPS1.2 contains H100 in addition to F88 and V123 (Figure 3.6A). However, even though it still produced both GPP and FPP to similar ratios, its activity was extremely low compared to the other characterized *Rosaceae* G/FPPS1s (Figure S3.6). These results suggest that the emergence of cytosolic GPPS activity occurred at least twice in phylogenetically distantly related plant species within *Rosaceae* and *Boraginaceae* families via different but convergent evolutionary mechanisms.

The non-canonical biosynthetic pathway for geraniol production in roses requires both GPP availability and NUDX1.1a enzyme in the cytosol. While the NUDX1.1a specialization appeared during diversification of *Rosa* genus (Conart *et al.* 2022), our results show that the ability to synthesize the cytosolic GPP preceded NUDX1.1a acquisition in several *Rosaceae* genus indicating that two evolutionary events enabled high geraniol production in rose flowers.

The discovery of bifunctional G/FPPSs suggests that plants have two types of homodimeric GPPSs in addition to heterodimeric ones, consisting of a catalytically inactive SSU interacting with a LSU. All previously characterized homodimeric GPPSs are evolutionarily related to GGPPSs forming a GGPPS-like group (Tholl, 2015), while recently isolated LeGPPS (Ueoka *et al.* 2020) and *Rosaceae* G/FPPSs identified in this study are evolutionarily related to FPPSs. Within this FPPSs-like group, there are two types of enzymes: bifunctional enzymes like RcG/FPPS1, FvG/FPPS1 and PpG/FPPS1 (Figure 3.5C) and monofunctional enzymes with strict product specificity like LeGPPS. Moreover, members of the FPPSs-like group are localized in the cytosol (Figure 3.2B; Ueoka *et al.* 2020) in contrast to the heterodimeric GPPSs which are always found in the plastids and homodimeric GPPSs, most of which were shown to be localized in mitochondrial (Rai *et al.* 2013). Further studies of GPPSs and monoterpene production will uncover how widely the recruitment of FPPSs to produce GPP was used in plants.

IV- Materials and Methods

Plant material and growth conditions

Rosa chinensis “Old Blush” (OB) obtained from Guillot nursery owner and *Nicotiana benthamiana* plants were cultivated in growth chambers under white fluorescent lamps with a light intensity of 150-200 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ with 50-60 % relative humidity at 21-22°C during the 16-h day period and 17°C during the 6-h night period. OB plants were cultivated in a soil mix containing 3/10 perlite, 1/10 pouzzolane and 6/10 Klasmann BP-substrate (Klasmann-Deilmann GmbH, Germany). *N. benthamiana* plants were cultivated in a soil mix containing 3/8 perlite and 5/8 KlasmannTS3 substrate.

Chemicals and synthesis of 1-deoxy-[5,5- $^2\text{H}_2$]-D-xylulose ($^2\text{H}_2$ -DOX)

All chemicals, authentic VOC standards, inhibitors and labelled precursors were purchased from Sigma-Aldrich (St. Louis, MO, USA) otherwise stated. $^2\text{H}_2$ -DOX was synthesized as described in ref. (Meyer *et al.* 2004). Chrysal professional 3 (Chrysal) was purchased from Chrysal France.

Inhibitory and stable isotope-labelling experiments

For labelling experiments, OB flowers at stage 5 (See Figure S3.5A) were harvested at 9H00. To favour precursors uptake, pedicel was removed by cutting in the middle of receptacle as transversal cross section and roses were then placed in 1 % chrysal solution for control treatment or 1 % chrysal solution containing 2 mg/ml (3R,S)-[2- ^{13}C]-mevalonolactone or 1mg/ml $^2\text{H}_2$ -DOX in 1-ml plastic tubes. VOCs emitted from flowers were collected every 3h for 54h before GC-MS analysis as described below. For the calculation of VOCs labelling by each precursor, the percentage of labelled versus unlabelled VOCs was determined based on specific ions for each VOC as listed in (Fig. S7). For inhibitory experiment, OB flowers at stage 5 were harvested at 9:00 h, pedicel was removed as described above and each flower was cut in two equal longitudinal cross sections. One part was placed in 1% chrysal solution in 1-ml plastic tubes for control treatment and the other part was placed similarly in control solution containing either 1 mg/mL of mevinolin or 1 mg/mL fosmydomycin. 24 hours following treatment, roses parts were placed in volatile collection system and VOCs were collected for 3 h before by GC/MS quantification of VOCs. For calculation each rose part treated with inhibitor was compared to its corresponding part treated with control treatment.

VOC collection from the Headspace of OB flowers

VOCs emitted from OB flowers were collected as previously described (Boachon *et al.* 2019) with minor modifications. Briefly, 1 flower (or 1/2 flower part) per sample were placed in 1-L glass jar and VOCs were pumped out with flow of 100 ml/min during the indicated times for each experiment and trapped on glass cartridges containing 30 mg of Porapak Q 80/100 mesh. Inlet air was filtered on similar cartridges filed with 100 mg of Porapak Q. Volatiles were eluted with 200 μl of hexane spiked with 50 μM of camphor used as internal standard before GC-MS analysis.

GC-MS analysis

Samples were analyzed on an Agilent 8890 gas chromatograph system (Agilent Technologies), equipped with multiple purpose sampler (Gerstel, Germany) and connected to an Agilent 5977B mass detector (Agilent Technologies). Samples were run. Samples were injected in the injector set at 250°C,

on a HP-5MS-UI column (30 m x 0.25 mm x 0.25 μm) (Agilent Technologies) using a program consisting of 0.5 min at 50°C, followed by 15 or 20°C/min to 320°C, then 5min at 320°C, with a flow of 1.8mL/min of He as carrier gas. Mass spectrometer ionization was set at 250 °C, energy was set at 70 eV and mass spectrum was scanned from 30 to 300 amu. Products were identified based on their retention times and electron ionization mass spectra compared to those of authentic standards or those present in the NIST2017 and WILEY libraries. Quantification of compounds was performed using the Mass Hunter quantitative software (Agilent Technologies) using response factors of commercially available authentic standards relative to the internal standard and normalized to the weight of tissues.

Isotopic $^{12}\text{C}/^{13}\text{C}$ analysis of VOCs by GC-C-IRMS

OB flowers were harvested and treated as described above for inhibitory experiments. To obtain sufficient amounts of compounds to be analyzed, VOCs collection was upscaled by using 8 to 10 flowers per sample with 24 hours collection period and VOCs were eluted with 100 μL of hexane spiked with 50 μM of camphor used as internal standard. Samples were analyzed by gas chromatography-isotope ratio mass spectrometry (GC-IRMS) on a 6890N GC gas chromatograph system (Agilent Technologies) connected to an Isoprime GV isotope ratio mass spectrometer (Elementar UK Ltd, Cheadle, UK) via the GC5MK1 GV combustion or pyrolysis interface. All samples were analysed in duplicate and verified by bracketing the analytical sequence with citronellol standard and geraniol standard every six analyses. The standards were calibrated to the international references V-PDB using the EA-IRMS technique. An amount of 1 μL of liquid sample was injected using a 7683 Agilent Technologies auto-sampler into the 6890N gas chromatograph equipped with an Agilent HP-5 column (60 m x 0.320 mm x 0.25 μm I.D film thickness) used in chromatographic separation. The injection port was held at 250°C, fitted with a split liner containing glass wool, and operated in split mode. The oven temperature program started at 50°C during 1 min, increased to 325°C at a rate of 10 °C min^{-1} , and held for 5 min. Compounds were separated at a flow rate of 1 $\text{mL}\cdot\text{min}^{-1}$. A heart split valve connected to the outlet of the GC column enabled the components to be obtained either in the combustion/pyrolysis tube or in the FID detector. This process allows elimination of the solvent peak as well as unwanted compounds. The combustion of carbon in carbon dioxide was undertaken using a tube containing copper oxide maintained at 850°C in a furnace. Silver wool was placed in the second portion of the tube to trap halogens and sulfur. A Nafion water trap was placed between the combustion furnace and the IRMS to remove water produced during the combustion. $^{12}\text{C}/^{13}\text{C}$ isotope ratios $\delta(13\text{C})$ of VOCS were calculated as previously described (Cuchet *et al.* 2021). To verify that the headspace collection procedure did not favor lighter or heavier VOC isotopes, $\delta(13\text{C})$ values of VOCs from a mix of authentic standards were calculated and compared when run directly on GC-C-IRMS or after being collected in the volatile collection system similarly as the rose samples.

GPSS and FPSS activities on purified subcellular compartments

20 g of OB petals at flower stage 5 were collected at 9:00 h, ground with a blender for 5 sec with ice cold buffer A (0.5 M sorbitol, 20 mM HEPES, 10 mM KCl, 1 % PVP40, 10% glycerol, 5 mM 2-Mercaptoethanol, pH 7) and filtered through two layers of Miracloth (Merck Millipore) to obtain crude

extract proteins. All steps were realized on ice or at 4°C. The crude extract was centrifuge 15 min at 2 500 g to pellet plastids. Supernatant was carefully collected and centrifuged 40 min at 15 000 g to obtain mitochondria and peroxyzomes in the pellet and cytosol in the supernatant. The pellet containing plastids was carefully resuspended in 4 ml of ice-cold buffer B (0.66 M sorbitol, 20 mM HEPES pH 7) and overlaid with a discontinues Percoll gradient consisting of 2 ml 80/20 % (v/v) Percoll/buffer B then 4 ml of 20/80 % (v/v) Percoll/buffer B. Separation of plastids was realized by centrifugation at 5 000 g for 20 min. Purified plastids were carefully collected at the interface of both gradient solutions and washed twice with buffer B with centrifugation at 3 000 g for 5 min before being resuspended in essay buffer. Mitochondria where resuspend in essay buffer. Total protein concentrations of each subcellular fraction were quantified using the Bradford method (Bradford, 1976). Purity of each fraction was verified by measuring activities of marker enzymes: peroxisomal catalase (Aebi, 1984), mitochondrial fumarase (Huang *et al.* 2015), cytosolic alcohol dehydrogenase (Tong and Lin, 1988), and plastidic chlorophyll content by a confocal microscopy (D'Andrea *et al.* 2014).

Identification of candidate genes, sequence analysis, phylogeny and synteny analysis

OB and *Rosids* *trans*-short-chain IDSs where identified by BLAST search using previously characterized IDSs on available OB genomes (Raymond *et al.* 2018; Hibrand Saint-Oyant *et al.* 2018) and Genome Database for Rosaceae (GDR, www.rosaceae.org) (Jung *et al.* 2019). *FPPS1* and *FPPS2* genes from *Rosids* were identified by BLAST search on GDR and Phytozome.v13 database (<https://phytozome-next.jgi.doe.gov/>) and wered used for comparative synteny with *RcG/FPPS1* and *RcFPPS2* using the synteny tool of Phytozome.v13 (Goodstein *et al.* 2012). Multiple sequence alignments were computed using the MUSCLE (codon) algorithm (gap open, -2.9; gap extend, 0; hydrophobicity multiplier, 1.2; clustering method, UPGMA) implemented in MEGA X(Kumar *et al.* 2018). Based on alignments, trees were reconstructed with MEGA X using a maximum likelihood algorithm and the Jones-Taylor-Thornton model with a gamma distribution of rate among sites (JTT+G). All site coverage was used. Ambiguous bases were allowed at any position. Bootstrap resampling analysis with 1000 replicates were performed to evaluate the topology of the generated trees.

Sampling procedure for Gene expression analysis and metabolite profiling

For analysis of OB flowers compared to nine Rose hybrids (Figure 3.2), petals from the indicated plants were harvested on flowers stage 5 corresponding to open flower with yellow stamen visible. Part of the samples were frozen in liquid nitrogen and stored at -80 °C for latter RNA extraction and part of the samples were placed in hexane containing 50 µM camphor used as internal standard pour metabolite profiling. For analysis OB flowers during developmental stages (Figure 3.4 and Figure S3.5), OB petals where harvested at 9:00h at different developmental stages as shown in (Figure S3.5A) : closed buds with closed sepals and closed coloured petals visible (stage 1), closed buds with sepals starting to open and closed red petals (Stage 2), buds starting to open with sepals curved and closed red petals (Stage 3), open flowers with stamens not visible day1 (stage 4), open flower with yellow stamen visible and flat petals day1 (stage 5), open flower with brown stamen and petals starting to curve day2 (stage 6). Samples were immediately frozen in liquid nitrogen and stored at -80 °C for latter RNA extraction and

quantification of prenyl phosphates. Flowers at the same flower stages were collected similarly for VOCs emission analysis. For analysis of OB flowers during a day/night cycle (Figure 3.4 and Figure S3.5) petals were collected every 3h for 34h starting at 6:00 on day1 with flowers on stage 4 which evolved in stage 5 during day1 and in stage 6 on day2 as illustrated on Fig. S5B. Samples of day1 12H00 and 24H00 were also used from transcriptomic analysis. Petals from at least four roses were pooled per sample. Samples were immediately frozen in liquid nitrogen and stored at - 80°C for latter RNA extraction and quantification of prenyl phosphates. Flowers at the same flower stages were collected similarly for VOCs emission analysis during a day/night cycle.

Gene expression analysis (transcriptomic and RT-qPCR) compared to VOCs profiling

For RNA extraction, tissues were ground in Nitrogen with mortar and pestle. 50 mg of petals per sample were then extracted using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer instruction using the on-column DNase treatment. RNA sample qualities were then verified with a bioanalyzer (Agilent Technologies) following manufacturer's guidelines. For RNA-seq analysis, for each developmental stage, five replicates with a RIN above 8 were sequenced by BGI Tech Solutions (Hong Kong, China) using the DNB-Seq platform. About 30 million 150 bp paired-end reads were obtained per sample. Reads were checked with fastQC v0.11.9 and cleaned using fastp v0.21 (Chen *et al.* 2018). Reads were then mapped onto the *R. chinensis* genome v2 (Raymond *et al.* 2018) using STAR v2.7.5a (Dobin *et al.* 2013), transcript assembly and coverage values were obtained using StringTie v2.1.4 (Pertea *et al.* 2015). TPMs were derived from coverage using a custom Perl script. For RT-qPCR analysis, reverse transcription (RT) was performed with the All-In-One MasterMix kit (Applied Biological Materials) according to the manufacturer's protocol using 1 µg of total RNA per sample. cDNA samples were then diluted 10 times with ultrapure water. For quantitative real-time PCR (qPCR), 2.5 µl of diluted cDNA were mixed with 5 µL of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), 1 µL of each forward and reverse specific primers (Fig. SX for the list and sequences of primers) and ultrapure water to a final reaction volume of 10 µl. Each sample were run with two technical replicates. qPCR parameters were as followed: 95°C for 30 sec and 45 cycles of (95 °C for 5 sec and 60 °C for 20 sec). Quantification of relative transcripts accumulation was calculated using the efficiency of each gene-specific primers and the EΔCt method (Pfaffl, 2001).

Generation of vector constructs

To produce recombinant proteins heterologously in bacterial system and biochemically characterize OB and rosid *IDSs* including mutants, their coding sequences were synthesized by Genscript (Piscataway) into the pET-30a(+) vector between KpnI and Sall restriction sites. To study subcellular localization, CDSs of OB *IDSs* with modified 3' ends to remove stop codons were PCR-amplified from OB flower cDNA and inserted in Gateway binary vectors in frame with the CDS of eGFP into pB7FWG2 or with the CDS of RFP into pH7RWG2 (Karimi *et al.* 2007) by Gateway LR reaction using LR-Clonase™II (Invitrogen). To characterize the activity of candidate genes *in planta* through transitory expression in *N. benthamiana* leaves, the CDS of all studied *IDSs* including mutants and *RcNUDX1.1a* were PCR-amplified and inserted into the plant expression vector pCAMBIA3300u by the using the USER cloning

system (New England Biolabs) as previously described (Boachon *et al.* 2015). The *RcG/FPPS1_RNAi* construct was generated by using the Sol Genomics Network VIGS Tool (<http://vigs.solgenomics.net/>) (Fernandez-Pozo *et al.* 2015) which implemented the OB genome upon request to identify the best target region of the *RcG/FPPS1* CDS and confirm that the designed double stranded RNA (dsRNA) trigger would not result in off-target interference. *RcG/FPPS1_RNAi* construct included two spliced *RcG/FPPS1* complementary DNA fragments corresponding to nucleotides 57-557 and 357-557 (in the antisense orientation) to create a hairpin structure. The sequence was then synthesized by Genscript (Piscataway) into pUC57 vector with EcoRV restriction sites. The *RcG/FPPS1_RNAi* obtained sequence was then cloned into pCAMBIA3300 between BamHI and KpnI restriction sites by standard cloning ligation. Primers used for cloning are listed in Figure S3.8.

Sub-cellular localisation

Fluorescent protein fusion constructs were transformed into *Agrobacterium tumefaciens* strain GV3101. After overnight culture in LB containing the appropriate antibiotics at 28°C with 200 rpm shaking, the bacteria were centrifuged 8 min at 4 000 g at room temperature and washed 3 times in MgCl₂ 10 mM, MES 10 mM, pH 5.7 before being incubated with 500 µM acetosyringone and then diluted to OD_{600nm} = 0.6 for each construct. OB petals were co-transformed by agro-infiltration with cultures of equal density of agrobacteria harboring the genes of interest in a ratio of 1/1 (v/v). 4 days after, petal pieces were excised for observation by laser scanning confocal microscopy. Cell imaging was performed using a LSM780 confocal system microscope (Zeiss, Germany). Images were recorded with a Plan-Apochromat 20x/0.8 M27 objective. Excitation/emission wavelengths were 488/493-556 nm for eGFP constructs, and 561/590-628 nm for mCherry constructs, respectively. Images were processed via contrasts and brightness corrections with ImageJ software (NIH, USA). Vectors CD3-999 (plastid targeted mCherry) and CD3-991 (mitochondria targeted mCherry) (Nelson *et al.* 2007), purchased from the Arabidopsis Biological Resource Center (<https://abrc.osu.edu/>), and pCAMBIA2300 (untargeted 35S:eGFP) were used as colocalization markers.

In vitro characterization of IDSs

E. coli strain BL21(DE3)pLysS were transformed with the recombinant protein expression vector pET-30 a (+) carrying the CDSs of studied IDSs. Transformed colonies were cultivated in LB containing chloramphenicol (50 µg/ml) and kanamycin (50 µg/ml). From liquid preculture, cultures of 250 ml were inoculated and set at OD_{600nm} = 0.4 before IPTG (1 mM) induction over night at 16 °C with 200 rpm shaking. Bacteria were then harvested by centrifugation at 3 000 g for 5 minutes and the pellets were conserved at -80°C for one night. Pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7, 500 mM NaCl, 2 mM DTT, 8 % glycerol v/v, 10 mM imidazole, 0.25 mg/ml lysozyme) and lysed by sonication for 3 min in ice (10 s sonication, 20 s stop, x 18 times). The resulting supernatant was mixed with Ni-NTA agarose resin (Qiagen) for 1 h. The resin was rinsed five times with washing buffer (50 mM Tris-HCl pH 7, 500 mM NaCl, 2 mM DTT, 8% v/v glycerol, 50 mM imidazole) and finally resuspended in the same buffer containing 250 mM imidazole. Proteins were desalted by passing through a PD10 desalting column (GE Healthcare) equilibrated with the assay buffer (25 mM MOPSO pH 7, 10 mM MgCl₂, 5% v/v

glycerol) and quantified with the Bradford method (Bradford, 1976). All steps of purification were conducted on ice. Purified proteins were aliquoted and stored at -20 °C. Biochemical characterization of each IDS in vitro was performed in 100 µl assay buffer (25 mM MOPSO pH 7, 10 mM MgCl₂, 5% v/v glycerol) with 100 ng to 5 µg of purified proteins and incubated from 5 to 20 minutes at 30°C with slowly shaking. Incubations were stopped by adding 100 µl of MeOH:H₂O (containing 10 mM NH₄OH) (7:3) and vortexed for 30 sec. Enzymatic Kinetic parameters were determined similarly by using 5 ng of RcG/FPPS1 and RcFPPS2 proteins (and the respective mutants) and 50 ng of RcGGPPS.LSU1, RcGGPPS.SSU, RcGGPPS.HOMO and co-purified RcGGPPS.LSU1 + RcGGPPS.SSU proteins. Enzymatic reactions were performed in assay buffer containing concentrations of IPP, DMAPP and GPP ranging from 0.5, to 50 µM for varying substrate and 60 µM for the stable substrate in 100 µl reaction volume at 30° C for 3 min. Incubations were stopped as described above. Kinetic parameters were calculated according to the Lineweaver-Burk plot model. Products were analysed by LC-MS/MS.

In planta characterization of IDSs

Plant expression constructs carrying the CDS of the studied IDSs were transformed into the hypervirulent *Agrobacterium* LBA4404. Transformed colonies were grown in LB media containing rifampicin (50 µg/ml), gentamicin (20 µg/ml) and kanamycin (50 µg/ml). After an overnight culture at 28°C with 200 rpm shaking, the bacteria were centrifuged 8 min at 4 000 g at room temperature and washed 3 time in resuspension buffer (MgCL₂ 10 mM, MES 10 mM, pH 5.7). 4-5 weeks old *N. benthamiana* were co-transformed by agro-infiltration with cultures of equal density of agrobacteria harboring the genes of interest in a ratio of 1/3 P19, 1/3 *NUDX1-1a* and 1/3 *IDS* (v/v), or 2/3 P19 and 1/3 *IDS* for single gene expression to a final OD_{600nm} = 1.2. 4 days post-infiltration the transformed *N. benthamiana* leaves were ground in liquid nitrogen and stored at -80 °C. For quantification of geraniol glycosides, 200 mg of tissues were resuspended in 400 µl of MeOH:H₂O (75:25) containing 10 µM of lavandulyl acetate used as internal standard. Samples were sonicated for 10 min at 40°C, vortexed for 30 s, centrifuged 5 min at 4 000 g to pellet cellular debries and supernatant was collected for LC-MS/MS analysis. For capsidiol quantification, 200 mg of tissues were resuspended in 400 µl of hexane containing 10 µM of camphor used as internal standard. Samples were sonicated for 10 min at 40°C, vortexed for 30 s, centrifuged 5 min at 4 000 g to pellet cellular debries and supernatant was collected for GC-MS analysis.

LC-MS/MS Analysis

All LC-MS/MS analysis were performed on an Agilent 1260 infinity II LC system coupled to an Agilent Ultivo triple quadrupole mass spectrometer (Agilent Technologies, Santa Clara).

For analysis of products from in vitro characterization of IDS, samples were run on a Poroshell 120 HPH-C18 column (50 mm × 2.1 mm, particle size 1.9 µm, Agilent) heated at 40 °C. The mobile phases consisted of 10 mM ammonium bicarbonate as solvent A, and acetonitrile as solvent B, both containing 0.15% v/v NH₄OH. Two microliters of reaction mixture were injected for each sample. Chromatography was run at 0.6 ml min flow rate and separation was achieved with a gradient starting with 2% B reaching 95% B in 2 min, followed by an isocratic phase at 98% B until 5 min and return at 2% B at 5.10 min with

equilibration until 10 min. Mass spectrometer tunings were as follow: capillary voltage 5500 V, gas temperature 350 °C, gas flow 12 l/min, and nebulizer 55 psi. Products detection was achieved in negative mode and MRM modes.

For the quantification of prenyl phosphates in OB flowers, 400 mg of petals were ground in liquid nitrogen and mixed with 1 ml of ice cold MeOH:H₂O (20 mM NH₄OH) 7:3 containing 10 µM of 1-naphthyl phosphate used as internal standard. Samples were then sonicated for 10 minutes at 4°C, vortexed 30s and centrifuged 5 min at 4 000 g. Supernatants were then reduced under with nitrogen flow on ice until 100 µl. Samples were analysed by LC-MS/MS as described above but on a Poroshell 120 HPH-C18 column (100 mm × 2.1 mm, particle size 1.9 µm, Agilent).

For geraniol glycosides quantification in *N. benthamiana* tissues samples were analyzed on a Poroshell 120 HPH-C18 column (100 mm × 2.1 mm, particle size 2.7 µm, Agilent) heated at 40 °C. The mobile phases consisted of water as solvent A, and acetonitrile as solvent B, both containing 0.1% v/v formic acid. One microliters of samples were injected and separation was achieved at a flow rate of 0.5 ml/min starting with an isocratic phase with 20% B for 0.5 min, followed by a gradient reaching 85% B at 10 min and then 98 % at 10.5 min, followed by isocratic phase with 98% B until 12 min and return at 20% B at 12.2 min with equilibration until 16 min. Mass spectrometer tunings were as follow: capillary voltage 6000 V, gas temperature 350 °C, gas flow 12 l/min, and nebulizer 55 psi. Products detection was achieved in positive mode and MRM modes.

MS/MS transitions and tunings for each compounds were as followed: 312.2 → 78.9 *m/z* for GPP with Fragmentor set at 70 V and Collision Energy set at 92 V, 381.3 → 78.8 *m/z* for FPP with Fragmentor set at 105 V and collision energy set at 124 V, 449.4 → 78.7 *m/z* for GGPP with Fragmentor set at 110 V and collision energy set at 148 V, 23 → 78.9 *m/z* for 1-Naphthyl phosphate with Fragmentor set at 64 V and Collision Energy set at 40 V, 137.1 → 81 *m/z* for geraniol glycosides with Fragmentor set at 80 V and Collision Energy set at 8 V, 197.2 → 137.1 *m/z* for lavandulyl acetate with Fragmentor set at 75 V, and collision energy set at 4 V.

Data analysis were performed with MassHunter quantitative software (Agilent Technologies). Compounds were quantified according to standard curves constructed with the use of authentic standards except for geraniol glycosides that were quantified relatively as previously described (Magnard *et al.* 2015; Sun *et al.* 2020).

General discussion |

I- Diversification of *NUDX1* in *Rosaceae* and in *Rosa*

An inventory of *NUDX1* gene copies and *Nudx1* clades was first done in *R. x wichurana*, *R. x hybrida* cv. 'Papa Meilland', and OB (Article 1). In this work, we found three clades, that we named *Nudx1-1* to *Nudx1-3* on the NJ tree (Figure 1.1). These clades were highly supported, with bootstraps of 100. By studying *NUDX1* copies of *F. vesca* and *P. micrantha*, we demonstrated that these clades were also present in the subfamily *Rosoideae* (Article 2). Indeed, the ML tree of *NUDX1* copies of *R. x wichurana*, *R. x hybrida* cv. 'Papa Meilland', and *F. vesca* also showed bootstraps of 100 (Figure 2.2). *NUDX1* copies of *P. micrantha* scaffolds could not be used for this tree because the online genome was a draft, but we found highly similar sequences by BLAST (Table S2.2). Thanks to the sequenced genome of OB and its assembly at the chromosome scale (Raymond *et al.* 2018; Hibrand Saint-Oyant *et al.* 2018), it was then possible to map *NUDX1* copies on chromosomes (Figure 2.3). Furthermore, it was also possible to map the alleles on each chromosome pair because three genomes were available online in the GDR (Jung *et al.* 2019). Indeed, because of the heterozygote status of the OB genome, reads were difficult to assemble and the corresponding online genome contains mismatches and errors (Raymond *et al.* 2018). As a consequence, authors did it also with haploid calluses and then obtained haplotype sequences of very good quality of assembly (Raymond *et al.* 2018; Hibrand Saint-Oyant *et al.* 2018). We also used the MINion (Oxford Nanopore) direct sequencing method of gDNA to obtain long reads without the assembly step (Figure 2.6). This technology is very useful to verify gene clusters and mapping. We used two diploid species: *R. moschata*, which emits geraniol, and *R. laevigata*, which does not. We found a very well conserved cluster on chromosome 4, which contains *NUDX1-1*, *NUDX1-2* and *NUDX1-3* copies. We also found dispersed copies of *NUDX1-2* on chromosomes 6 and 7. Finally, we also found a cluster of *NUDX1-1* copies on chromosome 2, but only in the rose species that produced geraniol.

In the same work, we were also able to compare the *NUDX1* copies of *Rosoideae* to the *NUDX1* copies of *Amygdaloideae*, the other major subfamily in the *Rosaceae* family, using the genomes of *P. persica* and *M. x domestica* published in the GDR (Jung *et al.* 2019). Surprisingly, all these sequences mapped in the same regions of OB's chromosome 4, but they have highly diverged. We named them *NUDX1-4* sequences (Figure 2.4). Evolution between duplicated genes is a common phenomenon, because gene duplication relaxes selection pressure and thus allows accumulation of mutations (Scannell and Wolfe, 2008). At the opposite, highly conserved sequences in *Nudx1-1*, *Nudx1-2* and *Nudx1-3* clades, indicate a high selection pressure, thus a putative conserved function. Obviously, due to the very low bootstrap of *Nudx1-4* clade, and due to some paraphyletic *NUDX1-4* sequences (Figure 2.2), this group should be studied in details in the future, functionally and phylogenetically, and revisited. Nevertheless, it seems that a first copy of *NUDX1* on chromosome 4 was at the origin of *NUDX1-4* sequences in *Amygdaloideae*, and, in parallel, at the origin of all the *NUDX1* sequences in *Rosoideae*. In the same way, the single copy of *NUDX1* in *A. thaliana* is located in the same microsyntenic region (unpublished result). In *Rosoideae*, the closest sequences, most at the base of the ML tree, belonged to the *Nudx1-*

3 clades (Figure 1.1; Figure 2.2). We could thus conclude that this clade contains ancestral sequences, with perhaps, the same function in *Rosoideae*.

More precisely, in *Rosoideae*, we found several copies of *NUDX1-1* and *NUDX1-2* genes, that we named *a* to *c* (Figure 2.3; Figure 2.6). *NUDX1-1a* was found on chromosome 2 forming a cluster of 4 genes (99 % of identity) and one pseudogene in OB, and two copies in *R. moschata*. In OB, the other chromosome 2 contained a null allele, *i.e.* not a single copy of *NUDX1* at this location. *NUDX1-1b* was found on chromosome 4 in all species of *Rosoideae* that we studied (Figure 2.3; Figure 2.4). *NUDX1-2a* was also found on chromosome 4 in all species of *Rosoideae*. *NUDX1-2b* and *c* were respectively found on chromosomes 6 and 7 (Table S1.8; Figure S1.4; Figure 2.3; Figure 2.4; Figure 2.6).

The *NUDX1-1a* cluster on chromosome 2 was not found in roses without geraniol and in other *Rosoideae* (Table S2.2). Additionally, we cloned as much as possible gDNAs in more than 30 species of wild roses. These gDNAs clearly formed two subclades: *Nudx1-1a* subclade including *NUDX1-1a* homologs, and *Nudx1-1b* subclade including *NUDX1-1b* homologs of OB (Figure 2.5; Figure S2.2). On the ML tree, the *Nudx1-1b* subclade was more ancestral than the *Nudx1-1a* subclade. It included gDNAs of all rose species. At the opposite, *Nudx1-1a* subclade included only gDNAs of roses that produced geraniol. We could thus conclude that *NUDX1-1b* gene was ancestral compare to *NUDX1-1a* gene, and that the geraniol production is a “recent trait” in wild roses.

Because we have found several copies of *NUDX1-1a* in OB and *R. moschata*, we tried to verify whether or not the copy number of this gene was variable in wild roses. For this purpose, we used 12 wild roses that we compared by qPCR to OB, knowing that OB had seven genes that could be amplified with our primers (five copies of *NUDX1-1a*, two copies of *NUDX1-1b*, two copies of *NUDX1-2a*, one copy of *NUDX1-2b*, and one copy of *NUDX1-2c*). The result was that the copy number was very variable from a wild rose to another. Furthermore, roses that had more copies than OB produced geraniol, and roses that had less, no geraniol. There was an exception concerning *R. hugonis* that we could not explain. Gene duplication is a common event, for example cytochrome P450 gene in *A. thaliana* have about 255 duplications, and constitute one of the major gene family with 76 sequences considered as tandem duplications (Cannon *et al.* 2004). Indeed, when several identical copies of sequences are present on a DNA strand, some errors can occur and create tandem or interspersed copies (Żmieńko *et al.* 2014; Lye and Purugganan, 2019). For example, nonallelic homologous recombination between two identical sequences may create an unequal crossing-over. Microhomology-mediated break-induced replication may also occur during replication. In *M. x domestica*, clusters of *O-METHYLTRANSFERASE* genes were interpreted as duplications due to hairpins structures of MITEs that had provoked DNA slippage during replication (Han and Korban, 2007). This mechanism could be at the origin of the duplication on chromosome 2 because we found MITEs between *NUDX1-1a* copies (Figure 2.6; Figure S2.4). The sequence repetition on chromosome 2 looked like the sequence repetition of *MATE1* in *Zea mays* (Maron *et al.* 2013). Indeed, in maize, there are three interspersed copies of *MATE1* with *gypsy*, *copía* and *mutator* TEs repeated in the intergenic regions. Authors demonstrated that the different inbred lines

of maize have different numbers of copies. This is referred to as Copy Number Variations (CNVs), *i.e.* variation of number of gene copies between individuals (Lye and Purugganan, 2019), or between inbred lines (Maron *et al.* 2013). They argue that these different lines correspond to different adaptations to different soils, because MATE1 is involved in aluminium tolerance. In another work, DeBolt (2010) demonstrated that several CNVs appeared in only five generations in *A. thaliana* when a selection pressure was applied, making CNVs a very strong mechanism of adaptation. Such link between CNVs and adaptation is also documented in *Homo sapiens* evolution (Hsieh *et al.* 2019). We cannot be sure that the repetition of *NUDX1-1a* corresponded to CNVs in roses because of two arguments. Firstly, there are three individuals of OB sequenced in the GDR, and they have the same copy number, thus no CNV. Secondly, cultivated roses are clones, propagated by cuttings and grafting in botanical gardens, even for wild roses in collection. Nevertheless, the differences of copy number between OB, *R. moschata*, *R. laevigata* and other wild species of *Rosa* could well correspond to ancestral CNVs, that could correspond to adaptations of different individuals of a common ancestor. Evidences for the role of CNVs in functional divergence contributing to adaptations *via* secondary metabolism evolution at the inter-species level had already been exemplified (Shirai and Hanada, 2019). It could therefore be interesting to check for such *NUDX1-1a* CNVs in wild populations of roses and in different areas to gain knowledge on adaptation due to terpenols. Furthermore, examples of CNVs are numerous in domesticated plants (Lye and Purugganan, 2019), but not in wild species, as only one is documented in *Picea spp.* (Prunier *et al.* 2017). CNVs could thus be a source of speciation in the wild.

As we concluded before, *NUDX1-1b* was more ancestral than *NUDX1-1a*. We could then suppose that *NUDX1-1b* was duplicated in some wild roses from chromosome 4 to chromosome 2. Around these genes, we found several identical transposons: MITE *P580.2030* and *Copia R24588* (Fig. 6 in Article 2). However, these transposons were in different orders. This could indicate a kind of shuffling of the different DNA fragment during duplication, which is a strong clue for transposon-mediated *trans*-duplication, like in Pack-MULE (Packed Mutator-like Element) transposition for example.

Most of the Pack-MULEs are less than 5 kb in size and flanked by opposite Tandem Inverted Repeats (TIRs) greater than 100 pb, which is the case in OB (Figure S2.4; TIRs are the two opposite sequences of a MITE). Pack-MULEs are composed of mixed fragments of different genes or sequences and often include introns, have a gene microsynteny not shared between species, exhibit expression, which does not resemble those of the parental sequences, and is often higher and tissue-specific (Ferguson *et al.* 2013; Cerbin and Jiang, 2018; Zhao *et al.* 2016, 2018). All these criteria are compatible with the hypothesis of a Pack-MULE in *Rosa*. Furthermore, on chromosome 2 only, MITE *P580.2030* was interrupted by MITE *G13554*, which could be interpreted as an event more recent than the *trans*-duplication of the Pack-MULE.

II- Specialization of NUDX1 in terpene production

The Nudix hydrolase proteins were first described as enzymes that clean cells from potentially deleterious metabolites, and modulate the concentration of intermediates in some biochemical pathways (Bessman *et al.* 1996). They are found in all living organisms and consist of diphosphohydrolases of nucleoside diphosphates linked to some moiety *X* (Nudix). Their substrates are diverse including oxidized dNTPs, nucleotide sugars and alcohols, or capped RNAs (review in Mildvan *et al.* 2005; McLennan, 2006; Kraszewska, 2008; Srouji *et al.* 2017). One of these Nudix hydrolases has been studied since a long time. It was named MutT (Bessman *et al.* 1996) or NudA (McLennan, 2006) in *Escherichia coli*, 8-oxo-dGTPase (Sakumi *et al.* 1993) or MTH1 (Gad *et al.* 2014) in *Homo sapiens*, AtNUDT1 (Dobrzanska *et al.* 2002) or AtNUDX1 (Yoshimura *et al.* 2007) in *A. thaliana*. Several substrates were proposed for this NUDX1 enzyme, but one of them comes up more frequently: 8-oxo-dGTP. This compound could cause mutations if incorporated into DNA sequences. NUDX1 removes one phosphate creating 8-oxo-dGMP, which cannot be integrated into DNA. However, this enzyme activity has been recently revisited in plants.

In garden roses, it has been shown that RhNUDX1 was involved in geraniol production (Magnard *et al.* 2015). Authors used an AFLP-differential display method on two cultivars with a close genetic background: the highly scented *R. x hybrida* cv. 'Papa Meilland' and the unscented *R. x hybrida* 'Rouge Meilland'. They found that *RhNUDX1* was the most expressed gene in the scented cultivar. It was also expressed during blooming in all tested cultivars producing geraniol. Using reverse genetic approach, *in vitro* enzymatic activity assays, and analysis of the progeny from a cross between OB that produces geraniol and *R. x wichurana* that does not, they demonstrated that the *RhNUDX1* expression was strongly linked to geraniol production. They also showed that GPP was dephosphorylated by RhNUDX1 producing GP, that was then dephosphorylated by an uncharacterized phosphatase to geraniol. In contrast, in other plants, GPP was involved in geraniol production by another terpene synthase dependent pathway. Indeed, *O. basilicum* (Iijima *et al.* 2004), *Cinnamomum tenuipilum* (Yang *et al.* 2005), *C. roseus* (Simkin *et al.* 2013), *Citrus sinensis* (Li *et al.* 2017), *C. plicata* (Xu *et al.* 2017), and *Pelargonium x hybridum* (Blerot *et al.* 2018) possess a geraniol synthase, which directly hydrolyzes GPP to geraniol. This enzyme belongs to TPSs family. While geraniol synthases are localized in plastids, RhNUDX1 was shown to reside in the cytosol, raising the question about the origin of the cytosolic GPP in roses. In parallel, Henry *et al.* (2015; 2018) demonstrated that AtNUDX1 was involved in the regulation of IPP concentration in *A. thaliana* leaves. Using *in vitro* enzymatic assays and heterologous transgenic experiments, they showed that AtNUDX1 dephosphorylated IPP and DMAPP into IP and DMAP, and another enzyme, isopentenyl phosphate kinase, phosphorylated IP to IPP, and DMAP to DMAPP. As a consequence, the regulation of IPP and DMAPP concentrations by these two enzymes could modulate the metabolic flux of terpene metabolism. It has been shown that AtNUDX1 can hydrolyze GPP (precursor of monoterpenes), and FPP (precursor of sesquiterpenes) with equal catalytic efficiencies. Since GPP is generally produced into plastids but AtNUDX1 was localized in the cytosol, this likely explains the absence of geraniol production *in planta*. On the other hand, production of some

sesquiterpenes via TPS independent pathway was not excluded, as both FPP and AtNUDX1 are present in the cytosol.

Substrate affinities of NUDX1 for IPP, GPP, FPP and 8-oxo-dGTP were discussed in several papers (Henry *et al.* 2018; Liu *et al.* 2018; Jemth *et al.* 2019; Article 1) raising the question of the diversification and/or specialization of this enzyme in the plant kingdom. Indeed, we demonstrated that in the flowers of a hybrid of *R. wichurana*, RwnUDX1-2c hydrolyzed specifically FPP to farnesyl monophosphate, which in turn gives rise to (*E,E*)-farnesol (Article 1). Another NUDX1 enzyme was recently characterized in *T. cinerariifolium* (Li *et al.* 2020). The authors showed that although *TcNudix1* gene was expressed specifically in ovarian trichomes of the flower, its expression could be also induced in leaves by methyl jasmonate treatment. *TcNudix1* protein contains a plastid transit peptide at the *N*-terminus, that was confirmed by transient expression with a reporter gene. Analysis of enzyme activity revealed that *TcNudix1* was capable of removing one phosphate from chrysanthemyl diphosphate producing chrysanthemyl monophosphate. Then, formation of chrysanthemol from chrysanthemyl monophosphate probably needed an endogenous phosphatase. Chrysanthemol is the precursor of six pyrethrins that protect ovaria and wounded leaves against phytophagous insects. Another example of evolution of the *NUDX1* gene family seemed to have occurred in *Pelargonium graveolens*. *PgNdx1* gene encodes the cytosolic protein with 65 % amino acid similarity with AtNUDX1, and 63 % with RhNUDX1, which can *in vitro* hydrolyze GPP into GP with high catalytic efficiencies (Bergman *et al.* 2021). However, no correlation was found between *PgNdx1* expression and geraniol levels in different *Pelargonium* cultivars. In contrast, some authors have isolated and characterized a geraniol synthase from *P. x hybridum*, which suggests that perhaps both pathways exist in this genus (Blerot *et al.* 2018). This could be similar to a situation in *R. x damascena*, where Önder *et al.* (2021) have cloned a putative geraniol synthase. In *Pelargonium*, when GPP, GP and geraniol concentrations were analyzed in leaf extracts of *Pelargonium* cultivars, strong correlation between these compounds was found (Bergman *et al.* 2021). Moreover, in a cultivar that produces (-)- β -citronellol, but not geraniol, GPP, GP and citronellyl monophosphate (CP), but not citronellyl diphosphate were detected. The authors discussed the possibility that GP was transformed into CP, and then into (-)- β -citronellol, and proposed to search for a double bond reductase that could convert GP to CP.

In summary, NUDX1 enzymes can therefore hydrolyze different diphosphates in the terpene pathways. They may have different activities and catalytic efficiencies, but their cellular compartmentation is also highly important and could give rise to specific metabolites. These specificities are based on different evolutionary histories from one species to another and from one family to another. In our work, we demonstrated that the CNV of *NUDX1-1a* corresponded to a gene dosage expression (Fig. 7 in Article 2), and that the promoter of *NUDX1-1a* contained a regulatory element, that we named *box38*, of the LTR region of *Copia R24588* which was sufficient for petal expression (Figs. 9, S6 and S7 in Article 2).

Concerning the gene dosage effect, we demonstrated that the number of copies of *NUDX1-1a* in botanical roses was responsible for its expression in a nonlinear model, and that the expression of *NUDX1-1a* was correlated with the concentration of geraniol and other compounds (Figure 2.7; Figure S2.3). For example, Perry *et al.* (2007) find that the copy number of *AMY1*, a gene encoding salivary

amylase in *Homo sapiens*, is higher in populations with high-starch diets than in populations with low-starch diets. Nevertheless, the correlation of the copy number and the concentration of amylase in saliva have an r^2 of 0.351 meaning that the relation is not strictly linear and depends of other factors. Models of gene dosage effect predicts that the copies could quickly diverged if the selection pressure is not very high. At the opposite, the copies can be fixed by adaptative natural selection if the number of copies increases fitness (Hahn, 2009). Thus, as *RcNUDX1-1a* copies had an identity varying only from 97.1 to 97.6 %, we can hypothesize that this gene, and thus acyclic terpene concentration, was very important in the adaptation and the evolution of *Rosa* species.

Concerning the *box38*, element, we demonstrated that it was necessary and sufficient to drive previously discovered petal-specific *NUDX1-1a* expression in petals of fully opened flowers (Figure 2.9). We also observed that it was differently repeated from a rose species to another, but always at 138 bp from the start codon (Figure S2.7). This *box38* is a part of the LTR of *Copia R24588* (Figure S2.6). LTR acts as a selfish transcription of the *Copia* elements and it is known that such sequences contain regulatory elements that can modify gene expression in *cis* and can contribute to neofunctionalization (Kobayashi *et al.* 2004; Grandbastien 2015; Galindo-González *et al.* 2017). Furthermore, *box38* location in the 5'-upstream region of the pseudogenes *^vNUDX1-2a* (Figure 2.6) suggests that they have been expressed at the origin. However, during transposition of the MULE from chromosome 4 to chromosome 2, the *box38* repeats were shuffled and placed in front of *NUDX1-1a* making its expression petal-specific. There is increasing evidence that TEs are a source of diversification of species and modification of gene expression, particularly in the *Rosaceae* (Gu *et al.* 2016, Wang *et al.* 2016, Daccord *et al.* 2017, Jiang *et al.* 2019). Indeed, several TE insertions have been described in *Rosaceae*, which modified transcription levels as a result of new binding sites for transcription factors or disruption of existing ones, new methylation/acetylation pattern, or hairpin structure formation (Han and Korban, 2007; Wang *et al.* 2009; Iwata *et al.* 2012; Gu *et al.* 2016; Hibrand Saint-Oyant *et al.* 2018; Morata *et al.* 2018; Ono *et al.* 2018; Zhang *et al.* 2019). Further investigations are thus necessary to understand the underlying mechanisms of petal-specific expression.

Whatever the action mechanisms of *box38*, petal-expression of *NUDX1-1a* was positively correlated with the geraniol production of rose flowers (Figure 1.2; Figure S2.3). Logically, it was also correlated with the compounds putatively derived from geraniol. Such pathway has not been studied in roses but in some other species. This the case for nerol, β -citronellol, neral, geraniol, and neryl acetate (Iijima *et al.* 2004; 2006; 2014; Ito & Honda, 2007; Yuan *et al.* 2011; Sato-Masumoto & Ito, 2014; Xu *et al.* 2017). (*E*)- β -ocimene and β -myrcene has never been involved in a geraniol pathway, but they also belong to acyclic monoterpenes. However, we also demonstrated that this correlation was not restricted to geraniol but also to some cyclic monoterpenes (limonene, α -phellandrene), acyclic sesquiterpenes ((*E,E*)-farnesol, (*E,E*)-farnesal, (*E,E*)- α -farnesene, (*Z,E*)- α -farnesene, (*E*)- β -farnesene), aliphatics ((*E*)-2-hexenol, (*E,Z*)-nonadienal), benzenoids (benzylalcohol), and phenylpropanoids (2-phenylethanol). It was also negatively correlated with DMT, 2-pentadecanone and 2-tridecanone. We could explain these correlations with three hypotheses.

First, it is possible that the activity of *NUDX1-1a* depends on a transcription factor. If it is up-regulated, *NUDX1-1a* activity would be up-regulated and geraniol concentration would increase if GPP is available. If other genes were regulated by this transcription factor, they were also overexpressed and some other compounds were produced with geraniol. They may up-regulate some genes but they also may down-regulate others in different pathways (Dudareva *et al.* 2013). Such pleiotropic effects with transcription factors that regulate the expression of scent genes are known. For example, Ben Zvi *et al.* (2012) overexpressed *PAP1* in *R. x hybrida* cv. 'Pariser Charme'. *PAP1* is a Myb transcription factor involved in production of anthocyanins in *A. thaliana*. They observed an increase of the transcription of several genes involved in color and scent, among which *AAT1* and *PAAS*. *AAT1* is a geraniol and citronellol acetyl transferase which can synthesize geranyl acetate, benzyl acetate, (*Z*)-3-hexenyl acetate and β -citronellyl acetate *in vitro*, *in planta*, or in transgenic plants with different efficiencies (Shalit *et al.* 2003; Guterman *et al.* 2006). *PAAS* is a phenylacetaldehyde synthase which synthesizes 2-phenylacetaldehyde leading to 2-phenylethanol (Kaminaga *et al.* 2006). In our work, geranyl acetate, neryl acetate and 2-phenylethanol were positively correlated to *NUDX1-1a* activity. Thus, we could imagine that an up-regulation of the transcription factors could lead to a modulation of expression of several biosynthetic gene.

In the second hypothesis, the volatile compounds could be correlated to *NUDX1-1a* activity only because they have been co-selected with geraniol for their roles in chemical communication with other species. In a meta-analysis work, (Junker and Blüthgen, 2010) demonstrated that flower scents have a dual function by attracting obligate flower visitors but repelling facultative ones. Considering attraction, 2-phenylethanol is one of the most active compounds known for flower-visiting insects, and particularly for bees (Raguso, 2004). Furthermore, some compounds are considered as pheromones in bees. It is the case for (*E*)- β -ocimene, geraniol, nerol, geranial, neral, nerolic acid, geranic acid, and (*E,E*)-farnesol (Trhlin and Rajchard, 2011). In our work, seven of these nine compounds were positively correlated with *NUDX1-1a* activity in roses. An intriguing result was the negative correlation between 2-tridecanone and 2-pentadecanone, and *NUDX1-1a* activity. Nevertheless, in *R. rugosa*, it has been demonstrated that 2-tridecanone is strongly repellent for foraging bumble bees, and is present in high amounts in pollen (Dobson *et al.* 1999). One could imagine that selection by pollinators decreased the concentration of such compound in petals to increase landing, but increased it in pollen to decrease foraging. Furthermore, volatile compounds can also repel enemies (Junker and Blüthgen, 2010). Thus, for example, the presence of (*E*)- β -farnesene and 2-phenylethanol could also be involved in defense as it is the case in several species (Gibson and Pickett, 1983; Schnee *et al.* 2002; Günther *et al.* 2019). In summary, the selection pressure could modulate several gene expressions because several compounds are active both for attraction and defense. As a consequence, the indirect correlation of *NUDX1-1a* with all these products could just correspond to a diffuse selection pressure on several genes by insects.

Concerning the third hypothesis, some compounds could be by-products of *NUDX1-1a* activity, and/or could be produced by the same pathway. This is the case for putative acyclic monoterpenes as we explained before, but it could also concern other compounds, and particularly acyclic sesquiterpenes.

In *Rosa*, RhNUDX1 can use FPP *in vitro* (Magnard *et al.* 2015), and this is also the case for, AtNUDX1 in *A. thaliana* (Henry *et al.* 2018). In our work we demonstrated that in a hybrid of *R. wichurana*, RwnNUDX1-2c used FPP to produce (*E,E*)-farnesol and derivatives *in planta* (Figure 1.2; Table 1.2). We also demonstrated that GPP and FPP were produced in the cytosol by the MVA pathway, with FPPS1 and FPPS2 activities (Article 3). Furthermore, it has been demonstrated in *Zea mays* that (*E*)- β -farnesene is a by-product of the (*E,E*)-farnesol synthase activity (Schnee *et al.* 2002). We could thus imagine that NUDX1-1a could use both GPP and some FPP *in planta*, and thus also produced acyclic sesquiterpenes.

III- An evolutive scenario of geraniol production in *Rosa*

Our results suggest that geraniol production in rose petals was due to two distinct evolutionary events during evolution of *Rosids*. The first event was the evolution in *Rosaceae* of a classical FPPS producing FPP into FPPS also producing GPP through point mutations of at least of two amino acids (F88 and V123) affecting the affinity of the protein for GPP binding. These mutations have led to a bi-functional enzyme, G/FPPS1, that could produce both GPP and FPP in the cytosol, the other enzyme, FPPS2, producing only FPP (Article 3). This specialization of G/FPPS1 occurred before the diversification of *Rosaceae*. Still before this diversification, *NUDX1* gene was *cis*-duplicated several times on chromosome 4 (Article 2). These copies then diverged during the evolution of *Amygdaloideae* subfamily, giving a heterogeneous NUDX1-4 group with unknown functions. In the *Rosideae* subfamily, the three ancestral copies were *trans*-duplicated and then gave three clades: Nudx1-1, Nudx1-2 and Nudx1-3. It is likely that Nudx1-3 clade represents the ancestral *NUDX1*. After these events, during the evolution of *Rosa*, *NUDX1-1a* copy specialized into the production of geraniol. It is possible that the *trans*-duplication on chromosome 2 was due to a Pack-MULE as discussed before. Indeed, this type of transposition could explain the shuffling of *NUDX1-1a* and LTR fragments of *Copia R24588* giving rise to the *box38* and thus the specific expression of *NUDX1-1a* in petals. *NUDX1-1a* was thus expressed in petals, and the encoding enzyme could metabolize GPP produced by G/FPPS1 to produce geraniol. However it is not known when during evolution G/FPPS1 acquired its specific expression in petal with a rhythmic pattern. This geraniol production has been selected in many *Rosa* species. These speciation events could have come from CNV in individuals. Furthermore, in some species, at least in a hybrid of *R. wichurana*, NUDX1-2c could have specialized into production of (*E,E*)-farnesol (Article 1).

The production of geraniol seems to have been selected in *Rosa*. We could assume that it is thanks to the action of pollinators, as the role of VOCs are very important in numerous flowers (review in Muhlemann *et al.* 2014). Indeed, geraniol, (*E,E*)-farnesol, and their derivatives are known to be involved in bee attraction. For example, geraniol is detected by honey bees, as it was proved by the signals obtained with electroantennogram assays (Larson *et al.* 2020). It is also known that the Nasonov secretion in bees is made of geraniol, nerol, geranial, neral, nerolic acid, geranic acid, and (*E,E*)-farnesol (Pickett *et al.* 1980). The Nasonov secretion is considered as an orientation pheromone for bees (Trhlin and Rajchard, 2011; Bortolotti and Costa, 2014). Thus, geraniol and/or (*E,E*)-farnesol could have been

involved in the selection pressure that conserved *Nudx1-1a* cluster, and perhaps *NUDX1-2c* in the hybrid of *R. wichurana*.

We thus provided evidences that cytosolic GPPS function was present before acquiring *NUDX1-1a* expression in *Rosaceae* petals. It means that the cytosolic production of GPP was selected before (or in parallel) the specialization of *Nudx1-1* and *Nudx1-2* clades. This is in agreement with the observations in other *Rosoideae*. Indeed, Aharoni *et al.* (2004) cloned a cytosolic linalool/nerolidol synthase which produced linalool (monoterpene) and nerolidol (sesquiterpene) in *Fragaria x ananassa* fruits. They also cloned a cytosolic pinene synthase which produced α -pinene, α -myrcene and other monoterpenes in *F. vesca* fruits. They concluded that there was a source of cytosolic GPP in strawberries. In another work, Hampel *et al.* (2007) observed a strong labelling of monoterpenes, such as linalool and α -pinene, when fruits of *Rubus idaeus*, were fed with labelled mevalonate, compared to labelled DOX. They concluded that monoterpenes could be synthesized either in the cytosol of raspberry fruits or using MVA-dependent precursors transported to plastids. Here we show in Roses that both scenarios could apply since GPP is formed in the cytosol where it is used by *NUDX1.1a* for geraniol production and is transported to the plastids where it is used to produce ionols. It could be interesting to generalize these observations to other *Rosoideae* and also to *Amygdaloideae*, in flowers and fruits. It is possible that *NUDX1-1b*, *NUDX1-2 b* and *c*, and *NUDX1-4* were involved in terpene production in such subfamilies.

Moreover, cytosolic GPP is not restricted to monoterpenes biosynthesis, as GPP can be involved in the prenylation of some compounds by prenyl transferases. Geranylation of compounds leads to numerous bioactive molecules, which are considered as an evolutionary advantage (Epifano *et al.* 2008). For example, in *L. erythrorhizon*, geranylated coumarins are produced in roots. The p-hydroxybenzoic acid is prenylated by GPP and allow production of shikonin (Ueoka *et al.* 2020). In *Canabis sativa*, cannabidiol and tetrahydrocannabinol are prenylated bioactive compounds (see review of prenylated compounds in Bruijn *et al.* 2020). Thus, one could imagine that the production of cytosolic GPP could also be the substrate of prenyl transferases for geranylation of yet unknown metabolites in *Rosaceae*.

Conservation of 2 FPPSs in *Rosaceae* genomes (Figure 3.5), one of which produces cytosolic GPP in addition to FPP, raised the question about the evolutive advantage brought by this conservation and also the specialisation of IDSs for specific biological processes in specific organs. Indeed, it is likely that FPPS2 maintains the essential functions of FPP biosynthesis for production of vitals primary metabolites in cells such as sterols while RcG/FPPS1 specialized into production of GPP and FPP for VOCs production, at least in flowers. However, having conserved the FPPS activity of RcG/FPPS1 (unlike LeGPPS in *L. erythrorhizon*) could also benefit to plant fitness. First, RcG/FPPS1 bifunctional activity is like to kill two birds with one as it can produce both precursors leading to sesquiterpenes, monoterpenes and ionones (though crosstalk) biosynthesis in petals while FPPS2 function could be focused on primary metabolism. In addition, RcG/FPPS1 could complement and/or support FPPS2 essential functions as it was observed in *Arabidopsis thaliana* where both FPPS1 and FPPS2 can complement the loss of the other, respectively, since only the double mutant is lethal (Closa *et al.* 2010). Subcellular co-localization

of cytosolic FPPS in rose with one producing GPP raise the question about the existence of potential FPP and GPP biosynthesis channelling as RcG/FPPS1 produced GPP could be consumed as substrate by RcFPPS2. Furthermore, this subcellular co-localization of two FPPS could result in the putative formation of heterodimer between G/FPPS1 and FPPS2 but has not been explored yet. Indeed, the mutation of ancestral Y88 (and potentially I123V) from *Rosids* into F88 in *Rosaceae* results in the loss of the Y88-Y88' hydrogen bond thus affecting the interaction between both RcG/FPPS1 subunits and ultimately RcG/FPPS1 product specificity. Thus, the interaction of both rose FPPS could affect their activity and thus product specificity. Studies investigating this question have been performed with the *Saccharomyces cerevisiae* FPPS (ERG20) to increase GPP production. This was achieved by suppressing one allele of ERG20 WT in a diploid strain to optimize improved ERG20 homodimers of mutant protein and thus GPP and monoterpene production (Ignea *et al.* 2014).

Because RcG/FPPS1 product specificity can be modified depending on the ratio of IPP and DMAPP substrates provided *in vitro* (Figure 3.3) it is probable that endogenous substrate availability is critical for RcG/FPPS1 activity *in planta*. *In vitro*, the excess of DMAPP leads to higher formation of GPP by RcG/FPPS1 and inversely, excess of IPP leads to higher FPP formation. Thus, the concentration and IPP/DMAPP ratio in rose cells, which could depend on flower stage or rhythmicity probably drives the product specificity of RcG/FPPS1 towards GPP or FPP production and the subsequent downstream products. In tobacco, the production of IPP/DMAPP ratio from the MEP pathway stands in between 5:1 and 7:1 (Tritsch *et al.* 2010) which corresponds approximately to the ratio produced by the HDR enzyme (Rohdich *et al.* 2002). The IPP/DMAPP ratio produced by the MVA pathway has received less attention. The last enzymatic step of this pathway produces only IPP which then is converted into DMAPP by the action of IDI. An example of potential regulation of terpenes production by IDI was reported in *Artemisia annua*, which accumulates more than 20 monoterpenes in glandular trichomes (Polichuk *et al.* 2010; Ruan *et al.* 2016). Its plastidic *AaIDI-1* is 5-fold more expressed in trichomes than in roots and its encoded enzyme produces a 7:1 ratio of DMAPP/IPP *in vitro* (Ma *et al.* 2017). Increasing DMAPP quantity at the expense of IPP would mathematically favour GPP production instead of GGPP in plastid. Further investigations on the IPP/DMAPP ratio in rose flowers and on the involvement of IDI and its developmental and rhythmic expression pattern in petals will help to better understand the regulation of GPP and FPP derived VOC biosynthesis. In addition, the challenging quantification of cellular and subcellular IPP/DMAPP ratio would greatly help to better understand how the production of GPP, FPP or GGPP derived from both MVA and MEP pathways are regulated.

As a conclusion, we demonstrated that cytosolic GPP production appeared in the *Rosaceae* before the functionalization of the Nudx1 clade, and before the production of GP (thus geraniol) by NUDX1-1a. We have described the major stages of this evolution in *Rosoideae* and *Rosa*. We have also demonstrated the importance of TEs in this scenario, both for gene duplication, and for gene neofunctionalization. In a structural point of view, we explained the effect of some mutations on the enzyme activity.

Appendix 1 |

Supplementary materials for chapter 1

Table S1.1 | Sequences in RNAseq transcriptome with homologies to *NUDX1* sequences.

| Sequence name | Sequence length (bp) | Corresponding <i>NUDX1</i> sequence | Expression in the cultivars (FPKM) | | | |
|---------------|----------------------|-------------------------------------|------------------------------------|--------|--------|--------|
| | | | Rw | OW9035 | OB | OW9047 |
| c34287_g1_i1 | 859 | <i>NUDX1-1</i> | 27.8 | 31.4 | 9347.2 | 6687.3 |
| c58359_g1_i1 | 274 | <i>NUDX1-3</i> (partial) | 0.0 | 0.2 | 1.5 | 0.0 |
| c72373_g1_i1 | 164 | <i>NUDX1-3</i> (partial) | 0.7 | 0.3 | 3.3 | 0.0 |
| c92297_g1_i1 | 1294 | <i>NUDX1-2</i> | 1433.0 | 1332.6 | 1.9 | 2.8 |

Table S1.2 | List of PCR primers.

| Primer name | Primer sequence (5'-3') | Purpose |
|-------------------|--|---|
| M-NUDX F1 | ATGGGAAACGAGACAGTAGTAGT | Forward primer, for <i>NUDX1-1</i> amplification |
| M-NUDX R1 | TCATGTTGAAAAGGGTTAAATCC | Reverse primer, for <i>NUDX1-1</i> amplification |
| RwNUDX 32F | CACCATGGTCAACGAGACGGTG | Forward primer, for amplification, protein expression and heterologous expression of <i>NUDX1-2</i> |
| NUDX OB2 R3 | GATTA AAAAGGGAAAGGATTAATCC AGCC | Reverse primer, for amplification, protein expression and heterologous expression of <i>NUDX1-2</i> |
| RcNUDX1.3 F | ATGGAAAACGGCGCGT | Forward primer, for <i>NUDX1-3</i> amplification |
| RcNUDX1.3 R | TTACTGCCCTGAATTGGAATC | Reverse primer, for <i>NUDX1-3</i> amplification |
| NUDX qPCR F | AAGCCAAACCATCGCAGTAC | Forward primer, for qPCR amplification on <i>NUDX1-1</i> |
| NUDX qPCR R | GGAAGATTGTCCCACTCATACC | Reverse primer, for qPCR amplification on <i>NUDX1-1</i> |
| RwNUDX qPCR F | GAGCTTTGAGGAGTGTGCAAC | Forward primer, for qPCR amplification on <i>NUDX1-2</i> |
| RwNUDX qPCR R | GCTGTTGATGATCTGCCAAGG | Reverse primer, for qPCR amplification on <i>NUDX1-2</i> |
| RcNUDX1.3 qPCR F2 | GATCTTTATGCGAGCGGTTTC | Forward primer, for qPCR amplification on <i>NUDX1-3</i> |
| RcNUDX1.3 qPCR R2 | GTCGTCCCACTCGTACCAAC | Reverse primer, for qPCR amplification on <i>NUDX1-3</i> |
| Rc TCTP F | TTGGTCTTTGCCTACTACAAAGAGG | Forward primer, for qPCR control, housekeeping gene <i>TCTP</i> |
| Rc TCTP R | AAGCCAGTTGCTACTTCTTAGCACT | Reverse primer, for qPCR control, housekeeping gene <i>TCTP</i> |
| Rc Tub F | ATTGAGCGTCCCACCTACAC | Forward primer, for qPCR control, housekeeping gene <i>Tubulin</i> |
| Rc Tub R | AGCATGAAATGGATCCTTGG | Reverse primer, for qPCR control, housekeeping gene <i>Tubulin</i> |
| Rc EF1 F | GGTAAGGACCTTCACATC | Forward primer, for qPCR control, housekeeping gene <i>EF1-alpha</i> |
| Rc EF1 R | CAGCCTCCTTCTCAAACCTCT | Reverse primer, for qPCR control, housekeeping gene <i>EF1-alpha</i> |
| NUDX OB1 F | CACCTTGGTTCCGCGTGGATCCATGGG AAACGAGACAGTAGTAG | Forward primer, for protein expression of <i>NUDX1-1</i> |
| NUDX OB1 R | TCATGTTGAAAAGGATTAATCC | Reverse primer, for protein expression of <i>NUDX1-1</i> |
| RwNUDX F | CACCTTGGTTCCGCGTGGATCCATGGT CAACGAGACGGTGG | Forward primer, for protein expression of <i>NUDX1-2</i> |
| NUDX AT1 F | CACCTTGGTTCCGCGTGGATCCATGTC GACAGGAGAAGCGATA | Forward primer, for protein expression of <i>AtNUDX1</i> |
| NUDX AT1 R | TTAGTCTCCACCACCATGAG | Reverse primer, for protein expression of <i>AtNUDX1</i> |
| RhNUDX1-2_F1 | GATATAAAACATGGTCAACGAG | Forward primer for genetic marker for <i>NUDX1-2</i> |
| RhNUDX1-2_R1 | GGAAAGGATTAATCCAGCC | Forward primer for genetic marker for <i>NUDX1-2</i> |

Table S1.3 | Identity percentages of amino acid sequences among 13 rose NUDX1 proteins and one *A. thaliana* protein.

| Protein | RhNUDX 1 | ReNUDX 1-1a | ReNUDX 1-1b | RwNUDX 1-1 | ReNUDX 1-2a | ReNUDX 1-2b | ReNUDX 1-2c | RwNUDX 1-2a | RwNUDX 1-2b | RwNUDX 1-2c | RwNUDX 1-2c' | ReNUDX 1-3 | RwNUDX 1-3 | AtnUDX1 |
|-------------|-------------|----------------|----------------|---------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|---------------|---------------|---------|
| RhNUDX1 | - | 98.0 | 98.0 | 98.0 | 70.9 | 76.2 | 76.2 | 72.2 | 74.2 | 75.5 | 76.2 | 70.3 | 70.3 | 64.3 |
| ReNUDX1-1a | 98.0 | - | 97.3 | 97.3 | 71.5 | 76.8 | 76.2 | 72.8 | 74.8 | 75.5 | 76.2 | 71.7 | 71.7 | 64.3 |
| ReNUDX1-b | 98.0 | 97.3 | - | 100.0 | 72.8 | 78.1 | 78.1 | 74.2 | 76.2 | 77.5 | 78.1 | 72.4 | 72.4 | 66.4 |
| RwNUDX1-1 | 98.0 | 97.3 | 100.0 | - | 72.8 | 78.1 | 78.1 | 74.2 | 76.2 | 77.5 | 78.1 | 72.4 | 72.4 | 66.4 |
| ReNUDX1-2a | 70.9 | 71.5 | 72.8 | 72.8 | - | 90.0 | 86.0 | 98.7 | 94.0 | 85.3 | 86.0 | 66.9 | 66.9 | 60.8 |
| ReNUDX1-2b | 76.2 | 76.8 | 78.1 | 78.1 | 90.0 | - | 92.0 | 90.0 | 92.0 | 91.3 | 92.0 | 71.7 | 71.7 | 62.2 |
| ReNUDX1-2c | 76.2 | 76.2 | 78.1 | 78.1 | 86.0 | 92.0 | - | 87.3 | 92.0 | 97.3 | 98.0 | 69.0 | 69.0 | 62.2 |
| RwNUDX1-2a | 72.2 | 72.8 | 74.2 | 74.2 | 98.7 | 90.0 | 87.3 | - | 95.3 | 86.7 | 87.3 | 67.6 | 67.6 | 60.8 |
| RwNUDX1-2b | 74.2 | 74.8 | 76.2 | 76.2 | 94.0 | 92.0 | 92.0 | 95.3 | - | 91.3 | 92.0 | 69.7 | 69.7 | 63.6 |
| RwNUDX1-2c | 75.5 | 75.5 | 77.5 | 77.5 | 85.3 | 91.3 | 97.3 | 86.7 | 91.3 | - | 99.3 | 68.3 | 68.3 | 62.2 |
| RwNUDX1-2c' | 76.2 | 76.2 | 78.1 | 78.1 | 86.0 | 92.0 | 98.0 | 87.3 | 92.0 | 99.3 | - | 69.0 | 69.0 | 62.9 |
| ReNUDX1-3 | 70.3 | 71.7 | 72.4 | 72.4 | 66.9 | 71.7 | 69.0 | 67.6 | 69.7 | 68.3 | 69.0 | - | 100.0 | 61.9 |
| RwNUDX1-3 | 70.3 | 71.7 | 72.4 | 72.4 | 66.9 | 71.7 | 69.0 | 67.6 | 69.7 | 68.3 | 69.9 | 100.0 | - | 61.9 |
| AtnUDX1 | 64.3 | 64.3 | 66.4 | 66.4 | 60.8 | 62.2 | 62.2 | 60.8 | 63.6 | 62.2 | 62.9 | 61.9 | 61.9 | - |

Table S1.4 | Summary of QTLs for sesquiterpenoids detected with nonparametric Kruskal-Wallis rank-sum test in the OW progeny. In bold, QTLs also detected by Interval Mapping.

| Compound | Year | QTL characteristics | | | |
|----------------------------------|-------------|------------------------|--------------------------|-----------------------|----------------|
| | | LG ^a | MM ^b | Position ^c | K ^d |
| <i>E,E</i> -farnesol | 2015 | A1 | Rh12GR_98623_1205 | 67.068 | *** |
| | 2014 | A3 | Rh12GR_51455_751 | 19.414 | **** |
| | 2015 | A5 | Rh12GR_51891_352 | 14.073 | *** |
| | 2015 | B5 | RhMCRND_7474_560 | 61.167 | *** |
| | 2015 | B6 | Rh12GR_12629_1266 | 47.96 | *** |
| | 2014 | B7 | Rh12GR_21458_519 | 51.173 | ***** |
| | 2015 | B7 | Rh88_33312_1396 | 50.506 | ***** |
| <i>E,E</i> -farnesal | 2015 | A1 | Rh12GR_59097_179 | 65.735 | **** |
| | 2014 | A3 | RhK5_18608_170 | 43.023 | ***** |
| | 2015 | A3 | Rh12GR_54386_321 | 49.708 | **** |
| | 2015 | B6 | Rh12GR_37919_235 | 47.294 | *** |
| | 2014 | B7 | Rh88_33312_1396 | 50.506 | ***** |
| | 2015 | B7 | Rh88_33312_1396 | 50.506 | ***** |
| <i>E</i> - β -farnesene | 2015 | A1 | Rh12GR_98623_1205 | 67.068 | *** |
| | 2015 | A4 | Rh12GR_24441_2313 | 4.685 | **** |
| | 2015 | B5 | RhMCRND_29430_292 | 59.825 | **** |
| | 2014 | B6 | Rh12GR_150_3109 | 39.257 | ***** |
| | 2015 | B6 | Rh12GR_12629_1266 | 47.960 | **** |
| | 2014 | B7 | Rh88_33312_1396 | 50.506 | ***** |
| 2015 | B7 | Rh88_33312_1396 | 50.506 | ***** | |
| <i>E,E</i> - α -farnesene | 2015 | A5 | Rh12GR_8671_347 | 34.127 | **** |
| | 2015 | B5 | RhMCRND_7474_560 | 61.167 | *** |
| | 2014 | B6 | Rh12GR_79916_1503 | 42.591 | **** |
| | 2015 | B6 | Rh12GR_12629_1266 | 47.96 | ***** |
| | 2014 | B7 | Rh12GR_21856_430 | 45.164 | ***** |
| | 2015 | B7 | Rh88_33312_1396 | 50.506 | ***** |
| <i>Z,E</i> - α -farnesene | 2014 | A7 | RhK5_5399_322 | 0 | *** |
| | 2014 | B6 | Rh12GR_14196_283 | 56.663 | **** |
| | 2015 | B6 | Rh12GR_12629_1266 | 47.960 | ***** |
| | 2014 | B7 | Rh12GR_21856_430 | 45.164 | ***** |
| | 2015 | B7 | Rh88_33312_1396 | 50.506 | ***** |

^a Linkage group (LG): A female map; B, male map

^b Closest molecular marker (MM) associated

^c Position on the LG (cM)

^d K value: Significance levels: *** 0.01, **** 0.005, ***** 0.001, ***** 0.0005, ***** 0.0001

Table S1.5 | Summary of QTLs for sesquiterpenoids detected with Interval Mapping in the OW progeny. The QTLs were detected in the male parent (Rw) and in the female parent (OW).

| Compound | Year | QTL characteristics | | | | | |
|----------------------------------|------|---------------------|-----------------|------------------------|-----------------------|-------------------|-----------------|
| | | PT ^a | LG ^b | LOD score ^c | Position ^d | MM ^e | r ^{2f} |
| <i>E,E</i> -farnesol | 2014 | 2.7 | B7 | 22.36 | 51.173 | Rh12GR_21458_519 | 49.9 |
| | 2015 | 2.7 | B7 | 27.87 | 50.506 | Rh88_33312_1396 | 62.2 |
| <i>E,E</i> -farnesal | 2014 | 2.6 | A3 | 4.43 | 19.414 | Rh12GR_51455_751 | 12.8 |
| | 2014 | 2.7 | B7 | 11.77 | 50.506 | Rh88_33312_1396 | 61.9 |
| | 2015 | 2.6 | B7 | 27.68 | 50.506 | Rh88_33312_1396 | 30.5 |
| <i>E</i> - β -farnesene | 2014 | 2.6 | B6 | 3.02 | 48.627 | RhMCRND_11020_192 | 8.9 |
| | 2014 | 2.6 | B7 | 28.43 | 51.173 | Rh12GR_21458_519 | 58.5 |
| | 2015 | 2.6 | B7 | 28.55 | 50.506 | Rh88_33312_1396 | 63.1 |
| <i>E,E</i> - α -farnesene | 2015 | 2.6 | B6 | 2.82 | 47.96 | Rh12GR_12629_1266 | 9.4 |
| | 2014 | 2.6 | B7 | 16.08 | 45.164 | Rh12GR_21856_430 | 39.2 |
| | 2015 | 2.6 | B7 | 38.92 | 50.506 | Rh88_33312_1396 | 74.3 |
| <i>Z,E</i> - α -farnesene | 2015 | 2.8 | B6 | 2.51 | 47.96 | Rh12GR_12629_1266 | 8.4 |
| | 2014 | 2.6 | B7 | 8.17 | 45.164 | Rh12GR_21856_430 | 22.3 |
| | 2015 | 2.8 | B7 | 36.94 | 50.506 | Rh88_33312_1396 | 72.4 |

^a The threshold of the LOD score was defined by a permutation test (PT)

^b Linkage group (LG): A female map; B, male map

^c QTLs with a LOD higher than the threshold LOD were considered

^d position on the linkage group (cM)

^e closest molecular marker (MM) associated ^f % of explanation r²

Table S1.6 | X-ray Diffraction data collection and refinement statistics.

| | RhNUDX1 | RhNUDX1/GPP |
|---|------------------------|-------------------------|
| PDB code | 6YPB | 6YPF |
| Wavelength (Å) | 0.9800 | 0.9700 |
| Resolution range (Å) | 39.02-1.70 (1.76-1.70) | 104.30-1.45 (1.50-1.45) |
| Space group | P 31 2 1 | P 31 2 1 |
| Unit cell | | |
| a, b, c | 48.674 48.674 103.132 | 49.2873 49.2873 104.286 |
| α, β, γ | 90 90 120 | 90 90 120 |
| Total reflections | 179864 (18021) | 117623 (10638) |
| Unique reflections | 16181 (1581) | 26639 (2589) |
| Multiplicity | 11.1 (11.4) | 4.4 (4.1) |
| Completeness (%) | 99.78 (99.56) | 99.45 (98.69) |
| Mean I/sigma(I) | 6.59 (0.86) | 10.66 (0.69) |
| R-merge (%) | 0.2217 (1.748) | 0.05642 (0.7807) |
| CC1/2 | 0.996 (0.476) | 0.997 (0.512) |
| CC* | 0.999 (0.803) | 0.999 (0.823) |
| Reflections used in refinement | 16172 (1574) | 26592 (2565) |
| Reflections used for R-free | 1625 (159) | 2007 (194) |
| R-work | 0.2083 (0.3798) | 0.2030 (0.3701) |
| R-free | 0.2429 (0.4166) | 0.2169 (0.4159) |
| CC(work) | 0.947 (0.740) | 0.953 (0.723) |
| CC(free) | 0.936 (0.618) | 0.943 (0.704) |
| Number of non-hydrogen atoms | 1135 | 1182 |
| macromolecules | 1057 | 1055 |
| ligands | - | 19 |
| solvent | 78 | 108 |
| RMS (bonds) | 0.006 | 0.007 |
| RMS (angles) | 0.83 | 0.77 |
| Ramachandran favored (%) | 97.58 | 98.37 |
| Ramachandran outliers (%) | 0.00 | 0.00 |
| Rotamer outliers (%) | 0.00 | 0.00 |
| Average B-factor (Å²) | 30.16 | 35.84 |
| macromolecules | 29.69 | 34.49 |
| ligands | - | 70.00 |
| solvent | 36.44 | 43.05 |

Statistics for the highest-resolution shell are shown in brackets.

Table S1.7 | Summary of values analysis during 20ns molecular dynamics. The values of diagram plotted on Figure S3.S6 are averaged on the whole 20ns trajectory +/- standard deviation; RMSD, Root mean Square Deviation.

| | Protein C α RMSD (Å) | | | Distance Mg ²⁺ -phosphate group (Å) | | | Positional ligand RMSD (Å) | | | Predicted binding energy (kCal.mol ⁻¹) | | |
|------------|-----------------------------|-------------|-------------|--|-------------|-------------|----------------------------|-------------|-------------|--|--------------|--------------|
| | 8-oxo-dGTP | GPP | FPP | 8-oxo-dGTP | GPP | FPP | 8-oxo-dGTP | GPP | FPP | 8-oxo-dGTP | GPP | FPP |
| RhNUDX1 | 1.72 ± 0.23 | 1.32 ± 0.13 | 2.12 ± 0.24 | 1.87 ± 0.05 | 1.79 ± 0.03 | 1.79 ± 0.03 | 0.44 ± 0.05 | 2.06 ± 0.27 | 1.62 ± 0.22 | -7.07 ± 0.15 | -7.54 ± 0.11 | -8.65 ± 0.19 |
| RwNUDX1-2c | 2.69 ± 0.46 | 1.50 ± 0.35 | 1.40 ± 0.25 | 1.78 ± 0.03 | 1.85 ± 0.05 | 1.82 ± 0.04 | 1.19 ± 0.32 | 1.97 ± 0.21 | 1.71 ± 0.24 | -5.95 ± 0.17 | -7.63 ± 0.15 | -8.83 ± 0.18 |
| AtNUDX1 | 1.98 ± 0.38 | 1.63 ± 0.37 | 1.64 ± 0.28 | 1.87 ± 0.05 | 1.79 ± 0.04 | 1.80 ± 0.04 | 0.56 ± 0.16 | 1.55 ± 0.40 | 2.01 ± 0.21 | -7.20 ± 0.15 | -7.65 ± 0.17 | -8.93 ± 0.13 |
| RcNUDX1-1a | 1.92 ± 0.57 | 2.77 ± 0.52 | 1.95 ± 0.26 | 1.87 ± 0.06 | 1.78 ± 0.03 | 1.82 ± 0.06 | 0.57 ± 0.11 | 1.85 ± 0.17 | 1.76 ± 0.31 | -6.63 ± 0.17 | -7.49 ± 0.10 | -8.43 ± 0.17 |
| RcNUDX1-1b | 1.67 ± 0.24 | 1.56 ± 0.16 | 1.93 ± 0.34 | 1.87 ± 0.05 | 1.78 ± 0.04 | 1.81 ± 0.04 | 0.62 ± 0.11 | 1.45 ± 0.17 | 1.70 ± 0.16 | -6.78 ± 0.17 | -7.11 ± 0.15 | -8.51 ± 0.09 |

Table S1.8 | Accession numbers of the *NUDX1* sequences.

| Sequence name | GenBank accession number | Genome ID ^a |
|--------------------|--------------------------|---|
| <i>RcNUDX1-1a</i> | - | RcHm_v2.0_Chr2g0142051/61 ^b RcHm_v2.0_Chr2g0142071 RcHm_v2.0_Chr2g0142081 ^c RcHm_v2.0_Chr2g0142111 RcHm_v2.0_Chr2g0142121 |
| <i>RcNUDX1-1b</i> | - | RcHm_v2.0_Chr4g0436181 |
| <i>RcNUDX1-2a</i> | - | RcHm_v2.0_Chr4g0436151 |
| <i>RcNUDX1-2b</i> | - | RcHm_v2.0_Chr6g0244161 |
| <i>RcNUDX1-2c</i> | - | RcHt_S2031.3 |
| <i>RcNUDX1-3</i> | - | RcHm_v2.0_Chr4g0436191 |
| <i>RwNUDX1-1</i> | MT362556 | - |
| <i>RwNUDX1-2a</i> | MT362557 | - |
| <i>RwNUDX1-2b</i> | MT362558 | - |
| <i>RwNUDX1-2c</i> | MT362559 | - |
| <i>RwNUDX1-2c'</i> | MT362560 | - |
| <i>RwNUDX1-3</i> | MT362561 | - |

^a, Reference of the genes are according to the GDR database (<https://www.rosaceae.org/>)

^b, A STOP codon is interrupting the ORF of this sequence, this sequence is not shown in Figure S1.

^c, Sequence shown in Figure 1.

Appendix 1

| | | | | | | | | | | | | | | | | | | | | | | |
|------------------------|---|---|----|---|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|----|----|-----|
| | 1 | | 10 | | 20 | | | | | | | | | | | | | | | | | |
| RcHm_v2.0_Chr2g0142071 | M | G | N | E | T | V | V | V | A | E | T | S | G | S | I | K | V | A | V | V | V | C |
| RcHm_v2.0_Chr2g0142081 | M | G | N | E | T | V | V | V | A | E | T | S | G | S | I | K | V | A | V | V | V | C |
| RcHm_v2.0_Chr2g0142111 | M | G | N | E | T | V | V | V | A | E | T | S | G | S | I | K | V | A | V | V | V | C |
| RcHm_v2.0_Chr2g0142121 | M | G | N | E | T | V | V | V | A | E | T | S | G | S | I | K | V | A | V | V | V | C |
| | | | | | | 30 | | | | | | | | | | | | | | 40 | | |
| RcHm_v2.0_Chr2g0142071 | L | L | R | G | Q | N | V | L | L | G | R | R | R | S | S | L | G | D | S | T | F | S |
| RcHm_v2.0_Chr2g0142081 | L | L | R | G | Q | N | V | L | L | G | R | R | R | S | S | L | G | D | S | T | F | S |
| RcHm_v2.0_Chr2g0142111 | L | L | R | G | Q | N | V | L | L | G | R | R | R | S | S | L | G | D | S | T | F | S |
| RcHm_v2.0_Chr2g0142121 | L | L | R | G | Q | N | V | L | L | G | R | R | R | S | S | L | G | D | S | T | F | S |
| | | | | | | | 50 | | | | | | | | | | | | | | 60 | |
| RcHm_v2.0_Chr2g0142071 | L | P | S | G | H | L | E | F | G | E | S | F | E | E | C | A | A | R | E | L | K | E |
| RcHm_v2.0_Chr2g0142081 | L | P | S | G | H | L | E | F | G | E | S | F | E | E | C | A | A | R | E | L | K | E |
| RcHm_v2.0_Chr2g0142111 | L | P | S | G | H | L | E | F | G | E | S | F | E | E | C | A | A | R | E | L | K | E |
| RcHm_v2.0_Chr2g0142121 | L | P | S | G | H | L | E | F | G | E | S | F | E | E | C | A | A | R | E | L | K | E |
| | | | | | | | | 70 | | | | | | | | | | | | | | 80 |
| RcHm_v2.0_Chr2g0142071 | E | T | D | L | D | I | G | K | I | E | L | L | T | V | T | N | N | L | F | L | D | E |
| RcHm_v2.0_Chr2g0142081 | E | T | D | L | D | I | G | K | I | E | L | L | T | V | T | N | N | L | F | L | D | E |
| RcHm_v2.0_Chr2g0142111 | E | T | D | L | D | I | G | K | I | E | L | L | T | V | T | N | N | L | F | L | D | E |
| RcHm_v2.0_Chr2g0142121 | E | S | D | L | D | I | G | K | I | Q | L | L | T | V | T | N | N | L | F | L | D | E |
| | | | | | | | | | | | | | | | | | | | | | | 90 |
| RcHm_v2.0_Chr2g0142071 | A | K | P | S | Q | Y | V | A | V | F | M | R | A | V | L | A | D | P | R | Q | E | P |
| RcHm_v2.0_Chr2g0142081 | A | K | P | S | Q | Y | V | A | V | F | M | R | A | V | L | A | D | P | R | Q | E | P |
| RcHm_v2.0_Chr2g0142111 | A | K | P | S | Q | Y | V | A | V | F | M | R | A | V | L | A | D | P | R | Q | E | P |
| RcHm_v2.0_Chr2g0142121 | A | K | P | S | Q | Y | V | A | V | F | M | R | A | V | L | A | D | P | R | Q | E | P |
| | | | | | | | | | | | | | | | | | | | | | | 100 |
| RcHm_v2.0_Chr2g0142071 | Q | N | I | E | P | K | F | C | D | G | W | G | W | Y | E | W | D | N | L | P | K | P |
| RcHm_v2.0_Chr2g0142081 | Q | N | I | E | P | K | F | C | D | G | W | G | W | Y | E | W | D | N | L | P | K | P |
| RcHm_v2.0_Chr2g0142111 | Q | N | I | E | P | K | F | C | D | G | W | G | W | Y | E | W | D | N | L | P | K | P |
| RcHm_v2.0_Chr2g0142121 | Q | N | I | E | P | K | F | C | D | G | W | G | W | Y | E | W | D | N | L | P | K | P |
| | | | | | | | | | | | | | | | | | | | | | | 110 |
| RcHm_v2.0_Chr2g0142071 | L | F | W | P | L | E | N | V | V | Q | D | G | F | N | P | F | P | T | | | | |
| RcHm_v2.0_Chr2g0142081 | L | F | W | P | L | E | N | V | V | Q | D | G | F | N | P | F | P | T | | | | |
| RcHm_v2.0_Chr2g0142111 | L | F | W | P | L | E | N | V | V | Q | D | G | F | N | P | F | P | T | | | | |
| RcHm_v2.0_Chr2g0142121 | L | F | W | P | L | E | N | V | V | Q | D | G | F | N | P | F | P | T | | | | |

Figure S1.1 | Alignment of amino acid sequences of four *RcNUDX1-1a* copies present in *R. chinensis* cv. 'Old Blush'. RcHm_v2.0_Chr2g0142051/61 was not aligned, as A STOP codon is interrupting the ORF.

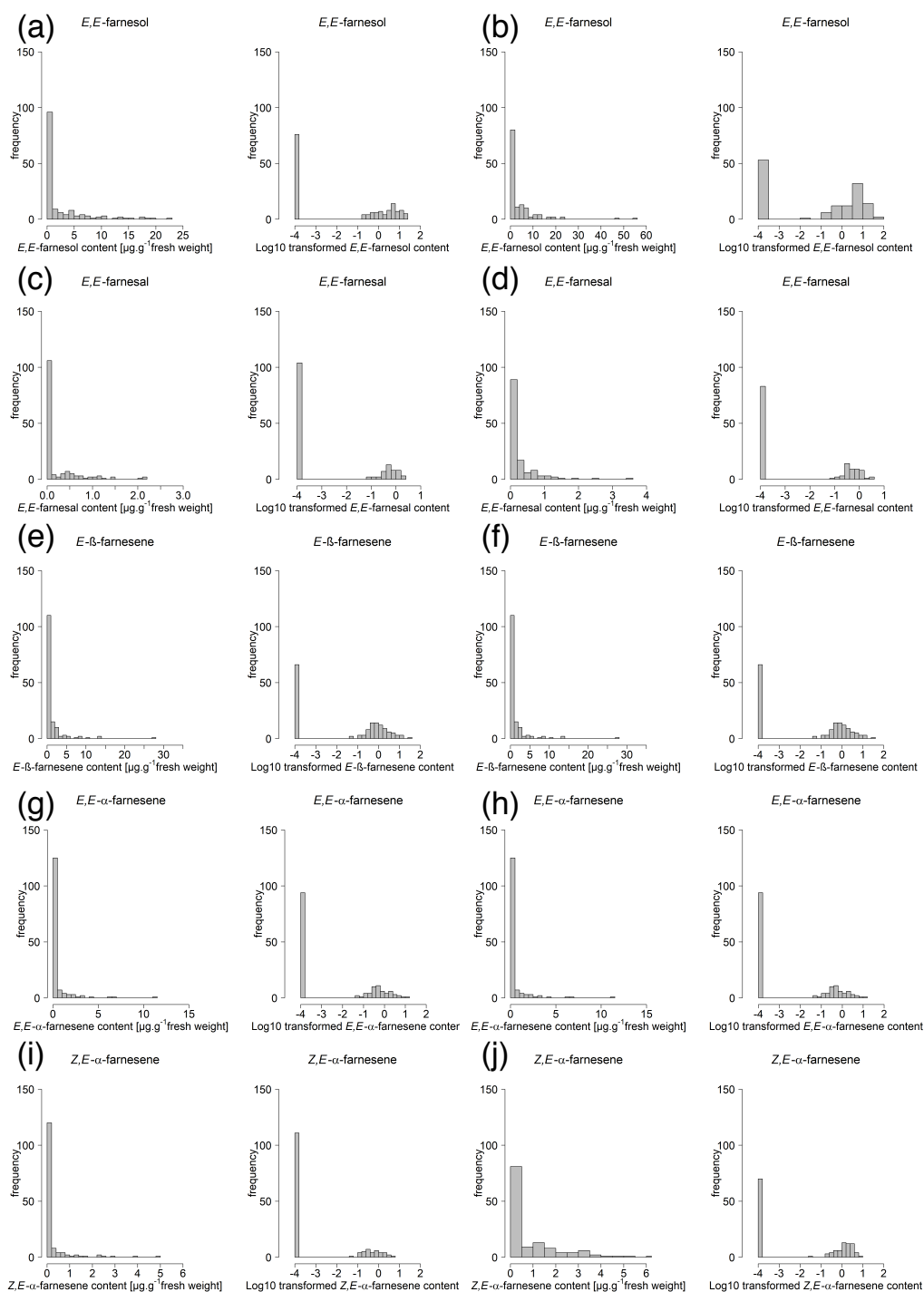


Figure S1.2 | Distribution in the OW progeny (raw data and logarithm transformed data) of volatile compound amounts for *E,E*-farnesol (a and b), *E,E*-farnesal (c and d), *E*- β -farnesene (e and f), *E,E*- α -farnesene (g and h) and *Z,E*- α -farnesene (i and j), Samples were collected in 2014 (a, c, e, g, i) and 2015 (b, d, f, h).

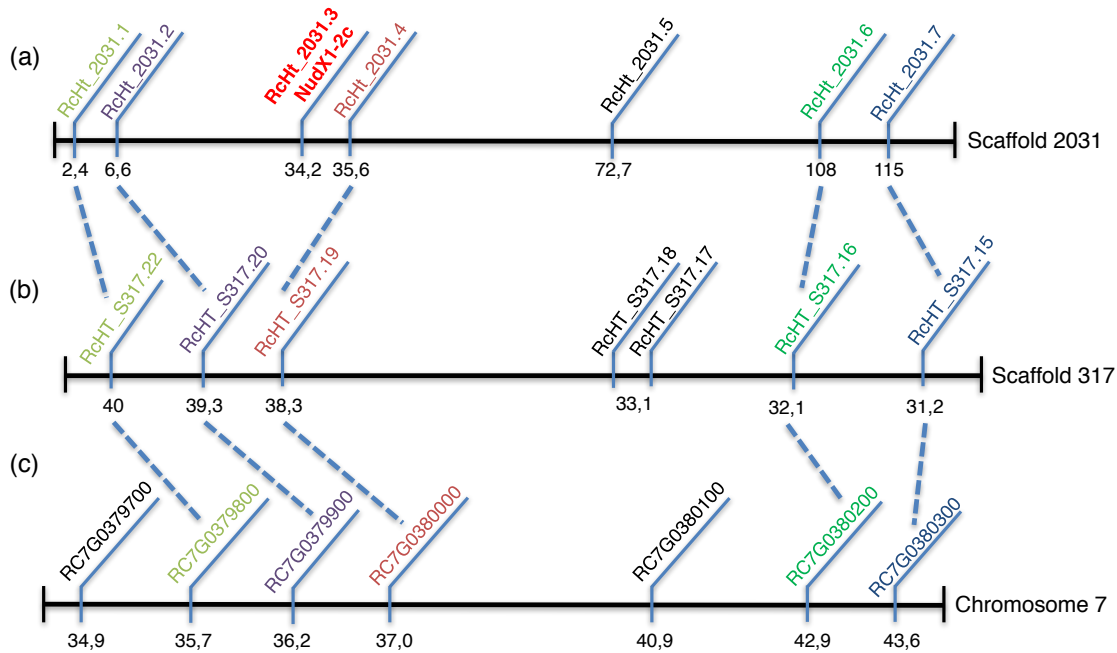


Figure S1.4 | Micro-synteny analysis at the *NUDX1-2c* locus between **a** and **b**) the heterozygous sequence of *R. chinensis* cv. 'Old Blush' (OB) and **c**) the homozygous reference sequence obtained from OB. *NUDX1-2c* is located on scaffold 2031 from OB (in red). Five genes surrounding *NUDX1-2c* present highly similar sequences (e value below e^{-100} from BLASTP results) on scaffold 317 (from OB) and on chromosome 7 (from homozygous reference sequence). Highly similar sequences are linked with dotted lines and the genes are in the same color. Positions are indicated in kb for OB (a and b) and in Mb for the homozygous reference sequence (c). Reference of the genes are according to the GDR database (<https://www.rosaceae.org/>).

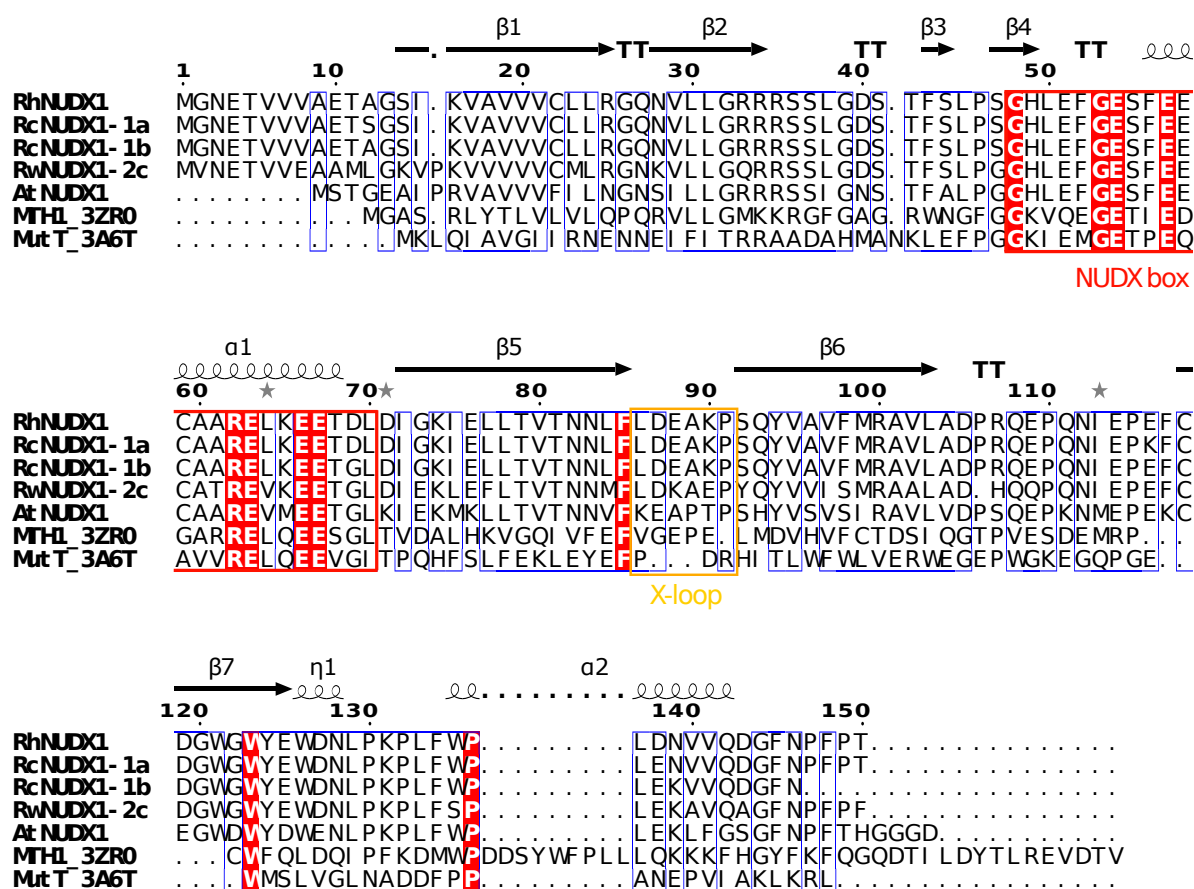


Figure S1.5 | Amino acid sequences alignment and structure indication of NUDX1 proteins for protein structure modelling using ESPrIPT 3.0 (Robert and Gouet, 2014). *RhNUDX1*: NUDX1 protein of *R. x hybrida* cv. ‘Papa Meiland’; *RcNUDX1-1a* and *RcNUDX1-1b*: two NUDX1 proteins of *R. chinensis* cv. ‘Old Blush’; *RwNUDX1-2*: NUDX1 protein from *R. x wichurana*; and *AtNUDX1*: NUDX1 protein from *A. thaliana*; *MTH1_3ZR0*: NUDX1 protein from *Homo sapiens* (Protein database bank ID 3ZR0); *MutT_3A6T*: NUDX1 protein from *E. coli* (Protein database bank ID 3A6T). Conserved amino acids are highlighted in red. NUDX box is highlighted in red and X-loop in yellow.

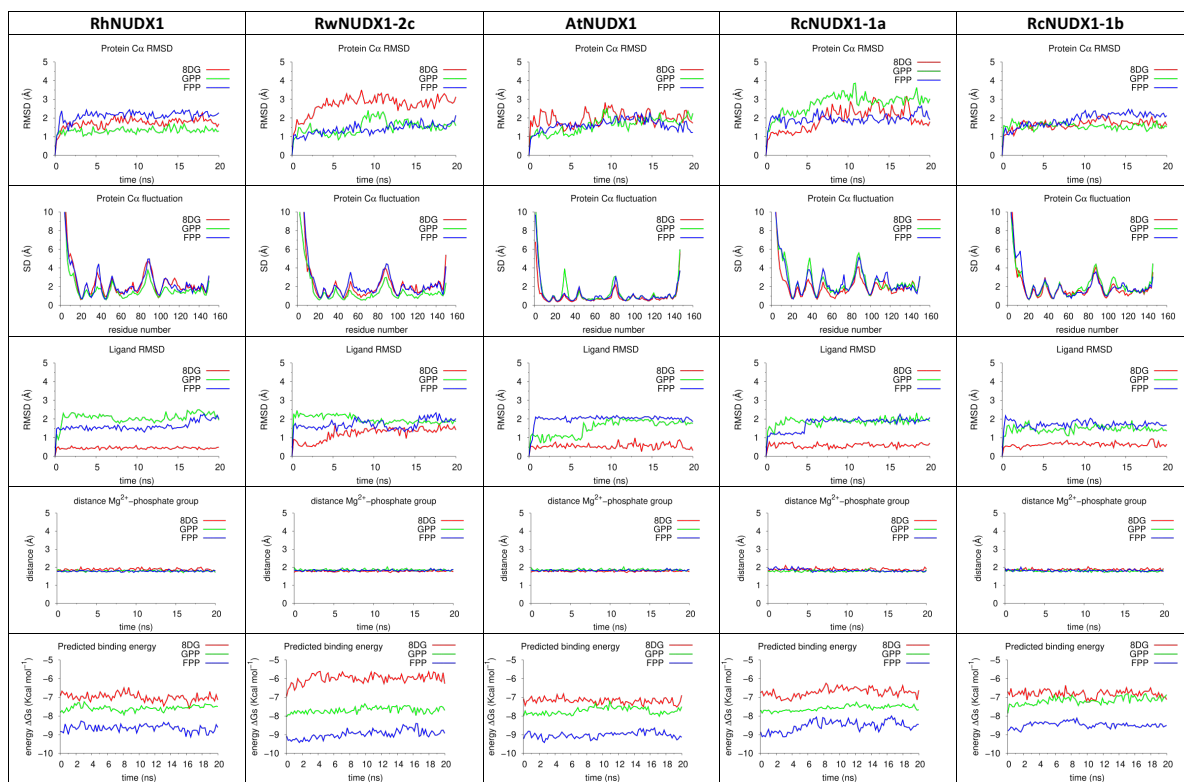


Figure S1.6 | Analysis of the 20ns trajectories for RhNUDX1, RwnUDX1-2c, AtNUDX1, RcNUDX1-1a and RcNUDX1-1b. Each row corresponds to a specific kind of analysis for all the protein modelled. Lines are coloured according to substrate modelled (red, 8-oxo-dGTP, 8DG; green, GPP; blue, FPP). First row, protein C α RMSD (root mean square deviation); Second row, protein C α fluctuation over sequence (SD, standard deviation); Third row, ligand position RMSD; Fourth row, Mg²⁺-oxygen phosphate of ligand; Fifth row, ligand binding energy estimated by PRODIGY-LIG.

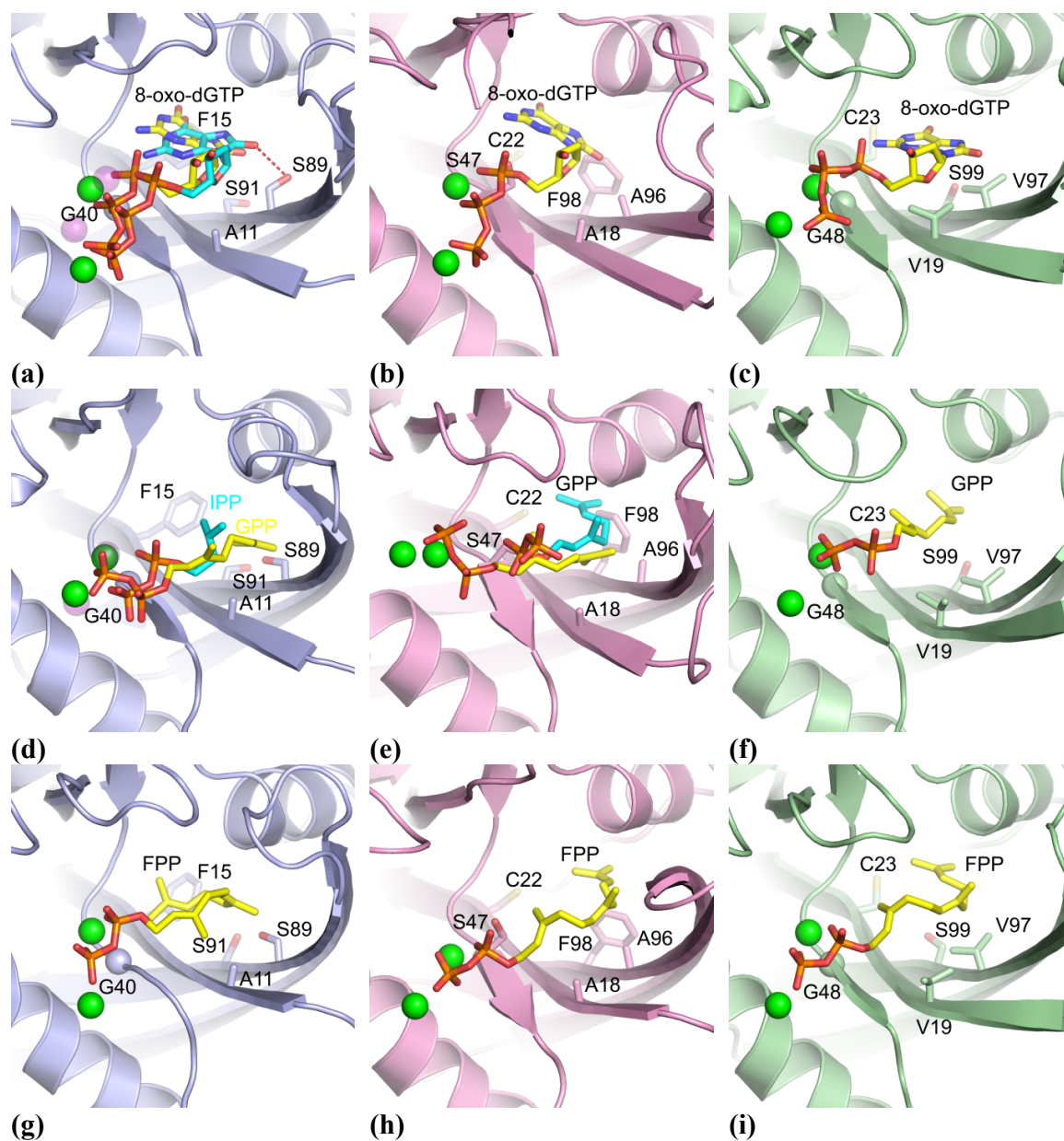


Figure S1.7 | Models of NUDX1- substrate interactions. Interaction of AtNUDX1 (blue) with 8- oxo-dGTP (a), GPP (d) and FPP (g). Interaction of RhNUDX1 (pink) with 8-oxo-dGTP (b), GPP (e) and FPP (h). Interaction of RwNUDX1-2c (green) with 8-oxo-dGTP (c), GPP (f) and FPP (i). Magnesium ions involved in ligand coordination are shown as green spheres. GPP and FPP are shown as sticks in yellow (geranyl and farnesyl moieties) and orange (diphosphate moiety). The ligands of known crystal structure are superimposed in cyan in a (8-oxo-dGTP from PDB code 6FL4), d (IPP from PDB code 6DBZ) and e (GPP from PDB code 5GP0). Hydrogen bond between S89 and 8-oxo-dGTP in AtNUDX1 is drawn as a dashed line.

Appendix 2 |

Supplementary materials for chapter 2

Supplementary dataset table S8 is given as a digital form (.xlsx)

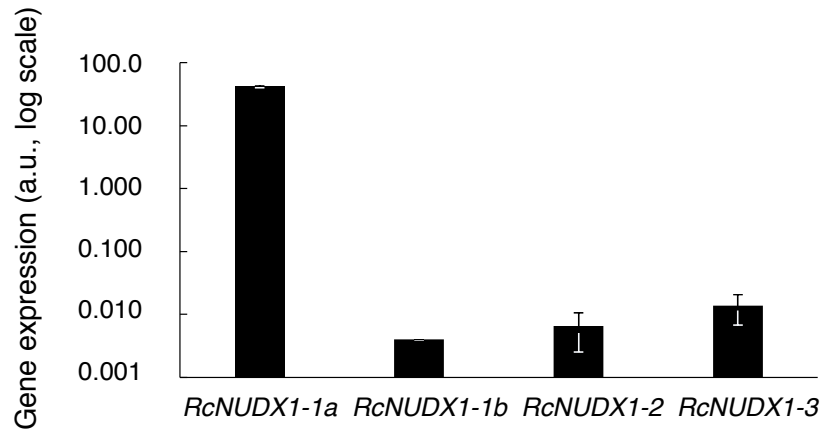


Figure S2.1 | Expression of *RcNUDX1* genes in petals of full-opened flowers of Old Blush. The forward primer was designed to target a conserved sequence of the last exon of both genes including two different non-contiguous nucleotides, while reversed primers was targeted to a specific and less conserved sequence in the 3'-UTR sequence (primers FP1- RP1 to FP4-RP4, and FP6-RP6 for reference gene, table S7). *RcNUDX1-2* represents the expression of *RcNUDX1-2a*, *RcNUDX1-2b*, and *RcNUDX1-2c*, because the primers bound the three sequences in qRT-PCR. Error bars correspond to SD.

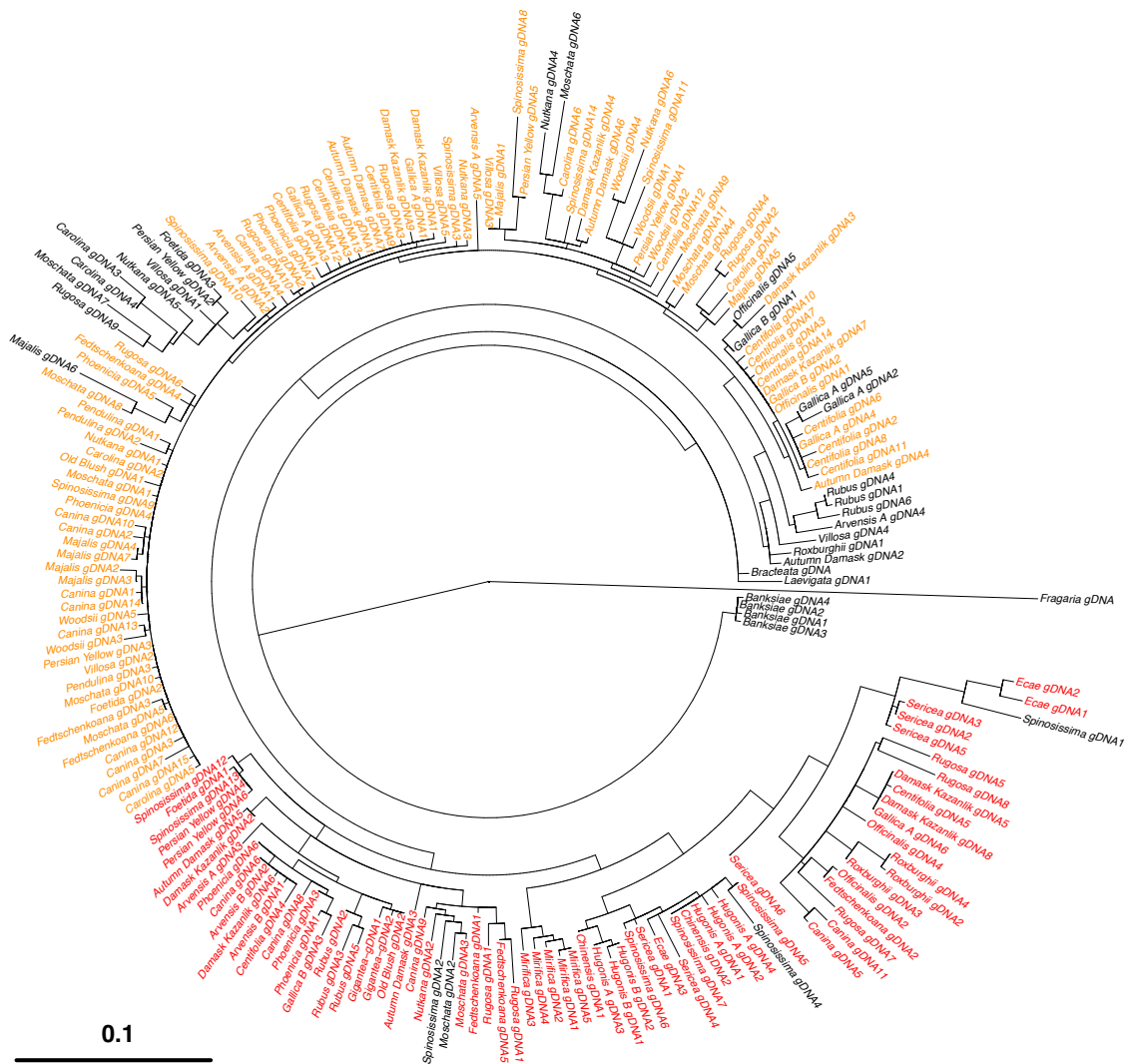


Figure S2.2 | Extended ML tree of genomic sequences of the Nudx1-1 clade. The tree was made with gDNA sequences (ATG to STOP including introns; FP7-RP7 primers were used for cloning; table S7, Clones_IntronExonStructure.fasta, Alignment_OldBlush_MLtree.fasta); Orange branches are gDNAs that matched with *RcNUDX1-1a* by blast analysis (i.e. their identity percentages with *RcNUDX1-1a* and *RcNUDX1-1b* differ by more than 1% in favor of *RcNUDX1-1a*), red branches, with *RcNUDX1-1b* (i.e. their identity percentages differ by more than 1% in favor of *RcNUDX1-1b*), and dark branches, with both (i.e. their identity percentages differ by less than 1%). Scale bar represent substitution per site.

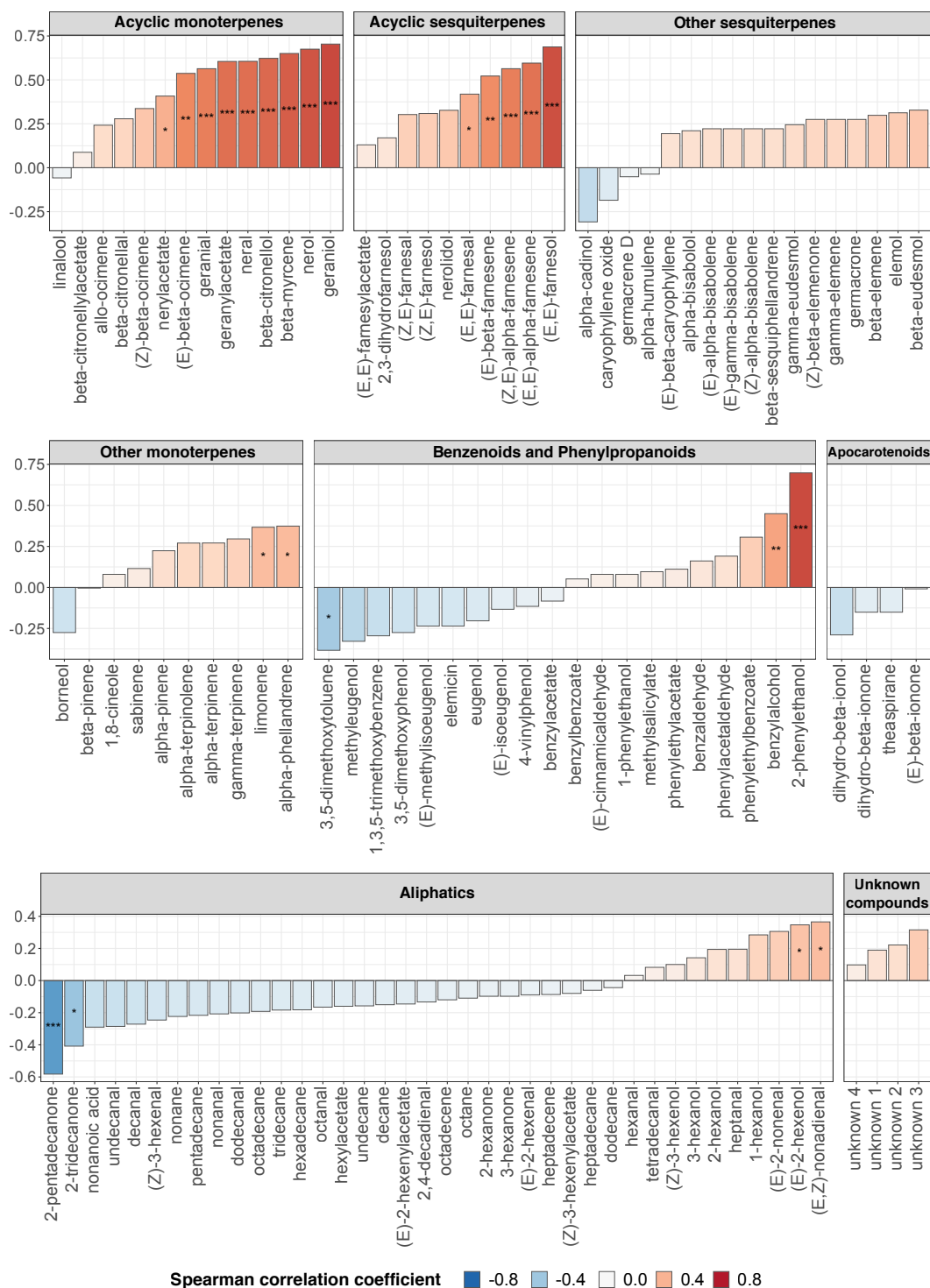


Figure S2.3 | Heatmap correlation between expression of *NUDX1-1* homologs and concentrations of volatile compounds in 34 accessions of botanical roses. Blue to red colors and histogram heights are redundant information. Statistical significances are indicated by stars (P -values: ***<0.001; **0.001-0.01; *0.01-0.05). Compound concentrations, correlation coefficients and P -values are given in tables S8 and S11. Primers FP8-RP8 were used for amplification and primers FP5-RP5 and FP6-RP6 for reference genes (table S7).

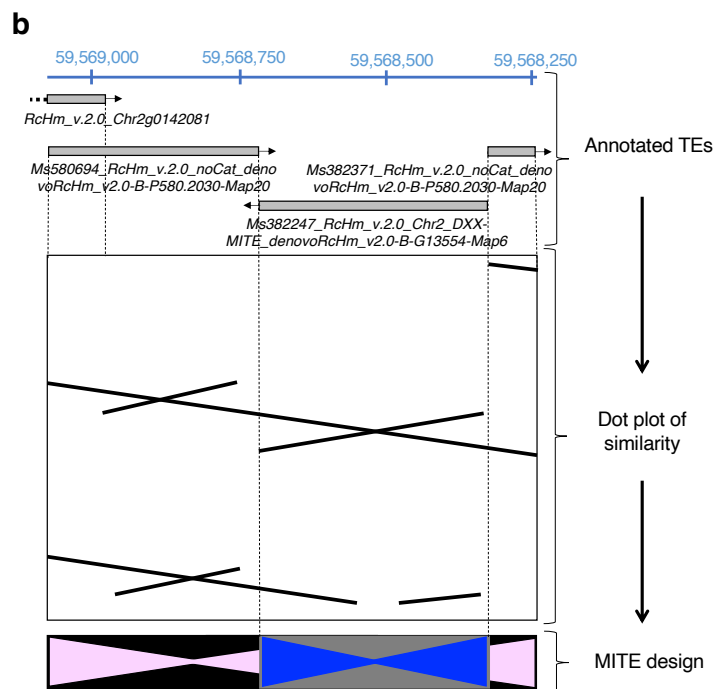
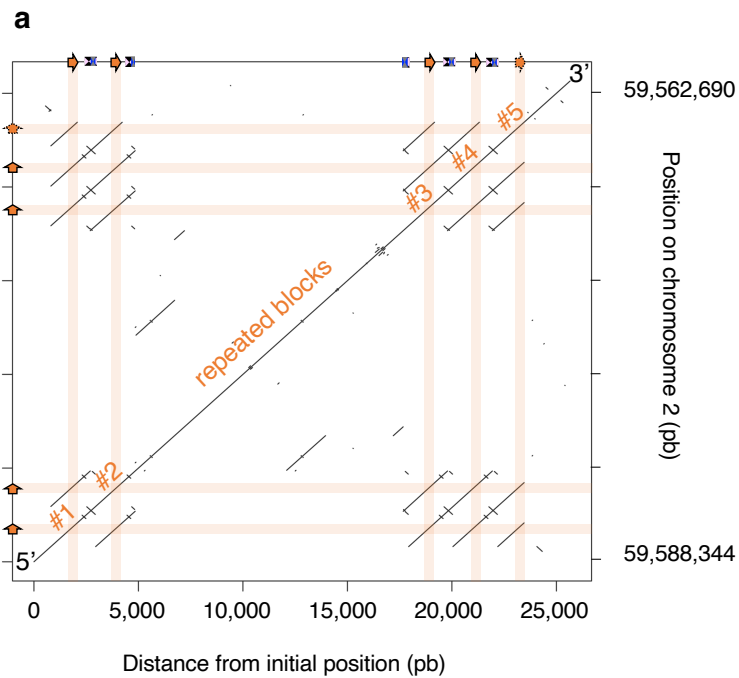


Figure S2.4 | Study of intergenic homologies in Old Blush chromosome 2. **a.** Dot-plot of similarity showing repeated blocks (in green), *RcNUDX1-1a* copies (in orange, Y axis on the left), and MITE positions in the intergenic regions (upside X axis). Downside X axis and Y axis on the right are the positions and the lengths of the DNA sequences on chromosome 2. **b.** Workflow showing the design of repeated MITEs in intergenic regions. Annotated TEs were search in the GDR (table S12), then characterized on a dot-plot of similarity, and finally interpreted as MITEs.

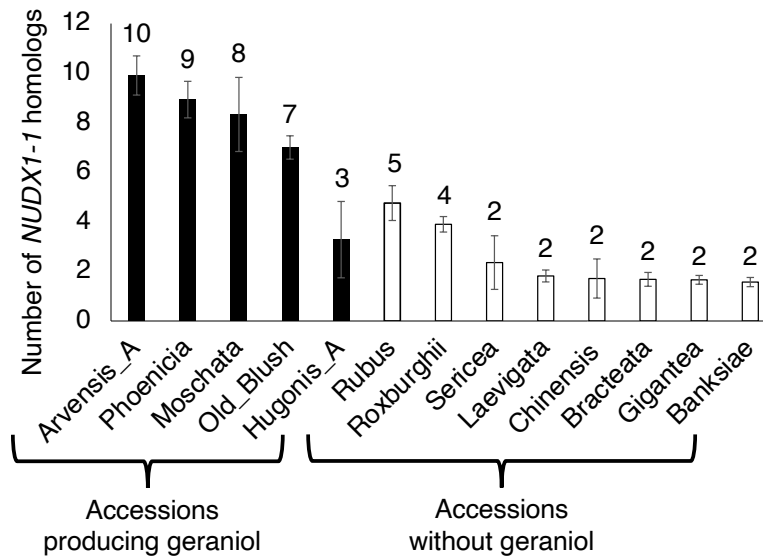
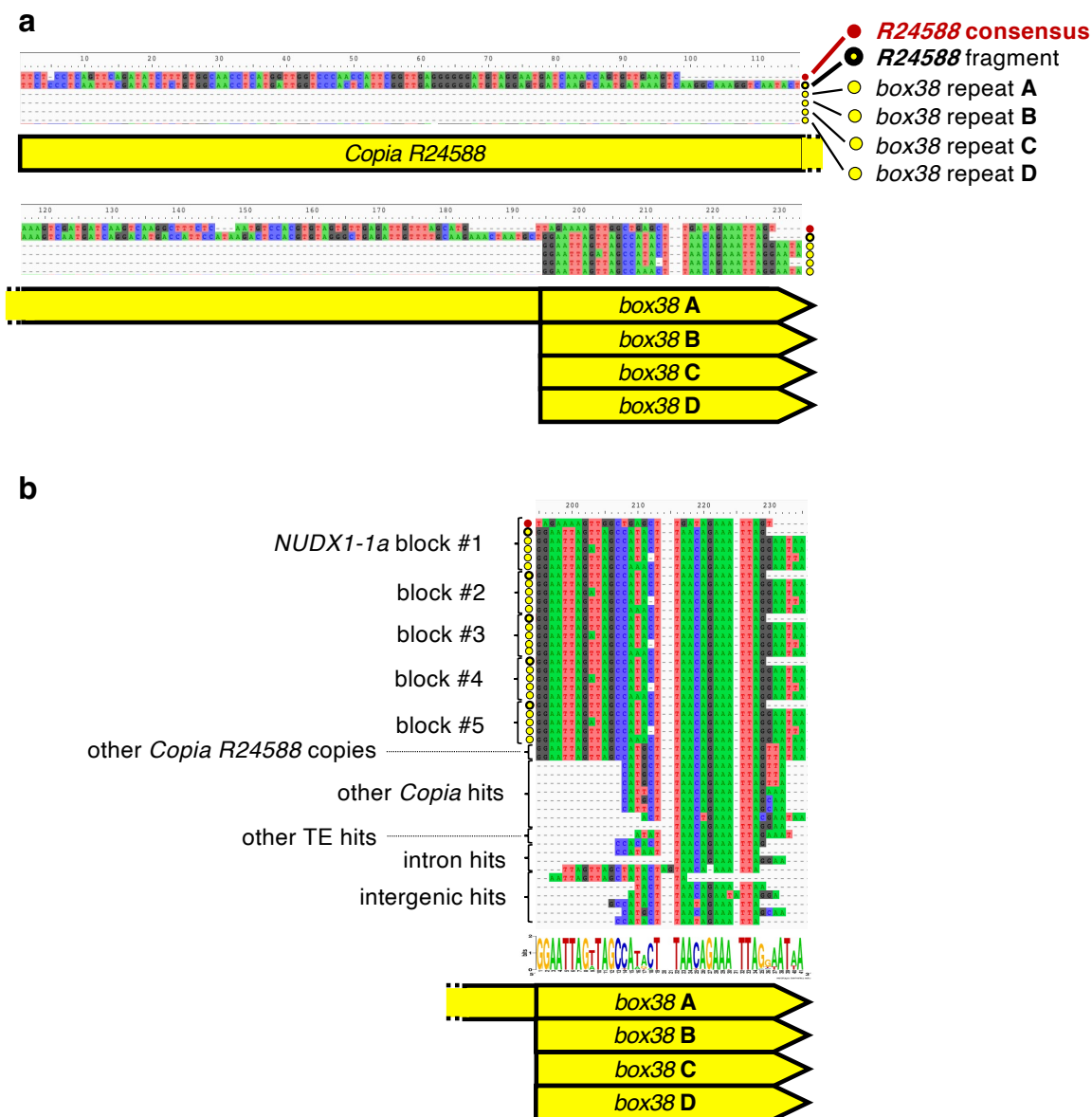


Figure S2.5 | Estimation of the number of copies of *NUDX1-1* homologs in 12 accessions of wild diploid rose, and Old Blush. The number of sequences of *NUDX1-1* homologs was estimated by qPCR, verifying that the primer (FP8-RP8; table S7) amplified seven homologs in the reference Old Blush (five *RcNUDX1-1a* copies and two *RcNUDX1-1b* alleles; table S2, Figure S3.3). The chosen wild roses were all diploid to facilitate the calculation based on Ct values obtained in PCR for Old blush, and avoid caveats. Black bars indicate accessions that produce geraniol in petals, white bars, that do not produce geraniol (table S8). Error bars correspond to SD.



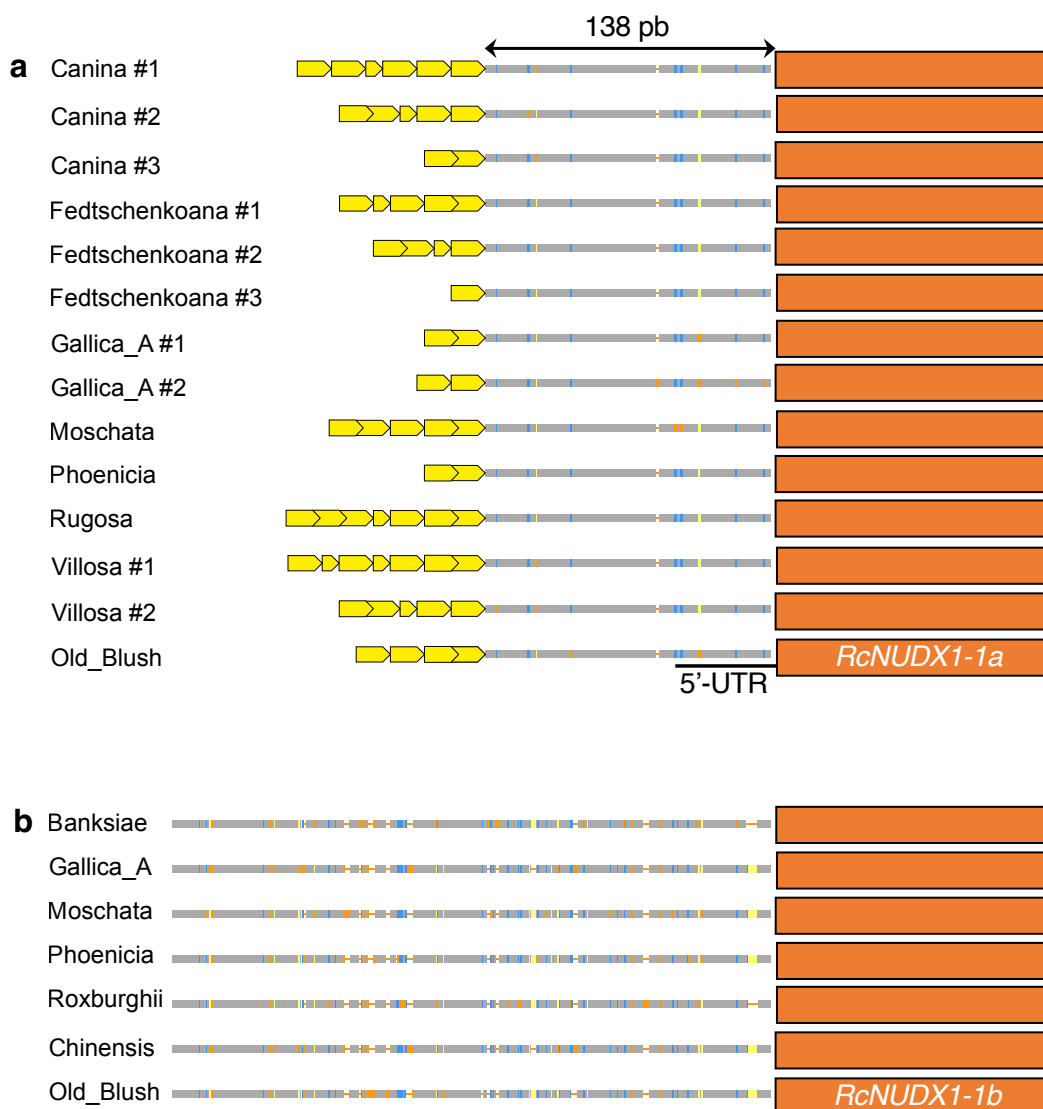


Figure S2.7 | Organization of the *box38* repetitions upstream the coding region of *NUDX1- 1a* homologs. a. Repetition of *box38* (large yellow arrows) at the 3'-end of *Copia R24588* fragment. The last *box38* is at 138 pb from the ATG codon of all *NUDX1-1a* homologs (orange boxes). The 5'- UTR of Old Blush of 55 pb is indicated. **b.** Organization of the same region upstream the coding region of *NUDX1-1b* homologs. The orange boxes correspond to the coding regions. Sequences were obtained after PCR amplification with a primer in *Copia R24588* and a primer in the coding sequence (FP9-RP9 and FP10-RP10 primers; table S7).

Table S2.1 | Dataset of botanical rose accessions.

| Accession name ^a | Botanical or horticultural name | Synonym | Location and accession number | Botanical classification | Horticultural data | Geographical origin | Plody |
|-----------------------------|--|-------------------------------|---|--------------------------|---|--|-------|
| Avensis A | <i>R. avensis</i> Huds. 1762 | Field Rose | Roseate de Saint-Clair (Caluire, France) #78 | Synsylvae Section | Wild species | West and South Europe | 2n |
| Avensis B | <i>R. avensis</i> Huds. 1762 | Field Rose | Roseate de Loubar [†] (Les Brettes, France) #3-7-37 | Synsylvae Section | Wild species | West and South Europe | 2n |
| Aurum Damask | <i>R. damasceana</i> var. <i>sempervirens</i> (Loisel. & Michel) Rowlee | Aurum Damask | Laboratoire BVPAM [†] (Saint-Etienne, France) #11 | Gaillancae Section | Heritage rose, Damask Group, Cultivated before 1633 | Unknown | 4n |
| Banksiae | <i>R. banksiae</i> var. <i>normalis</i> Reg. 1878 | Lady Banks Rose | Roseate de Saint-Clair (Caluire, France) # | Banksianae Section | Wild species | South East China | 2n |
| Bracteata | <i>R. bracteata</i> Wendt 1798 | Macartney Rose | Roseate de Saint-Clair (Caluire, France) #4 | Bracteatae Section | Wild species | East Asia | 2n |
| Canina | <i>R. canina</i> L. 1753 | Dog Briar | Wild Briar (Mornant, France) | Caninae Section | Wild species | Europe, South West Asia & North Africa | 5n |
| Carolina | <i>R. carolina</i> L. | Caroline's Rose | Roseate de Saint-Clair (Caluire, France) #21 | Carolinae Section | Wild species | North East of North America | 4n |
| Centifolia | <i>R. x centifolia</i> L. 1753 | Cabbage Rose | Roseate de Saint-Clair (Caluire, France) #46 | Gaillancae Section | Heritage rose, Cabbage Group, Cultivated before 1318 | Unknown | 4n |
| Chinensis | <i>R. chinensis</i> Jacq. var. <i>spontanea</i> (Rehd. & Wils.) T. T. Yu & T. C. Ku 1885 | Henry's Crimson China | Parc de la Tête d'Or (Lyon, France) | Chinese Section | Wild species | West China and India | 2n |
| Damask Kazanlik | <i>R. x damasceana</i> L. <i>triglobata</i> (Dieck) Keller ex Asch. & Graeb. | Kazanlik Damask Rose | Laboratoire BVPAM [†] (Saint-Etienne, France) #16 | Gaillancae Section | Heritage rose, Damask Group, Cultivated before 1612 | South East Europe & Middle East | 4n |
| Ecae | <i>R. ecae</i> Aitch. | Mrs. Aitchison's Rose | Roseate de Saint-Clair (Caluire, France) #32 | Cinnamomae Section | Wild species | South East Asia & North China | ? |
| Fedtschenkoana | <i>R. fedtschenkoana</i> Regel 1878 | Mrs. Fedtschenko's Rose | Roseate de Saint-Clair (Caluire, France) #5 | Cinnamomae Section | Wild species | Central Asia | 4n |
| Foetida | <i>R. foetida</i> J. Herm. 1911 | Austrian Briar | Roseate de Loubar [†] (Les Brettes, France) #1-5-15 | Pimpinellifoliae Section | Wild species | Asia Minor & Middle East | 4n |
| Gallica A | <i>R. gallica</i> L. 1762 | Red Rose | Roseate de Saint-Clair (Caluire, France) #53 | Gaillancae Section | Wild species | Asia Minor & Europe | 4n |
| Gallica B | <i>R. gallica</i> L. 1762 | Red Rose | Roseate de Saint-Clair (Caluire, France) #10-6-20 | Gaillancae Section | Wild species | Asia Minor & Europe | 4n |
| Gigantea | <i>R. gigantea</i> (Collet ex Ciep.) 1888 | <i>R. x odorata gigantea</i> | Roseate de Saint-Clair (Caluire, France) #63 | Chinese Section | Wild species | South of Central Asia | 2n |
| Hugonis A | <i>R. hugonis</i> Hemsl. 1905 | Father Hugo's Rose | Roseate de Saint-Clair (Caluire, France) #69 | Pimpinellifoliae Section | Wild species | China | 2n |
| Hugonis B | <i>R. hugonis</i> Hemsl. 1905 | Father Hugo's Rose | Roseate de Loubar [†] (Les Brettes, France) #10-7-22 | Pimpinellifoliae Section | Wild species | China | 2n |
| Laevigata | <i>R. laevigata</i> Michx. 1803 | Cherokee Rose | Roseate de Saint-Clair (Caluire, France) #64 | Laevigatae Section | Wild species | South East Asia | 2n |
| Majalis | <i>R. majalis</i> Herm. 1762 | Cinnamon Rose | Roseate de Saint-Clair (Caluire, France) #33 | Cinnamomae Section | Wild species | North East Europe & Russia | 2n |
| Mirica | <i>R. sieboldii</i> Minckley (Greene) Cockerell | <i>Hesperoceras miriflora</i> | Roseate de Saint-Clair (Caluire, France) #56 | Hesperoceras Subgenus | Wild species | West of North America | 2n |
| Muschata | <i>R. moschata</i> Herm. 1888 | Musk Rose | Roseate de Saint-Clair (Caluire, France) #55 | Cinnamomae Section | Wild species | South China & India | 2n |
| Nukiiana | <i>R. nukiiana</i> C. Presl. | Brook Rose | Roseate de Saint-Clair (Caluire, France) #53 | Cinnamomae Section | Wild species | North West America | 2n |
| Onitidis | <i>R. gallica</i> var. <i>onitidis</i> Andr. | Provenç Rose | Roseate de Saint-Clair (Caluire, France) #52 | Gaillancae Section | Heritage rose, Gallica Group, Cultivated before 1160 | Unknown | 4n |
| Old Blush | <i>R. chinensis</i> QJ 1951 | Blush | Laboratoire BVPAM [†] (Saint-Etienne, France) #21 | Chinese Section | Heritage rose, China Group, Cultivated before 1731 | Unknown | 4n |
| Pearl and Yellow | <i>R. pendula</i> L. 1759 | Pearl and Yellow | Roseate de Saint-Clair (Caluire, France) #49 | Pimpinellifoliae Section | Wild species | South of Central Europe | 2n |
| Praegeri | <i>R. praegeri</i> Erosa 1849 | Praegeri Rose | Roseate de Saint-Clair (Caluire, France) #66 | Synsylvae Section | Heritage rose, Hybrid Foetida Group, Cultivated before 1837 | Middle East & Asia Minor | 3n |
| Rochburghii | <i>R. roxburghii</i> Tratt. var. <i>plena</i> Rehd. 1828 | Chestnut rose | Roseate de Saint-Clair (Caluire, France) #77 | Fraxinifoliae Subgenus | Wild species | Middle East | ? |
| Rubus | <i>R. rubus</i> Lév. & Vanot 1908 | Raspberry Rose | Roseate de Saint-Clair (Caluire, France) #67 | Synsylvae Section | Wild species | West Asia | 2n ? |
| Ruscica | <i>R. ruscica</i> Thunb. 1784 | Ruscica Rose | Roseate de Saint-Clair (Caluire, France) #59 | Cinnamomae Section | Wild species | China | 2n |
| Sarica | <i>R. sericea</i> Lindl. 1820 | Himalayas Rose | Roseate de Saint-Clair (Caluire, France) #72 | Cinnamomae Section | Wild species | North China, Russia & West Asia | 2n |
| Sponsoissima | <i>R. spinosissima</i> L. 1753 | Apple Rose | Roseate de Saint-Clair (Caluire, France) # | Pimpinellifoliae Section | Wild species | North East India & South China | 4n |
| Villosa | <i>R. villosa</i> L. 1753 | Apple Rose | Roseate de Saint-Clair (Caluire, France) #20 | Caninae Section | Wild species | Europe to Asia | 4n |
| Woodsi | <i>R. woodsii</i> Lindl. 1826 | Woods Rose | Roseate de Saint-Clair (Caluire, France) #43 | Cinnamomae Section | Wild species | Central & South Europe & Asia Minor | 4n |
| | | | | | | West of North America | 2n |

^a These accession names are used in the tables and figures of the present work.

Table S2.6 | In vitro activities of NUDX1-1 using GPP as substrate.

| Accession | Enzyme name | K_m (M) | K_{cat} (sec⁻¹) | K_{cat}/K_m (sec⁻¹.M⁻¹) |
|------------------|--------------------|--------------------------|---|--|
| Old Blush | RcNUDX1-1a | 2.97E-06 ± 5.63E-07 | 0.13 ± 0.01 | 4.39E+04 ± 5.10E+03 |
| Moschata | RmNUDX1-1a | 2.27E-06 ± 2.60E-07 | 0.11 ± 0.01 | 4.87E+04 ± 2.84E+03 |
| Old Blush | RcNUDX1-1b | 2.68E-06 ± 2.37E-07 | 0.32 ± 0.01 | 1.18E+05 ± 6.53E+03 |
| Moschata | RmNUDX1-1b | 2.57E-06 ± 3.32E-07 | 0.29 ± 0.01 | 1.14E+05 ± 9.67E+03 |

Table S2.7 | List of primers.

| Primer # | Primer sequence | Use | Tm |
|----------|---|--|------|
| FP1 | 5'-TGGAGAACGTGGTTCAGGA-3' | qRT-PCR of <i>RcNUDX1-1a</i> | 59°C |
| RP1 | 5'-CAGCCTGCTGTGCACTACATA-3' | qRT-PCR of <i>RcNUDX1-1a</i> | 59°C |
| FP2 | 5'-CCACTCTTCTGGCCTTTGGA-3' | qRT-PCR of <i>RcNUDX1-1b</i> | 59°C |
| RP2 | 5'-CCTGCTGGGCCTATGAATCA-3' | qRT-PCR of <i>RcNUDX1-1b</i> | 59°C |
| FP3 | 5'-GTCGTCCCACTCGTACCAAC-3' | qRT-PCR of <i>RcNUDX1-2</i> | 60°C |
| RP3 | 5'-GATCTTTATGCGAGCGGTTTC-3' | qRT-PCR of <i>RcNUDX1-2</i> | 60°C |
| FP4 | 5'-GCTGTTGATGATCTGCCAAG-3' | qRT-PCR of <i>RcNUDX1-3</i> | 60°C |
| RP4 | 5'-GAGCTTTGAGGAGTGTGCAA-3' | qRT-PCR of <i>RcNUDX1-3</i> | 60°C |
| FP5 | 5'-GGGTAAGGAGAAGGTTACATC-3' | qRT-PCR of <i>EF1</i> (reference gene) | 64°C |
| RP5 | 5'-CAGCCTCCTTCTCAAACCTCT-3' | qRT-PCR of <i>EF1</i> (reference gene) | 64°C |
| FP6 | 5'-ATTGAGCGTCCCACCTACAC-3' | qRT-PCR of <i>TUB</i> (reference gene) | 58°C |
| RP6 | 5'-AGCATGAAATGGATCCTTGG-3' | qRT-PCR of <i>TUB</i> (reference gene) | 58°C |
| FP7 | 5'-ATGGGAAACGAGACAGTAGTAGT-3' | gDNA and cDNA cloning of <i>NUDX1</i> homologs | 58°C |
| RP7 | 5'-TCATGTTGGAAAAGGGTTAAATCC-3' | gDNA and cDNA cloning of <i>NUDX1</i> homologs | 58°C |
| FP8 | 5'-AAGCCAAACCATCGCAGTAC-3' | qPCR and qRT-PCR of <i>NUDX1</i> homologs | 58°C |
| RP8 | 5'-GGAAGATTGTCCCACTCATACC-3' | qPCR and qRT-PCR of <i>NUDX1</i> homologs | 58°C |
| FP9 | 5'-TCTGAAGTGCGAACTGACAA-3' | Cloning of upstream region <i>NUDX1-1b</i> homologs | 60°C |
| RP9 | 5'-AGCAAATACGACGCAATTTT-3' | Cloning of upstream region <i>NUDX1-1b</i> homologs | 60°C |
| FP10 | 5'-CTGGGAAGGGAAAAGGTGGA-3' | Cloning of upstream region <i>NUDX1-1a</i> homologs | 60°C |
| RP10 | 5'-TAATGCTGGAATTAGTTGGCCA-3' | Cloning of upstream region <i>NUDX1-1a</i> homologs | 60°C |
| FP11 | 5'-GGCTTAA[U]TATTTGTAATAATTCGTAACGTGAC-3' | Cloning of 138pb in <i>NUDX1-1a</i> promoter | 60°C |
| FP12 | 5'-GGCTTAA[U]TAGGGCTGAGATTGTTTTGC-3' | Cloning of 316pb in <i>NUDX1-1a</i> promoter | 60°C |
| FP13 | 5'-GGCTTAA[U]GCCTCTTCCCGTCTACTC-3' | Cloning of 521pb in <i>NUDX1-1a</i> promoter | 60°C |
| FP14 | 5'-GGCTTAA[U]GGTGTATGCAACATGAATGG-3' | Cloning of 1100pb in <i>NUDX1-1a</i> promoter | 60°C |
| FP15 | 5'-GGCTTAA[U]GCCATTTGGGAAGAAAATTA-3' | Cloning of 1500 pb in <i>NUDX1-1b</i> promoter | 60°C |
| RP11 | 5'-GGTTTAA[U]CTGTCTCGTTTCCCTTGCTT-3' | Reverse primer for FP11 to FP15 primers | 60°C |
| FP16 | 5'-GGCTTAA[U]CGGTTTGCATTTGGCTAGA-3' | Cloning dCAMV 35S promoter | 60°C |
| RP16 | 5'-GGTTTAA[U]GTAGAGAGAGACTGGTGATTTC-3' | Cloning dCAMV 35S promoter | 60°C |
| FP17 | 5'-AAAGGTACCATGGGAAACGAGACAGTAGTAGT-3' | Cloning of <i>RcNUDX1-1</i> into pET-30a(+) with <i>KpnI</i> | 58°C |
| RP17 | 5'-AAAGTCGACTCATGTTGAAAAGGGTTAAATCC-3' | Cloning of <i>RcNUDX1-1</i> into pET-30a(+) with <i>Sall</i> | 58°C |

Table S2.9 | Blast and alignment analysis of the gDNA and cDNA sequences cloned in wild and heritage roses ^a.

| Accession | gDNA number | Blast analysis | cDNA alignment | Remark |
|----------------|-------------|-------------------|----------------|--|
| Arvensis_A | gDNA1 | <i>RcNUDX1-1a</i> | None | cDNA1 did not correspond to any gDNA |
| | gDNA2 | <i>RcNUDX1-1a</i> | cDNA2 | |
| | gDNA3 | <i>RcNUDX1-1b</i> | None | |
| | gDNA4 | both | None | |
| | gDNA 5 | <i>RcNUDX1-1a</i> | None | |
| Arvensis_B | gDNA1 | <i>RcNUDX1-1b</i> | | no cDNA |
| | gDNA2 | <i>RcNUDX1-1b</i> | | |
| Autumn_Damask | gDNA1 | <i>RcNUDX1-1a</i> | cDNA3 | cDNA1 and cDNA2 did not correspond to any gDNA |
| | gDNA2 | Both | none | |
| | gDNA3 | <i>RcNUDX1-1b</i> | none | |
| | gDNA4 | <i>RcNUDX1-1a</i> | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | <i>RcNUDX1-1a</i> | none | |
| | gDNA7 | <i>RcNUDX1-1a</i> | cDNA3 | |
| Banksiae | gDNA1 | both | | no cDNA |
| | gDNA2 | both | | |
| | gDNA3 | both | | |
| | gDNA4 | both | | |
| Bracteata | gDNA1 | both | | no cDNA |
| Canina | gDNA1 | <i>RcNUDX1-1a</i> | none | |
| | gDNA2 | <i>RcNUDX1-1a</i> | none | |
| | gDNA3 | <i>RcNUDX1-1a</i> | none | |
| | gDNA4 | <i>RcNUDX1-1a</i> | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | <i>RcNUDX1-1b</i> | none | |
| | gDNA7 | <i>RcNUDX1-1a</i> | none | |
| | gDNA8 | <i>RcNUDX1-1b</i> | none | |
| | gDNA9 | <i>RcNUDX1-1b</i> | none | |
| | gDNA10 | <i>RcNUDX1-1a</i> | none | |
| | gDNA11 | <i>RcNUDX1-1b</i> | none | |
| | gDNA12 | <i>RcNUDX1-1a</i> | cDNA | |
| | gDNA13 | <i>RcNUDX1-1a</i> | none | |
| | gDNA14 | <i>RcNUDX1-1a</i> | cDNA | |
| | gDNA15 | <i>RcNUDX1-1a</i> | none | |
| Carolina | gDNA1 | <i>RcNUDX1-1a</i> | none | cDNA1 and cDNA3 did not correspond to any gDNA |
| | gDNA2 | <i>RcNUDX1-1a</i> | cDNA2 | |
| | gDNA3 | both | none | |
| | gDNA4 | both | none | |
| | gDNA5 | <i>RcNUDX1-1a</i> | none | |
| | gDNA6 | <i>RcNUDX1-1a</i> | none | |
| Centifolia | gDNA1 | <i>RcNUDX1-1a</i> | none | cDNA1 and cDNA3 did not correspond to any gDNA |
| | gDNA2 | <i>RcNUDX1-1a</i> | none | |
| | gDNA3 | <i>RcNUDX1-1a</i> | cDNA2 | |
| | gDNA4 | <i>RcNUDX1-1b</i> | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | <i>RcNUDX1-1a</i> | none | |
| | gDNA7 | <i>RcNUDX1-1a</i> | none | |
| | gDNA8 | <i>RcNUDX1-1a</i> | none | |
| | gDNA9 | <i>RcNUDX1-1a</i> | cDNA2 | |
| | gDNA10 | <i>RcNUDX1-1a</i> | none | |
| | gDNA11 | <i>RcNUDX1-1a</i> | none | |
| | gDNA12 | <i>RcNUDX1-1a</i> | none | |
| | gDNA13 | <i>RcNUDX1-1a</i> | none | |
| | gDNA14 | <i>RcNUDX1-1a</i> | none | |
| Chinensis | gDNA1 | <i>RcNUDX1-1b</i> | | no cDNA |
| | gDNA2 | <i>RcNUDX1-1b</i> | | |
| Damas_Kazanlik | gDNA1 | <i>RcNUDX1-1a</i> | none | cDNA1, 2, 3 and 4 did not correspond to any gDNA |
| | gDNA2 | <i>RcNUDX1-1b</i> | none | |
| | gDNA3 | <i>RcNUDX1-1a</i> | none | |
| | gDNA4 | <i>RcNUDX1-1a</i> | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | <i>RcNUDX1-1b</i> | none | |
| | gDNA7 | <i>RcNUDX1-1a</i> | none | |
| | gDNA8 | <i>RcNUDX1-1b</i> | none | |
| | gDNA9 | <i>RcNUDX1-1a</i> | none | |
| Ecae | gDNA1 | <i>RcNUDX1-1b</i> | none | cDNA1 and 2 did not correspond to any gDNA |
| | gDNA2 | <i>RcNUDX1-1b</i> | none | |
| | gDNA3 | <i>RcNUDX1-1b</i> | none | |
| Fedtschenkoana | gDNA1 | <i>RcNUDX1-1b</i> | none | |
| | gDNA2 | <i>RcNUDX1-1b</i> | none | |
| | gDNA3 | <i>RcNUDX1-1a</i> | none | |
| | gDNA4 | <i>RcNUDX1-1a</i> | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | <i>RcNUDX1-1a</i> | cDNA | |

Appendix 2

| | | | | |
|----------------|---|--|---|--|
| Foetida | gDNA1 gDNA2 gDNA3 | <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> both | none cDNA2 none | cDNA1 did not correspond to any gDNA |
| Gallica_A | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 gDNA6 | <i>RcNUDX1-1a</i> both <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> both <i>RcNUDX1-1b</i> | none none none cDNA2 none none | cDNA1 and 3 did not correspond to any gDNA |
| Gallica_B | gDNA1 gDNA2 gDNA3 | both <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> | none cDNA none | |
| Gigantea | gDNA1 gDNA2 | <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> | | no cDNA |
| Hugonis_A | gDNA1 gDNA2 gDNA3 gDNA4 | <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> | none none none none | cDNA was not searched for |
| Hugonis_B | gDNA1 gDNA2 | <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> | none none | cDNA was not searched for |
| Laevigata | gDNA1 | both | none | no cDNA |
| Majalis | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 gDNA6 gDNA7 | <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> both <i>RcNUDX1-1a</i> | none none cDNA2 none none none none | cDNA1 did not correspond to any gDNA |
| Mirifica | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 | <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> | | no cDNA |
| Moschata | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 gDNA6 gDNA7 gDNA8 gDNA9 gDNA10 gDNA11 | <i>RcNUDX1-1a</i> both <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> both both <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> | none none none none none none none none none cDNA1 none | cDNA2 and 3 did not correspond to any gDNA |
| Nutkana | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 gDNA6 | <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> both both <i>RcNUDX1-1a</i> | cDNA2 none none none none none | cDNA1 did not correspond to any gDNA |
| Officinalis | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 | <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> both | cDNA none cDNA none none | |
| Old Blush | gDNA1 gDNA2 | <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> | cDNA none | |
| Pendulina | gDNA1 gDNA2 gDNA3 | <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> | none none none | cDNA1 and 2 did not correspond to any gDNA |
| Persian Yellow | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 gDNA6 | <i>RcNUDX1-1a</i> both <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> | none none cDNA none none none | |
| Phoenicia | gDNA1 gDNA2 gDNA3 gDNA4 gDNA5 gDNA6 gDNA7 | <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1a</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1a</i> | none none none cDNA none none none | |
| Roxburghii | gDNA1 gDNA2 gDNA3 gDNA4 | both <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> <i>RcNUDX1-1b</i> | | no cDNA |

Appendix 2

| | | | | |
|--------------|--------|-------------------|-------|---|
| Rubus | gDNA1 | both | none | cDNA1 , 2, 3 and 4 did not correspond to any gDNA cDNA1 has a STOP codon |
| | gDNA2 | <i>RcNUDX1-1b</i> | none | |
| | gDNA3 | <i>RcNUDX1-1b</i> | none | |
| | gDNA4 | both | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | both | none | |
| Rugosa | gDNA1 | <i>RcNUDX1-1b</i> | none | cDNA1 did not correspond to any gDNA |
| | gDNA2 | <i>RcNUDX1-1a</i> | none | |
| | gDNA3 | <i>RcNUDX1-1a</i> | none | |
| | gDNA4 | <i>RcNUDX1-1a</i> | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | <i>RcNUDX1-1a</i> | none | |
| | gDNA7 | <i>RcNUDX1-1b</i> | none | |
| | gDNA8 | <i>RcNUDX1-1b</i> | none | |
| | gDNA9 | both | none | |
| | gDNA10 | <i>RcNUDX1-1a</i> | none | |
| | gDNA11 | <i>RcNUDX1-1b</i> | none | |
| | gDNA12 | <i>RcNUDX1-1a</i> | none | |
| Sericea | gDNA1 | <i>RcNUDX1-1b</i> | | no cDNA |
| | gDNA2 | <i>RcNUDX1-1b</i> | | |
| | gDNA3 | <i>RcNUDX1-1b</i> | | |
| | gDNA4 | <i>RcNUDX1-1b</i> | | |
| | gDNA5 | <i>RcNUDX1-1b</i> | | |
| | gDNA6 | <i>RcNUDX1-1b</i> | | |
| Spinosissima | gDNA1 | both | none | |
| | gDNA2 | both | none | |
| | gDNA3 | <i>RcNUDX1-1a</i> | none | |
| | gDNA4 | both | none | |
| | gDNA5 | <i>RcNUDX1-1b</i> | none | |
| | gDNA6 | <i>RcNUDX1-1b</i> | none | |
| | gDNA7 | <i>RcNUDX1-1b</i> | none | |
| | gDNA8 | <i>RcNUDX1-1a</i> | none | |
| | gDNA9 | <i>RcNUDX1-1a</i> | cDNA | |
| | gDNA10 | <i>RcNUDX1-1a</i> | none | |
| | gDNA11 | <i>RcNUDX1-1a</i> | none | |
| | gDNA12 | <i>RcNUDX1-1b</i> | none | |
| | gDNA13 | <i>RcNUDX1-1b</i> | none | |
| | gDNA14 | <i>RcNUDX1-1a</i> | none | |
| Villosa | gDNA1 | both | none | |
| | gDNA2 | <i>RcNUDX1-1a</i> | cDNA | |
| | gDNA3 | <i>RcNUDX1-1a</i> | none | |
| | gDNA4 | both | none | |
| | gDNA5 | <i>RcNUDX1-1a</i> | none | |
| Woodsii | gDNA1 | <i>RcNUDX1-1a</i> | | cDNA2 did not correspond to any gDNA |
| | gDNA2 | <i>RcNUDX1-1a</i> | | |
| | gDNA3 | <i>RcNUDX1-1a</i> | | |
| | gDNA4 | <i>RcNUDX1-1a</i> | | |
| | gDNA5 | <i>RcNUDX1-1a</i> | cDNA1 | |

^agDNA and cDNA in Clones_gDNAs_cDNAs.fasta and Clones_IntronExonStructure.fasta. Primers FP7-RP7 were used for cloning (Table S7).

Table S2.10 | qRT-PCR on NUDX1-1 homologs (a.u.).

| Accession | Average between technical replicates for each biological individual ^a | | | | | | Standard deviation on technical replicates | | | | | | Standard deviation on biological replicates | | |
|-----------------|--|--------------|--------------|--------------|--------------|--------------|--|--------------|--------------|--------------|--------------|--------------|---|--------------|--------------------|
| | Individual 1 | Individual 2 | Individual 3 | Individual 4 | Individual 5 | Individual 6 | Average | Individual 1 | Individual 2 | Individual 3 | Individual 4 | Individual 5 | Individual 6 | Individual 6 | Standard deviation |
| Aivensis_A | 354.2317 | 148.6937 | 341.7287 | 441.6900 | 153.6794 | 100.5113 | 256.7558 | 54.3835 | 25.1580 | 24.3327 | 19.9252 | 35.6275 | 23.5599 | | 139.7405 |
| Aivensis_B | 0.1690 | 0.0090 | | | | | 0.0890 | 0.1714 | 0.0078 | | | | | | 0.1131 |
| Autumn_Damask | 49.7938 | 76.3068 | | | | | 63.0503 | 7.2542 | 4.6647 | | | | | | 18.7475 |
| Banksiae | 0.0020 | 0.0038 | 0.0111 | | | | 0.0056 | 0.0013 | 0.0016 | 0.0031 | | | | | 0.0048 |
| Bracteata | 0.0019 | 0.0001 | | | | | 0.0010 | 0.0001 | 0.0001 | | | | | | 0.0012 |
| Canina | 110.3493 | 112.6709 | | | | | 111.5101 | 18.2037 | 30.8590 | | | | | | 1.6416 |
| Carolina | 410.5635 | 332.6452 | 415.0480 | 199.4842 | | | 339.4352 | 235.7078 | 212.4561 | 111.1834 | 93.1146 | | | | 100.6792 |
| Camifolia | 113.8567 | 300.3545 | | | | | 207.1056 | 37.1950 | 57.3606 | | | | | | 131.8738 |
| Chinensis | 0.0021 | 0.0003 | 0.0014 | 0.0001 | | | 0.0009 | 0.0003 | 0.0001 | 0.0003 | 0.0001 | | | | 0.0009 |
| Damask_Kazanlik | 42.0393 | 44.2255 | | | | | 43.1324 | 4.5818 | 0.9934 | | | | | | 1.5458 |
| Ecae | 0.0064 | 0.0057 | 0.0091 | 0.0223 | 0.0785 | 0.0800 | 0.0337 | 0.0042 | 0.0026 | 0.0086 | 0.0196 | 0.0201 | 0.0460 | | 0.0358 |
| Fedischenkoana | 88.0061 | 83.5712 | 90.0338 | | | | 87.2037 | 34.4741 | 29.6235 | 27.1324 | | | | | 3.3052 |
| Foetida | 22.2202 | 4.4365 | | | | | 13.3284 | 5.7478 | 0.4873 | | | | | | 12.5749 |
| Gallica_A | 103.61851 | 72.8843 | | | | | 88.2514 | 14.8342 | 5.6948 | | | | | | 21.7323 |
| Gallica_B | 105.7963 | 77.1806 | | | | | 91.4885 | 5.5422 | 4.7901 | | | | | | 20.2343 |
| Gigantea | 0.1042 | 0.0134 | 0.0260 | | | | 0.0166 | 0.0061 | 0.0066 | 0.0157 | | | | | 0.0021 |
| Hugonis_A | 12.6915 | 12.2668 | | | | | 12.4792 | 1.0898 | 2.4970 | | | | | | 0.3003 |
| Hugonis_B | 0.0215 | 0.0019 | | | | | 0.0117 | 0.0060 | 0.0006 | | | | | | 0.0138 |
| Laevigata | 0.0013 | 0.0019 | | | | | 0.0016 | 0.0004 | 0.0007 | | | | | | 0.0004 |
| Majalis | 26.5795 | 23.9745 | | | | | 25.2770 | 5.7974 | 3.5694 | | | | | | 1.8419 |
| Mirifica | 0.0001 | 0.0001 | | | | | 0.0001 | 0.0001 | 0.0001 | | | | | | 0.0001 |
| Moschata | 105.7031 | 113.8273 | 115.1166 | | | | 111.5490 | 42.7139 | 56.1093 | 23.9258 | | | | | 5.7446 |
| Nuttkana | 463.6421 | 280.9720 | 377.5620 | | | | 374.0687 | 83.4758 | 34.2111 | 85.2870 | | | | | 129.1672 |
| Officinalis | 118.3519 | 106.6708 | | | | | 112.5114 | 58.1246 | 51.9502 | | | | | | 8.2598 |
| Old_Blush | 68.4644 | 63.8831 | 53.1969 | 74.7161 | 44.7050 | | 60.9931 | 29.2368 | 4.5113 | 3.2522 | 7.0509 | 8.3188 | | | 12.0268 |
| Pendulina | 116.2249 | 233.4340 | | | | | 174.8294 | 28.0713 | 15.9876 | | | | | | 82.8793 |
| Persian_Yellow | 56.2920 | 13.4222 | | | | | 34.8271 | 22.4901 | 3.9386 | | | | | | 30.2711 |
| Phoenicia | 163.7674 | 146.5574 | | | | | 155.1724 | 39.7363 | 24.8994 | | | | | | 12.1834 |
| Roxburghii | 0.0032 | 0.0255 | | | | | 0.0143 | 0.0001 | 0.0077 | | | | | | 0.0157 |
| Rubus | 0.6419 | 0.6965 | | | | | 0.6692 | 0.0666 | 0.1376 | | | | | | 0.0385 |
| Rugosa | 53.0213 | 19.1390 | | | | | 36.0802 | 31.7928 | 1.9633 | | | | | | 23.9584 |
| Sencea | 0.0033 | 0.0048 | | | | | 0.0041 | 0.0009 | 0.0027 | | | | | | 0.0011 |
| Villosa | 132.2808 | 123.7985 | | | | | 128.0396 | 11.9713 | 15.2043 | | | | | | 5.9979 |
| Woodsii | 17.8098 | 16.7036 | 22.1107 | 22.5776 | | | 19.8005 | 2.9275 | 1.1745 | 0.4314 | 1.2656 | | | | 2.9778 |

^a 2 to 4 technical replicates for each species. Primers FP8-RP8 were used for amplification and primers FP5-RP5 and FP6-RP6, for reference genes (Table S7).

Table S2.11 | Statistical analysis between volatile compound concentrations and NUDX1-1 expression^a

| Compound name | Spearman's correlation coefficient | P-values | Compound name | Spearman's correlation coefficient | P-values |
|-------------------------|------------------------------------|----------|-------------------------|------------------------------------|----------|
| (E)-2-hexenal | -0.0888 | 0.6172 | beta-eudesmol | 0.3278 | 0.0583 |
| (E)-2-hexenol | 0.3469 | 0.0443 | beta-myrcene | 0.6504 | 3.1252 |
| (E)-2-hexenylacetate | -0.1457 | 0.4109 | beta-pinene | -0.0041 | 0.9815 |
| (E)-2-nonenal | 0.3060 | 0.0783 | beta-sesquiphellandrene | 0.2217 | 0.2074 |
| (E)-alpha-bisabolene | 0.2217 | 0.2074 | borneol | -0.2750 | 0.1154 |
| (E)-beta-caryophyllene | 0.1939 | 0.2717 | caryophyllene oxide | -0.1850 | 0.2947 |
| (E)-beta-farnesene | 0.5220 | 0.0015 | decanal | -0.2711 | 0.1208 |
| (E)-beta-ionone | -0.0086 | 0.9614 | decane | -0.1501 | 0.3966 |
| (E)-beta-ocimene | 0.5369 | 0.0010 | dihydro-beta-ionol | -0.2893 | 0.0969 |
| (E)-cinnamaldehyde | 0.0798 | 0.6535 | dihydro-beta-ionone | -0.1508 | 0.3945 |
| (E)-gamma-bisabolene | 0.2217 | 0.2074 | dodecanal | -0.2013 | 0.2534 |
| (E)-isoeugenol | -0.1330 | 0.4530 | dodecane | -0.0438 | 0.8056 |
| (E)-methylisoeugenol | -0.2356 | 0.1797 | elemicin | -0.2356 | 0.1797 |
| (E,E)-alpha-farnesene | 0.5956 | 0.0002 | elemol | 0.3126 | 0.0717 |
| (E,E)-farnesal | 0.4189 | 0.0136 | eugenol | -0.2035 | 0.2483 |
| (E,E)-farnesol | 0.6876 | 7.0050 | gamma-elemene | 0.2750 | 0.1154 |
| (E,E)-farnesylacetate | 0.1302 | 0.4629 | gamma-eudesmol | 0.2449 | 0.1626 |
| (E,Z)-nonadienal | 0.3654 | 0.0335 | gamma-terpinene | 0.2958 | 0.0893 |
| (Z)-3-hexenal | -0.2457 | 0.1612 | geranial | 0.5637 | 0.0005 |
| (Z)-3-hexenol | 0.1001 | 0.5731 | geraniol | 0.7039 | 3.3973 |
| (Z)-3-hexenylacetate | -0.0797 | 0.6537 | geranylacetate | 0.6050 | 0.0001 |
| (Z)-alpha-bisabolene | 0.2217 | 0.2074 | germacrene D | -0.0512 | 0.7735 |
| (Z)-beta-elemenone | 0.2750 | 0.1154 | germacrone | 0.2750 | 0.1154 |
| (Z)-beta-ocimene | 0.3368 | 0.0513 | heptadecane | -0.0604 | 0.7341 |
| (Z,E)-alpha-farnesene | 0.5638 | 0.0005 | heptadecene | -0.0867 | 0.6258 |
| (Z,E)-farnesal | 0.3027 | 0.0817 | heptanal | 0.1950 | 0.2689 |
| (Z,E)-farnesol | 0.3087 | 0.0756 | hexadecane | -0.1820 | 0.3028 |
| 1-hexanol | 0.2843 | 0.1031 | hexanal | 0.0321 | 0.8565 |
| 1-phenylethanol | 0.0798 | 0.6535 | hexylacetate | -0.1599 | 0.3662 |
| 1,3,5-trimethoxybenzene | -0.2948 | 0.0905 | limonene | 0.3669 | 0.0327 |
| 1,8-cineole | 0.0798 | 0.6535 | linalool | -0.0577 | 0.7454 |
| 2-hexanol | 0.1941 | 0.2713 | methyleugenol | -0.3282 | 0.0580 |
| 2-hexanone | -0.0975 | 0.5829 | methylsalicylate | 0.0958 | 0.5895 |
| 2-pentadecanone | -0.5822 | 0.0003 | neral | 0.6057 | 0.0001 |
| 2-phenylethanol | 0.6984 | 4.3716 | nerol | 0.6750 | 1.1908 |
| 2-tridecanone | -0.4081 | 0.0165 | nerolidol | 0.3269 | 0.0591 |
| 2,3-dihydrofarnesol | 0.1700 | 0.3362 | nerylacetate | 0.4079 | 0.0166 |
| 2,4-decadienal | -0.1330 | 0.4530 | nonanal | -0.2073 | 0.2392 |
| 3-hexanol | 0.1419 | 0.4232 | nonane | -0.2233 | 0.2041 |
| 3-hexanone | -0.0975 | 0.5829 | nonanoic acid | -0.2899 | 0.0962 |
| 3,5-dimethoxyphenol | -0.2750 | 0.1154 | octadecane | -0.1918 | 0.2769 |
| 3,5-dimethoxytoluene | -0.3832 | 0.0252 | octadecene | -0.1197 | 0.4998 |
| 4-vinylphenol | -0.1153 | 0.5160 | octanal | -0.1652 | 0.3501 |
| allo-ocimene | 0.2420 | 0.1679 | octane | -0.1094 | 0.5376 |
| alpha-bisabolol | 0.2105 | 0.2320 | pentadecane | -0.2162 | 0.2192 |
| alpha-cadinol | -0.3082 | 0.0761 | phenylacetaldehyde | 0.1910 | 0.2792 |
| alpha-humulene | -0.0363 | 0.8383 | phenylethylacetate | 0.1113 | 0.5307 |
| alpha-phellandrene | 0.3737 | 0.0294 | phenylethylbenzoate | 0.3063 | 0.0780 |
| alpha-pinene | 0.2237 | 0.2033 | sabinene | 0.1153 | 0.5160 |
| alpha-terpinene | 0.2710 | 0.1210 | tetradecanal | 0.0824 | 0.6429 |
| alpha-terpinolene | 0.2705 | 0.1216 | theaspirane | -0.1508 | 0.3945 |
| benzaldehyde | 0.1614 | 0.3616 | tridecane | -0.1834 | 0.2989 |
| benzylacetate | -0.0831 | 0.6401 | undecanal | -0.2853 | 0.1018 |
| benzylalcohol | 0.4497 | 0.0076 | undecane | -0.1575 | 0.3736 |
| benzylbenzoate | 0.0518 | 0.7708 | unknown 1 | 0.1895 | 0.2830 |
| beta-citronellal | 0.2786 | 0.1105 | unknown 2 | 0.2217 | 0.2074 |
| beta-citronellol | 0.6225 | 8.4461 | unknown 3 | 0.3157 | 0.0688 |
| beta-citronellylacetate | 0.0878 | 0.6212 | unknown 4 | 0.0975 | 0.5829 |
| beta-elemene | 0.2983 | 0.0865 | | | |

^a Statistics correspond to expressions in Table S10, and volatile concentrations in Table S8

Table S2.12 | List of selected sequences in intergenic regions in “Rosa chinensis Old Blush homozygous Genome v2.0 chromosomes”^a.

| Chromosome number | Identified gene and TE sequences in the GDR ^b | Identification and remarks | TE fragments: hit starts, hit ends (bp) ^c | Identity to consensus sequence (%) |
|---------------------------|---|--|--|------------------------------------|
| Chromosome 2 | RcHm_v2.0_Chr2g01421141 | Gene A used as an upstream marker for <i>Rosa</i> microsynteny | | |
| | RcHm_v2.0_Chr2g01421131 | Gene B used as an upstream marker for <i>Rosa</i> microsynteny | | |
| | ms382385 RcHm_v2.0_Chr2_RLC-chim_denovoRcHm_v2.0-B-R24588-Map5_reversed | <i>Copia</i> R24588 (226 bp) | 1905-2098 | 81.9 |
| | RcHm_v2.0_Chr2g0142121 | <i>NUDX1-1a</i> copy 1 - Exon 1 | 5124-5183 | 82.0 |
| | ms382217 RcHm_v2.0_Chr2_DTX-chim_denovoRcHm_v2.0-B-R29785-Map6 | Unknown | | |
| | RcHm_v2.0_Chr2g0142121 | <i>NUDX1-1a</i> copy 1 - Exon 2 | 5124-5183 | 82.0 |
| | ms3820616 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (MITE 352 bp) | 5-356 | 93.1 |
| | ms382250 RcHm_v2.0_Chr2_DXX-MITE_denovoRcHm_v2.0-B-G13554-Map6 | MITE G13554 (MITE 396 bp) | 2-398 | 91.5 |
| | ms382374 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (MITE 76 bp) | 349-424 | 85.3 |
| | ms382384 RcHm_v2.0_Chr2_RLC-chim_denovoRcHm_v2.0-B-R24588-Map5_reversed | <i>Copia</i> R24588 (226 bp) | 1905-2098 | 82.3 |
| | RcHm_v2.0_Chr2g0142111 | <i>NUDX1-1a</i> copy 2 - Exon 1 | 5124-5183 | 82.0 |
| | ms382217 RcHm_v2.0_Chr2_DTX-chim_denovoRcHm_v2.0-B-R29785-Map6 | Unknown | | |
| | RcHm_v2.0_Chr2g0142111 | <i>NUDX1-1a</i> copy 2 - Exon 2 | 5124-5183 | 82.0 |
| | ms382373 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (MITE 352 bp) | 5-356 | 93.1 |
| | ms382252 RcHm_v2.0_Chr2_DXX-MITE_denovoRcHm_v2.0-B-G14214-Map9 | MITE G13554 (MITE 163 bp) | 246-408 | |
| | RcHm_v2.0_Chr2g0142101 | [insert region] ncRNA | | |
| | RcHm_v2.0_Chr2g0142091 | [insert region] gene | | |
| | ms382375 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-R4060-Map16 | [insert region] unknown | 60-93 | 88.2 |
| | ms382390 RcHm_v2.0_Chr2_RLG_denovoRcHm_v2.0-B-R3289-Map8 | [insert region] Gypsy | 2291-3512 | 93.2 |
| | ms382252 RcHm_v2.0_Chr2_DXX-MITE_denovoRcHm_v2.0-B-G14214-Map9 | MITE G13554 (MITE 245 bp) | 5-250 | 90.6 |
| | ms382372 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (MITE 72 bp) | 11-77 (reverse hit == 352-418) | 87.8 |
| | ms382383 RcHm_v2.0_Chr2_RLC-chim_denovoRcHm_v2.0-B-R24588-Map5_reversed | <i>Copia</i> R24588 (226 bp) | 1905-2098 | 81.9 |
| | RcHm_v2.0_Chr2g0142081 | <i>ReNUDX1-1a</i> copy 3 - Exon 1 | 5124-5183 | 82.0 |
| | DTX-chim_denovoRcHm_v2.0-B-R29785-Map6 | Unknown | | |
| | RcHm_v2.0_Chr2g0142081 | <i>ReNUDX1-1a</i> copy 3 - Exon 2 | 5124-5183 | 82.0 |
| | ms380694 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (356 bp) | 5-356 | 93.4 |
| | ms382245 RcHm_v2.0_Chr2_DXX-MITE_denovoRcHm_v2.0-B-G13554-Map6 | MITE G13554 (MITE 396 bp) | 2-398 | 91.2 |
| | ms382371 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (76 bp) | 349-424 | 86.7 |
| | ms382382 RcHm_v2.0_Chr2_RLC-chim_denovoRcHm_v2.0-B-R24588-Map5_reversed | <i>Copia</i> R24588 (226 bp) | 1905-2098 | 81.8 |
| | RcHm_v2.0_Chr2g0142071 | <i>ReNUDX1-1a</i> copy 4 - Exon 1 | 5124-5183 | 82.0 |
| | DTX-chim_denovoRcHm_v2.0-B-R29785-Map6 | Unknown | | |
| | RcHm_v2.0_Chr2g0142071 | <i>ReNUDX1-1a</i> copy 4 - Exon 2 | 5124-5183 | 82.0 |
| | ms380693 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (356 bp) | 5-356 | 93.4 |
| | ms382245 RcHm_v2.0_Chr2_DXX-MITE_denovoRcHm_v2.0-B-G13554-Map6 | MITE G13554 (MITE 396 bp) | 2-398 | 91.2 |
| | ms382370 RcHm_v2.0_Chr2_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (MITE 76 bp) | 349-424 | 86.7 |
| | ms382381 RcHm_v2.0_Chr2_RLC-chim_denovoRcHm_v2.0-B-R24588-Map5_reversed | <i>Copia</i> R24588 (226 bp) | 1905-2098 | 82.4 |
| | RcHm_v2.0_Chr2g0142061 | <i>ReNUDX1-1a</i> copy 5 (partial) | 5124-5183 | 82.0 |
| | DTX-chim_denovoRcHm_v2.0-B-R29785-Map6 | Unknown | | |
| | RcHm_v2.0_Chr2g0142061 | <i>ReNUDX1-1a</i> copy 5 (partial) | 5124-5183 | 82.0 |
| | RcHm_v2.0_Chr2g0142051 | <i>ReNUDX1-1a</i> copy 5 (partial) | 5124-5183 | 82.0 |
| RcHm_v2.0_Chr2g0142041 | Gene C used as an upstream marker for <i>Rosa</i> microsynteny | | | |
| RcHm_v2.0_Chr2g0142011 | Gene D used as an upstream marker for <i>Rosa</i> microsynteny | | | |
| Chromosome 4 ^e | RcHm_v2.0_Chr4g0436231 | Gene E used as an upstream marker for <i>Rosa</i> microsynteny | | |
| | RcHm_v2.0_Chr4g0436201 | Gene F used as an upstream marker for <i>Rosa</i> microsynteny | | |
| | RcHm_v2.0_Chr4g0436191 | <i>ReNUDX1-3</i> | | |
| | RcHm_v2.0_Chr4g0436181 | <i>ReNUDX1-1b</i> - Exon 1 | | |
| | DTX-chim_denovoRcHm_v2.0-B-R29785-Map6 | Unknown | | |
| | RcHm_v2.0_Chr4g0436181 | <i>ReNUDX1-1b</i> - Exon 2 | 5124-5183 | 82.0 |
| | ms208216 RcHm_v2.0_Chr4_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (MITE 422 bp) | 5-421 | |
| | ms208214 RcHm_v2.0_Chr4_noCat_denovoRcHm_v2.0-B-P580.2030-Map20 | MITE P580.2030 (MITE 411 bp) | 16-424 | |
| | RcHm_v2.0_Chr4g0436151 | <i>ReNUDX1-2a</i> | | |
| | RcHm_v2.0_Chr4g0436141 | Gene G used as an upstream marker for <i>Rosa</i> microsynteny | | |
| RcHm_v2.0_Chr4g0436131 | Gene H used as an upstream marker for <i>Rosa</i> microsynteny | | | |

^a All sequences are numbered according to the GDR (www.rosaceae.org; Jung et al 2019, Raymond et al. 2018). Some TEs are not shown.

^b Exact position of the TE in the consensus sequence available in the GDR.

^c *Copia* R24588 is absent of this haplotype.

Table S2.13 | Characteristics of the sequences obtained by MinION sequencing of two wild species.

| Accession name | Run time (h) | Number of reads | Base sequenced (Gb) | NS50 (Kb) | Maximum length (Kb) | Number of sequences > 10 Kb | Accession number in SRA database | Reads used to check: |
|----------------|------------------|-----------------|---------------------|-----------|---------------------|-----------------------------|----------------------------------|--|
| Moschata | 25 [3, 173, 703] | 206,550 | 10.6 | 58.71 | 206,550 | 165,311 | PRJNA706580 | RRNUDX1-7b cluster 52016a8-7574-4387-804b-988103106219 af1055a5-7d34-4361-8b5b-df6f45188f4a 04d87998-2472-4090-84cc-966be7b6f5ac 8bfe1aa2-c7aa-4b7d-at2c-43024f91118fb 55aa7448-1b01-42a2-ad78-e78333c61e7c1 |
| Laevigata | 22 [3, 238, 007] | 225,254 | 8.9 | 45.78 | 225,254 | 125,979 | PRJNA706580 | RRNUDX1-1a cluster 0724eb00-6322-436c-d64a-665ad1d7eb60 f80149e4-7908-42be-86ab-c142894baec1 9631b369-f260-4c03-bdcb-bb80c90118d7 |

Table S2.14 | List of marker genes used for microsynteny maps.

| Code number ^a | Corresponding names in the GDR ^b |
|--------------------------|---|
| A | <i>RcHm_v2.0_Chr2g0142141, RC0G0033600, RcHt_S929.3, FvH4_6g32850, Prupe.3G053500, D17G1178400, MD09G1197500/7000</i> |
| B | <i>RcHm_v2.0_Chr2g0142131, RC0G0033700, RcHt_S929.2, FvH4_6g32840, Prupe.3G053600, D17G1178500, MD09G1197600</i> |
| C | <i>RcHm_v2.0_Chr2g0142041, RC0G0034500, RcHt_S929.1, FvH4_6g32830, Prupe.3G053700, D17G1178700, MD09G1197700</i> |
| D | <i>RcHm_v2.0_Chr2g0142011, RC0G0034600, FvH4_6g32820, Prupe.3G053800, MD17G1178800, MD09G1197800</i> |
| E | <i>RcHm_v2.0_Chr4g0436231, RC4G0402600, RcHt_S1291.19, RcHt_S866.15, FvH4_4g28360, Prupe.1G302500, MD13G1049400, MD16G1050400</i> |
| F | <i>RcHm_v2.0_Chr4g0436201, RC4G0402500, RcHt_S1291.17, RcHt_S866.16, FvH4_4g28350, Prupe.1G302700, MD13G1049200, MD16G1050300</i> |
| G | <i>RcHm_v2.0_Chr4g0436141, RC4G0401800, RcHt_S1291.10, RcHt_S866.22</i> |
| H | <i>RcHm_v2.0_Chr4g0436131, RC4G0401700</i> |
| J | <i>RcHm_v2.0_Chr6g0244121, RC6G0027100</i> |
| K | <i>RcHm_v2.0_Chr6g0244171</i> |
| L | <i>RcHm_v2.0_Chr6g0244181, RC6G0027600</i> |
| M | <i>RcHm_v2.0_Chr7g0219431, RC7G0379800, RcHt_S2031.1, RcHt_S317.21</i> |
| N | <i>RcHm_v2.0_Chr7g0219451, RC7G0379900, RcHt_S2031.2, RcHt_S317.20</i> |
| O | <i>RcHm_v2.0_Chr7g0219471, RC7G0380000, RcHt_S2031.4, RcHt_S317.19</i> |
| P | <i>RcHm_v2.0_Chr7g0219501, RC7G0380200, RcHt_S2031.6, RcHt_S317.16</i> |
| Q | <i>RcHm_v2.0_Chr4g0436031, RC4G0401100, FvH4_4g28230, Prupe.1G303900, MD13G1048200/300</i> |
| R | <i>RcHm_v2.0_Chr2g0142211, FvH4_6g32870, Prupe.3G53300, MD17G1178300, MD09G1196800</i> |
| S | <i>RcHm_v2.0_Chr2g0142991, FvH4_6g32800, Prupe.3G054000, MD17G1179000, MD09G1197900</i> |

^a These codes are used in Figs. 3,4,6.

^b All sequences are numbered according to the GDR www.rosaceae.org (Jung et al 2019).

Appendix 3 |

Supplementary materials for chapter 3

Appendix 3

| Control experiments | Crude extract | Cytosolic fraction | Plastidic fraction | Mitochondrial and peroxizomal fraction |
|--|----------------------|---------------------------|---------------------------|---|
| (Protein recovery as % to total) | 100.0 | (82.4 ± 1.4) | (0.4 ± 0.1) | (7.9 ± 0.5) |
| ADH activity (Cytosol) ¹ | 17.2 ± 7.3 | 39.1 ± 1.0 | ND | 1.8 ± 0.3 |
| Fumarase activity (Mitochondria) ¹ | 2.6 ± 0.4 | 1.2 ± 0.4 | 3.3 ± 0.1 | 17.9 ± 0.4 |
| Catalase activity (Peroxisome) ¹ | 93.8 ± 17.6 | 20.8 ± 9.8 | 64.2 ± 9.0 | 126.9 ± 5.2 |
| Chlorophyll measurements (Plastids) ² | 0.2 ± 0.1 | 0.2 ± 0.1 | 3.4 ± 1.3 | 0.3 ± 0.1 |

Figure S3.1 | Subcellular fractionation of rose petal crude extracts. Activities of marker enzymes and chlorophyll content measurement. Marker enzyme activities alcohol dehydrogenase (ADH), fumarase and catalase were assayed in cytosolic, plastidic and mitochondrial/peroxizomal fractions compared to crude extract. Data are means ± SEM, n = 3. ¹Specific activities in pKat mg⁻¹ of proteins. ²Chlorophyll content in each fraction as relative fluorescence units. ND, not detectable.

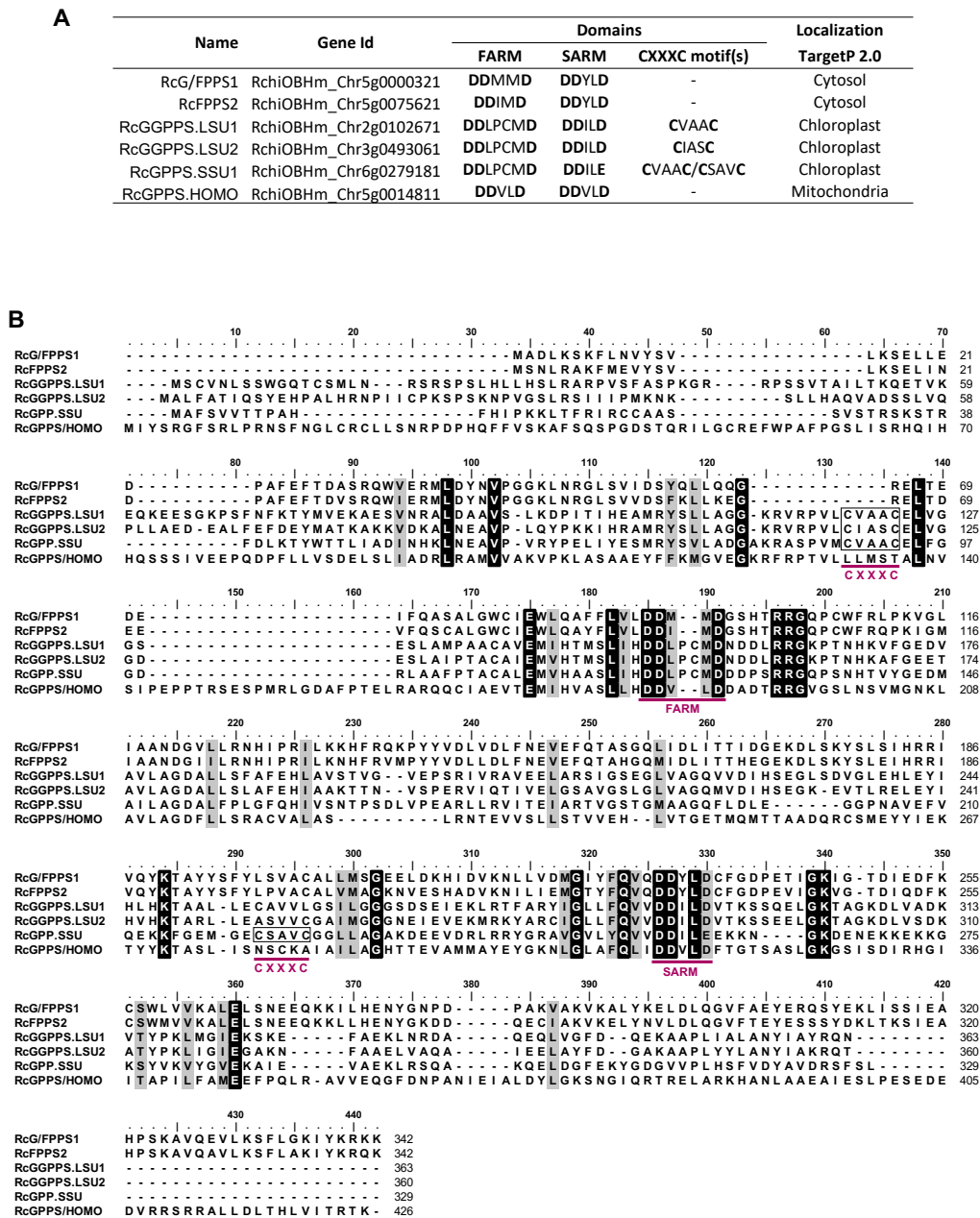


Fig. S2. The *trans*-short-chain IDSs member in RcOB and their characteristics.

(A) The list of *IDSs* present in the genome of RcOB, their names, gene identity, sequences of their conserved first aspartate motif (FARM), second aspartate motif (SARM) and CXXXC motifs and their predicted subcellular localization analyzed with TargetP 2.0. (B) ClustalW amino acid alignment of the RcOB IDS. Identical and similar amino acids are shown in black and grey, respectively. Conserved FARM, SARM and CXXXC motifs are highlighted.

Figure S3.2 | The *trans*-short-chain IDSs member in OB and their characteristics. (A) The list of *IDSs* present in the genome of OB, their names, gene identity, sequences of their conserved first aspartate motif (FARM), second aspartate motif (SARM) and CXXXC motifs and their predicted subcellular localization analyzed with TargetP 2.0. (B) ClustalW amino acid alignment of the OB IDS. Identical and similar amino acids are shown in black and grey, respectively. Conserved FARM, SARM and CXXXC motifs are highlighted.

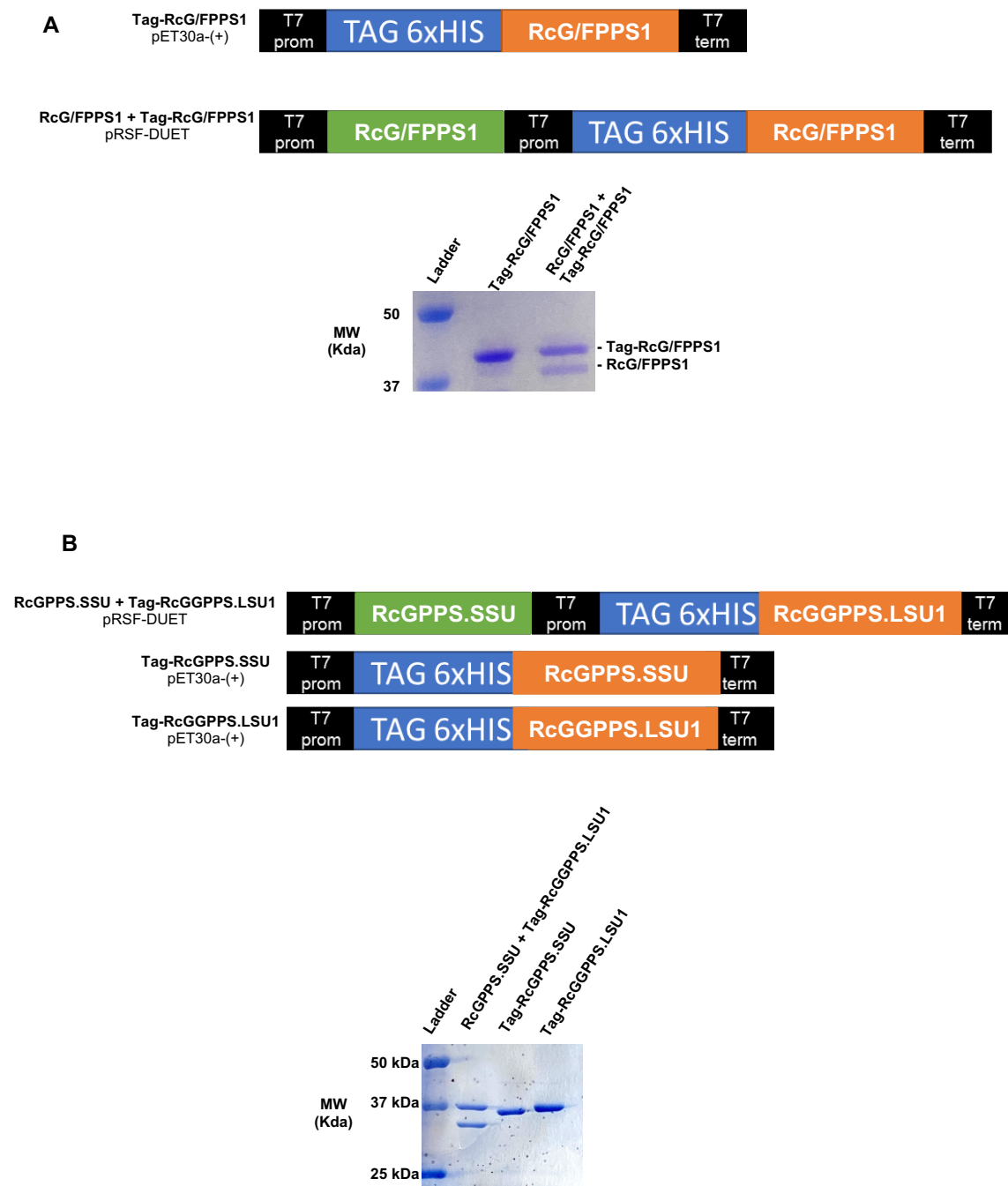


Figure S3.3 | RcG/FPPS1 is a homodimer and RcGPPS.SSU forms a heterodimer with RcGGPPS.LSU1. (A) Diagram on top shows the constructs used for purification of 6-HIS-Tagged RcG/FPPS1 alone and copurification of 6-HIS-Tagged Rc/GFPPS1 with untagged Rc/GFPPS1. SDS-PAGE gel of the corresponding purifications is presented on the bottom. (B) Diagram on top shows the constructs used for purification of 6-HIS-Tagged RcGGPPS.LSU1 or 6-HIS-Tagged RcGPPS.SSU alone and copurification of 6-HIS-Tagged RcGGPPS.LSU1 with untagged RcGPPS.SSU. SDS-PAGE gel of the corresponding purifications is presented on the bottom. MW = molecular weight standards.

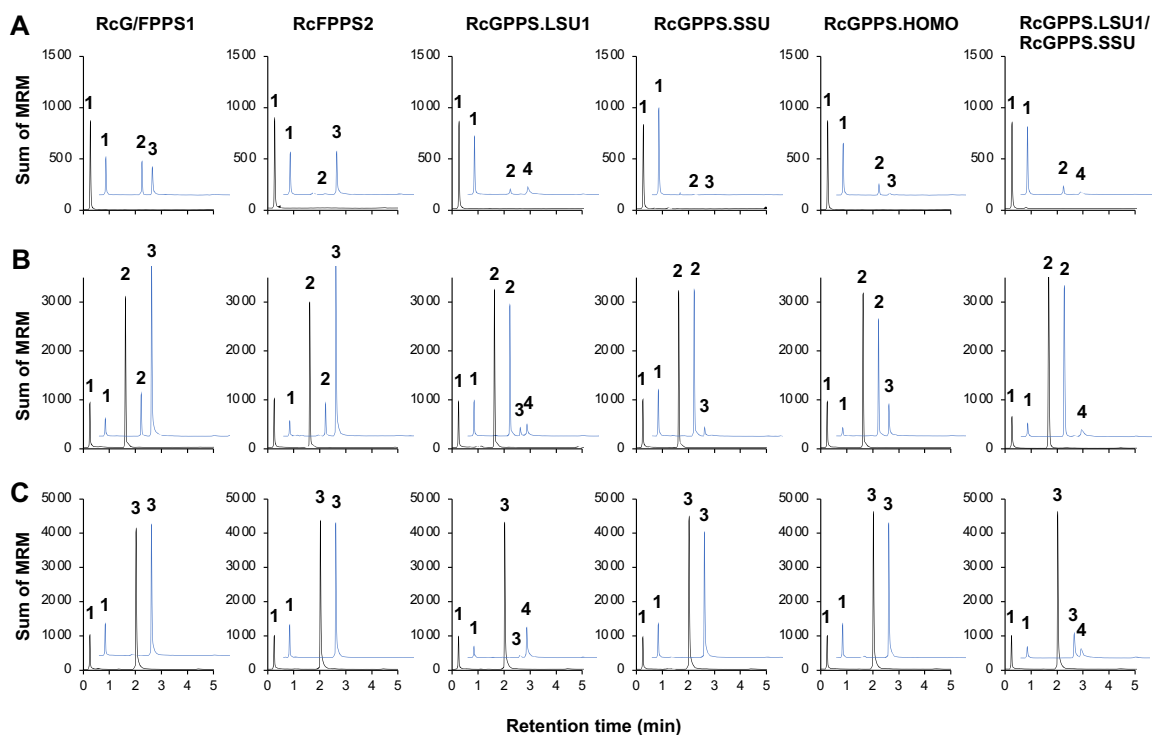


Figure S3.4 | Enzymatic characterization of the five rose IDS candidates. LC-MS/MS chromatograms of the reaction products from *in vitro* incubation of the five OB IDS candidates and the heterodimeric GPPS (RcGPPS.LSU1/RcGPPS.SSU) with (A) IPP and DMAPP, (B) IPP and GPP and (C) IPP and FPP. Chromatograms in blue represent the incubation of 5 μ g of each protein for 20 minutes with 10 μ M of substrates at 30 $^{\circ}$ C. Chromatograms in black represent the negative controls with incubation of the corresponding boiled proteins. 1 = DMAPP + IPP, 2 = GPP, 3 = FPP, 4 = GGPP.

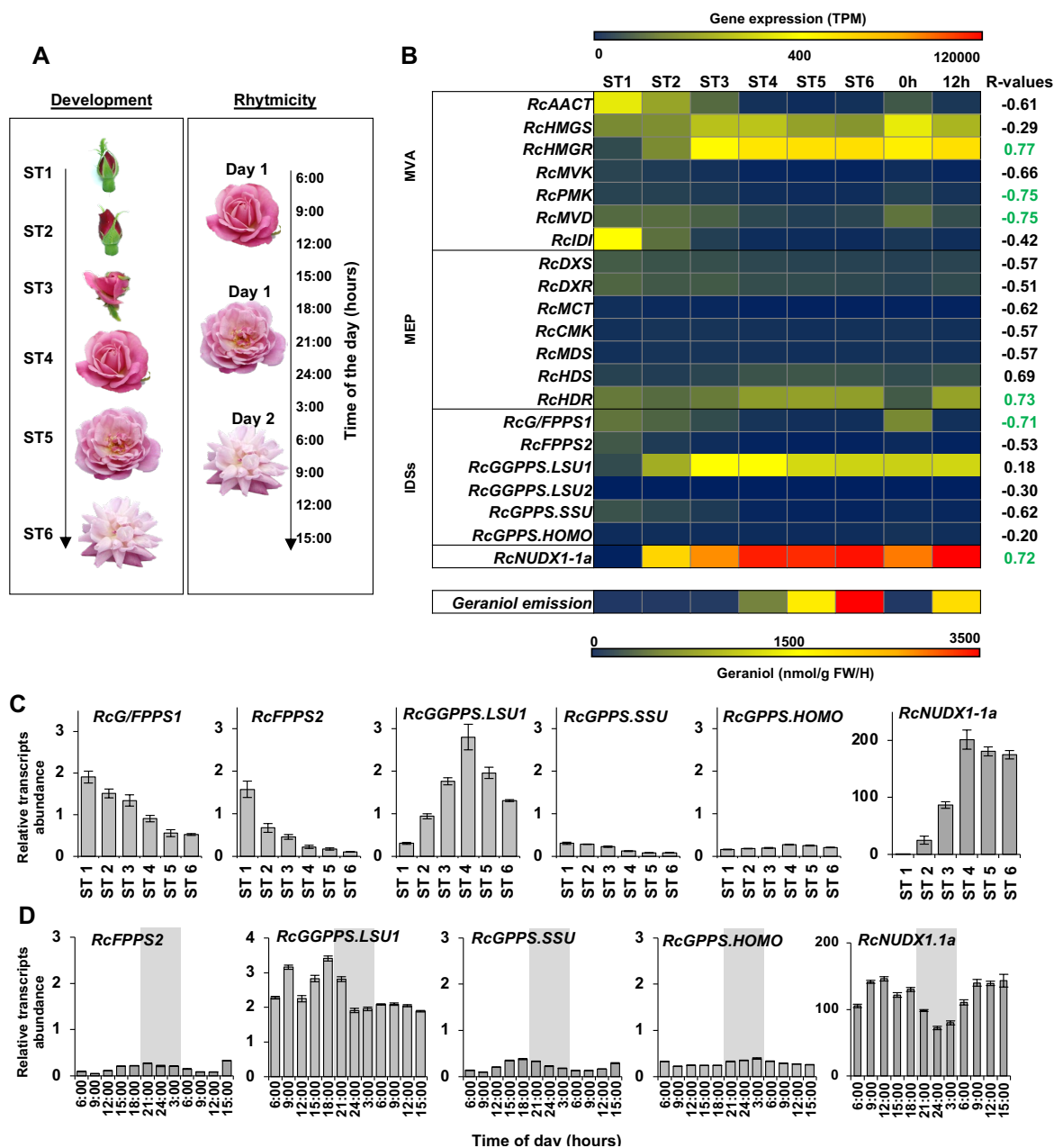


Figure S3.5 | Transcriptomic analysis of and metabolite profiling of OB genes during flower development and rhythmicity. (A) Pictures showing the six developmental stages (ST1 to 6) of OB flowers used for this analysis on left panel and development stages and times of collection for rhythmicity analysis on right panel. (B) Transcriptomic analysis of the genes involved in the MVA and MEP pathways, the six OB *IDS* candidates and *RcNUDX1.1a*. Heatmap on top panel shows the expression levels of the selected genes expressed as transcripts per million (TPM) from RNAseq analysis during development of OB flowers as described in (A) and at two time points during a day/night cycle (12:00 h and 24:00 h) of stage 5 flowers. Data are means from 4 biological replicates. Heatmap on bottom panel represents the levels of geraniol emission analyzed from the same developmental stages and time points by GC-MS. Data are means from 5 biological replicates. Spearman correlation coefficients (R-

Appendix 3

values) comparing the transcript levels of each gene to geraniol emission levels is shown on the right of top panel. R-values with significant p values (≤ 0.05) are highlighted in green. (C) and (D) RT-qPCR relative quantification of *IDSs* and *NUDX1.1a* transcript levels in OB flowers during six development stages (C) and a day/night cycle (D) as indicated in (A). Data are means \pm SEM, $n = 4$.

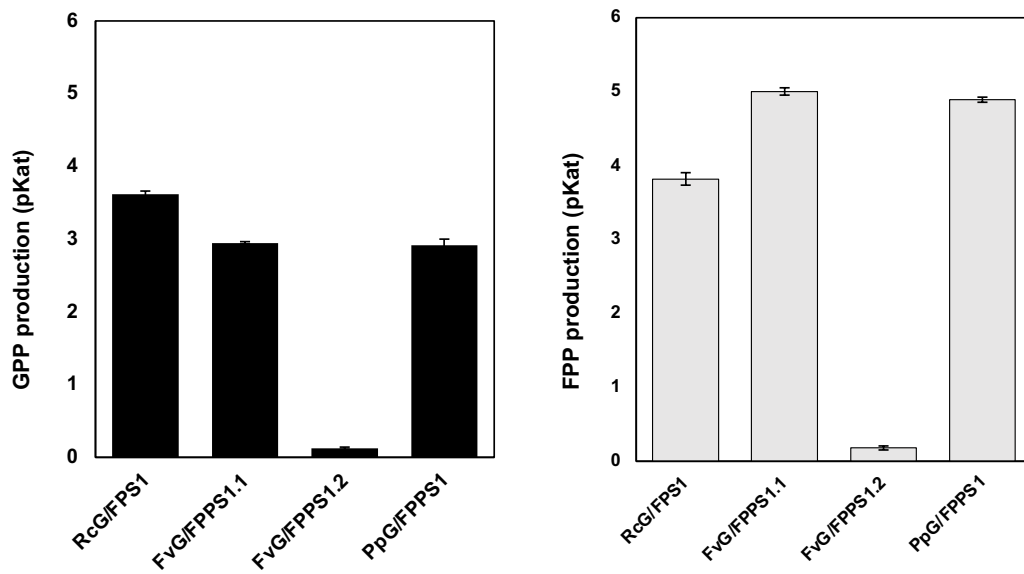


Figure S3.6 | FvFPPS1.2 having His residue at position 100 has low G/FPPS activity. (A) *Rosaceae* G/FPPS1s specific activities for GPP (left panel) and FPP (right panel) productions analyzed by LC-MS/MS. 250 ng of indicated enzymes were incubated for 20 minutes with 10 μ M DMAPP and IPP. Data are means of pKat \pm SEM, n = 3.

Appendix 3

| VOCs | Unlabelled | [2- ¹³ C]-mevalonolactone labelled | [² H ₂]-DOX labelled |
|-----------------|-----------------|---|---|
| Geraniol | 154 m/z (M + 0) | 155 m/z (M + 1) + 156 m/z (M + 2) | 156 m/z (M + 2) + 158 m/z (M + 4) |
| Germacrene D | 204 m/z (M + 0) | 205 m/z (M + 1) + 206 m/z (M + 2) + 207 m/z (M + 3) | 206 m/z (M + 2) + 208 m/z (M + 4) + 210 m/z (M + 6) |
| Dihydro-β-ionol | 196 m/z (M + 0) | 197 m/z (M + 1) + 198 m/z (M + 2) | 197 m/z (M + 1) + 198 m/z (M + 2) + 199 m/z (M + 3) |

Figure S3.7 | List of m/z ions used to calculate VOCs labelling.

Appendix 3

| Oligoname | Sequence(5'->3') | Usage |
|--------------------|---|--|
| RcNUDX1-1a_UCDS_F | GGCTTAA[UJATGGGAAACGAGACAGTAGT | cloning into pCAMBIA2300-U |
| RcNUDX1-1a_UCDS_R | GGTTTAA[UJTCATGTTGGAAAAGGGTTAAATC | cloning into pCAMBIA2300-U |
| RcFPPS1_UCDS_F | GGCTTAA[UJGCAATGGCGGATCTCAAGTC | cloning into pCAMBIA2300-U |
| RcFPPS1_UCDS_R | GGTTTAA[UJTCGAAGTTCATGACCCACA | cloning into pCAMBIA2300-U |
| RcFPPS2_UCDS_F | GGCTTAA[UJATGAGCAATTTAAGAGCCAAAGT | cloning into pCAMBIA2300-U |
| RcFPPS2_UCDS_R | GGTTTAA[UJCTACTTCTGCCTCTTGATATCT | cloning into pCAMBIA2300-U |
| RcGGPPS_LSU_UCDS_F | GGCTTAA[UJATGAGCTGTGTGAATCTGAGC | cloning into pCAMBIA2300-U |
| RcGGPPS_LSU_UCDS_R | GGTTTAA[UJACAGCAGAATCCCAACTCCT | cloning into pCAMBIA2300-U |
| RcGGPPS_SSU_UCDS_F | GGCTTAA[UJATGGCGTTCTCAGTGGTTAC | cloning into pCAMBIA2300-U |
| RcGGPPS_SSU_UCDS_R | GGTTTAA[UJCAAGGTAATGCAAAATGGTTCTGT | cloning into pCAMBIA2300-U |
| RcHOMO_UCDS_F | GGCTTAA[UJATGATATATTCCCGGGGATTTTC | cloning into pCAMBIA2300-U |
| RcHOMO_UCDS_R | GGTTTAA[UJTCGAAGTTCATGACCCACA | cloning into pCAMBIA2300-U |
| RcNUDX1-1a_qPCR_F | TGTTTTCATGAGGGCAGTGC | qPCR primer |
| RcNUDX1-1a_qPCR_R | CACGTTCTCCAAAGGCCAAA | qPCR primer |
| RcFPPS1_qPCR_F | TCATTGTCAATTCACCGCCG | qPCR primer |
| RcFPPS1_qPCR_R | GTCAGTTCCTCACCTGACA | qPCR primer |
| RcFPPS2_qPCR_F | ATTTCAAACCGCCCATGGAC | qPCR primer |
| RcFPPS2_qPCR_R | GCCGGTGAATTTCCAAGGAG | qPCR primer |
| RcLSU1_qPCR_F | CGGCTCTTCTTGAATGTGCA | qPCR primer |
| RcLSU1_qPCR_R | GCCTAATGTACCTCGCAAAA | qPCR primer |
| RcSSU_qPCR_F | TCACGGAGATTGCCAGAACT | qPCR primer |
| RcSSU_qPCR_R | ACACTCACCCATTTCCCCAA | qPCR primer |
| RcHOMO_qPCR_F | CTCTCTCGAGCTTGTGTTGC | qPCR primer |
| RcHOMO_qPCR_R | CAACGTTGATCAGCAGCAGT | qPCR primer |
| RcPP2A_qPCR_F | GATTCGTGATGCTGCTGCTA | qPCR primer |
| RcPP2A_qPCR_R | TTGCACGCAGAATTGTCATT | qPCR primer |
| RcEF1a_qPCR_qPCR_F | GGGTAAGGAGAAGGTTACATC | qPCR primer |
| RcEF1a_qPCR_qPCR_R | CAGCCTCCTTCTCAAACCTCT | qPCR primer |
| RcFPPS1_NcoI_F | GGTACCATTGGCGGATCTCAAGTCAAA | Cloning FPPS1 into pRSF-DUET-1 |
| RcFPPS1_NotI_R | AAAGCGGCCGCTACTTTTTCTCTTGTAAATCTTACC | Cloning FPPS1 into pRSF-DUET-1 |
| pET30a_UF | GGCTTAA[UJAAATTCGAACGCCAGCACAT | sub-cloning genscript pET30a+ vector into pCAMBIA2300-U |
| pET30a_UR | GGTTTAA[UJTTGTTAGCAGCCGGATCTCA | sub-cloning genscript pET30a+ vector into pCAMBIA2300-U |
| RcNUDX1-1a_ATG_F | AAAAAGCAGGCTTGATGGGAAACGAGACAGTAGTAG | cloning into pDON221 |
| RcNUDX1a_NSTOP_R | AAAAAGCAGGCTTGATGGGAAACGAGGTTAAATCCATC | cloning into pDON221 without stop codon = FP fusion in C-ter |
| RcFPPS1_ATG_F | AAAAAGCAGGCTTGATGGCGGATCTCAAGTCAAA | cloning into pDON221 |
| RcFPPS1_NSTOP_R | AGAAAAGCTGGGTTCTTTTCTCTTGTAAATCTTACC | cloning into pDON221 without stop codon = FP fusion in C-ter |
| RcFPPS2_ATG_F | AAAAAGCAGGCTTGATGAGCAATTTAAGAGCCAAAGT | cloning into pDON221 |
| RcFPPS2_NSTOP_R | AGAAAAGCTGGGTTCTTCTGCCTCTTGATATCTTTGC | cloning into pDON221 without stop codon = FP fusion in C-ter |
| RcLSU1_ATG_F | AAAAAGCAGGCTTGATGAGCTGTGTAATCTGAGC | cloning into pDON221 |
| RcLSU1_NSTOP_R | AGAAAAGCTGGGTTATTTTGCCTGTAAGCAATGTAAT | cloning into pDON221 without stop codon = FP fusion in C-ter |
| RcSSU_ATG_F | AAAAAGCAGGCTTGATGGCGTTCTCAGTGGTTAC | cloning into pDON221 |
| RcSSU_NSTOP_R | AGAAAAGCTGGGTTAAGACTAAAACCTCTATCAACTGC | cloning into pDON221 without stop codon = FP fusion in C-ter |
| RcHOMO_ATG_F | AAAAAGCAGGCTTGATGATATATTCCCGGGGATTTTC | cloning into pDON221 |
| RcHOMO_NSTOP_R | AGAAAAGCTGGGTTCTTCTGTTCTTGTAAATGACTAGTAGT | cloning into pDON221 without stop codon = FP fusion in C-ter |
| AtFPPS1_Kpn1_F | CGGGGTACCATTGGAGACCGATCTCAAGTCA | cloning AtFPPS1 into pET30a+ |
| AtFPPS1_Sal1_R | CGGCTGACCTACTCTGCCTCTGTAGATCTT | cloning AtFPPS1 into pET30a+ |

Figure S3.8 | List of primers used in this study.

Appendix 3

| | | | |
|---------------|-----------------------------|--------------------------------|---------------------------------|
| RcGERS1 | RchiOBHm_Ch5g0038101 | XP_024161123 | XM_02430535.2 |
| CrFPPS | | ADO95193.1 | HQ316638.1 |
| CrGPPS.LSU | | AGL91645.1 | JX417183.1 |
| CrGPPS.SSU | | AGL91646.1 | JX417184.1 |
| CrGPPS | | AGL91647.1 | JX417185.1 |
| MpFPPS | | AAK63847.1 | AF384040.1 |
| MpGPPS.LSU | | ABW86879.1 | EU108695.1 |
| MpGPPS.SSU | | ABW86880.1 | EU108696.1 |
| AtFPPS1 | At5g47770 | NM_124151 | Q09152-1 |
| AtFPPS2 | AT4G17190 | NM_117823 | Q43315-1 |
| AtGPPS/PPPS | At2g34630 | Y17376.1 | Q5HZ00-1 |
| AtGPPS.SSU-II | AT4G38460.1 | NP_195558.1 | Q39108 |
| AtGPPS.LSU | At4g36810 | | |
| SIFPPS | | NP_001234068 | NM_00124713.2 |
| SIGGPPS2 | Solyc04g079960 | DQ267903 | |
| SIGPPS/PPPS | Solyc08g023470 | NM_001247160.2 | |
| SIGPPS.SSU-I | Solyc07g064660 | XM_010326009 | |
| SIGPPS.SSU-II | Solyc09g008920 | | |
| RcGPPS/HOMO | RchiOBHm_Ch5g0014811 | A0A2P6Q5R6 | |
| RcG/FPPS1 | RchiOBHm_Ch5g0000321 | A0A2P6Q231 | XP_02415672.0.1 |
| RcFPPS2 | RchiOBHm_Ch5g0075621 | A0A2P6QLH7 | XM_02430430.2 |
| RcGGPPS.LSU1 | RchiOBHm_Ch2g0102671 | A0A2P6RMR0 | XM_02432653.2 |
| RcGGPPS.LSU2 | RchiOBHm_Ch3g0493061 | A0A2P6RGN4 | XM_02433531.0.2 |
| PtFPPS1 | | MK492685 | |
| PtGPPS1.LSU | | MK492686 | |
| PtGPPS2.LSU | | MK492687 | |
| PtGPPS.SSU1 | | MK492683 | |
| PtGPPS.SSU2 | | MK492684 | |
| AmGPPS.SSU | | AAS82859 | |
| AmGPPS.LSU | | AAS82860 | |
| CbGPPS.SSU | | AAS82870.1 | |
| HIGPPS.SSU | ACQ90681 | | |
| HIGPPS.LSU | ACQ90682 | | |
| AgGPPS1 | AAN01133 | | |
| AgGGPPS | AAL17614 | | |
| PalDS1 | ACZ57571 | | |
| PalDS2 | ACA21458.2 | | |
| PalDS3 | ACA21459 | | |
| QrIDS1 | CAC20852 | | |
| PbGPPS | ABV71395.1 | | |

Figure S3.9 | Accession numbers of genes used in this study.

Bibliography |

- Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., Lindahl E., **2015**. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19-25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Adal A.M., Mahmoud S.S., **2020**. Short-chain isoprenyl diphosphate synthases of lavender (*Lavandula*). *Plant Mol Biol* 102, 517–535. <https://doi.org/10.1007/s11103-020-00962-8>
- Adam R.P., **2007**. *Identification of essential oil components by gas chromatography/mass spectrometry*. Allured Publ. Corporation edition. pp 804
- Aebi H., **1984**. Catalase *in vitro*. *Methods in Enzymology* 105, 121-26. [https://doi.org/10.1016/S0076-6879\(84\)05016-3](https://doi.org/10.1016/S0076-6879(84)05016-3)
- Afonine P.V., Grosse-Kunstleve R.W., Echols N., Headd J.J., Moriarty N.W., Mustyakimov M., Terwilliger T.C., Urzhumtsev A., Zwart P.H., Adams, P.D., **2012**. Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Cryst. D* 68, 352-367. <https://doi.org/10.1107/S0907444912001308>
- Aharoni A., Giri A.P., Verstappen F.W.A., Berteau C.M., Sevenier R., Sun Z., Jongsma M.A., Schwab W., Bouwmeester H.J., **2004**. Gain and loss of fruit Flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* 16, 3110–3131. <https://doi.org/10.1105/tpc.104.023895>
- Ashaari N.S., Ab. Rahim M.H., Sabri S., Lai K.S., Song A.A.-L., Abdul Rahim R., Wan Abdullah W.M.A.N., Ong Abdullah J., **2020**. Functional characterization of a new terpene synthase from *Plectranthus amboinicus*. *PLoS ONE* 15, e0235416. <https://doi.org/10.1371/journal.pone.0235416>
- Axelsson E., Ratnakumar A., Arendt M.L., Maqbool K., Webster M.T., Perloski M., Liberg O., Arnemo J.M., Hedhammar A., Lindblad-Toh K., **2013**. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. 495, 360–364. <https://doi.org/10.1038/nature11837>
- Bach T.J., **1986**. Hydroxymethylglutaryl-CoA reductase, a key enzyme in phytosterol synthesis? *Lipids* 21, 82–88. <https://doi.org/10.1007/BF02534307>
- Bach T.J., Boronat A., Campos N., Ferrer A., Vollack K.-U., **1999**. Mevalonate Biosynthesis in Plants *Critical Reviews in Biochemistry and molecular Biology* 34, 107–122. <https://doi.org/10.1080/10409239991209237>
- Bach Thomas J., Lichtenthaler H.K., **1982**. Mevinolin: A highly specific Inhibitor of microsomal 3-Hydroxy-3-Methylglutaryl-Coenzyme A Reductase of radish plants. *Zeitschrift für Naturforschung C* 37, 46–50. <https://doi.org/10.1515/znc-1982-1-209>
- Bach T.J., Lichtenthaler H.K., **1982**. Inhibition of mevalonate biosynthesis and of plant growth by the fungal metabolite mevinolin. *Developments in plant biology* 64 18–24. [https://doi.org/10.1016/S0254-6299\(15\)30822-X](https://doi.org/10.1016/S0254-6299(15)30822-X)
- Baldermann S., Yang Z., Sakai M., Fleischmann P., Watanabe N., **2009**. Volatile Constituents in the Scent of Roses. *Floriculture and Ornamental Biotechnology* 3, 89–97.
- Banerjee A., Wu Y., Banerjee R., Li Y., Yan H., Sharkey T.D., **2013**. Feedback inhibition of deoxy-d-xylulose-5-phosphate synthase regulates the methylerythritol 4-phosphate pathway. *Journal of Biological Chemistry* 288, 16926–16936. <https://doi.org/10.1074/jbc.M113.464636>
- Barja M.V., Ezquerro M., Beretta S., Diretto G., Florez-Sarasa I., Feixes E., Fiore A., Karlova R., Fernie A.R., Beekwilder J., Rodríguez-Concepción M., **2021**. Several geranylgeranyl diphosphate synthase isoforms supply metabolic substrates for carotenoid biosynthesis in tomato. *New Phytol* 231, 255–272. <https://doi.org/10.1111/nph.17283>
- Bassard J.-E., Halkier B.A., **2018**. How to prove the existence of metabolons? *Phytochem Rev* 17, 211–227. <https://doi.org/10.1007/s11101-017-9509-1>
- Baudino S., Huguenev P., Caissard J.C., **2020**. Evolution of scent genes. *Biology of plant volatiles. Boca Raton: CRC Presse*. pp 217–234.
- Bellés X., Martín D., Piulachs M.-D., **2005**. The mevalonate pathway and the synthesis of juvenile hormone in insects. *Annu. Rev. Entomol.* 50, 181-199. <https://doi.org/10.1146/annurev.ento.50.071803.130356>
- Ben Zvi M.M., Shklarman E., Masci T., Kalev H., Debener T., Shafir S., Ovadis M., Vainstein A., **2012**. PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. *New Phytol.* 195, 335–345. <https://doi.org/10.1111/j.1469-8137.2012.04161.x>
- Bergman M.E., Bhardwaj M., Phillips M.A., **2021**. Cytosolic geraniol and citronellol biosynthesis require a Nudix hydrolase in rose-scented geranium (*Pelargonium graveolens*). *The Plant Journal* 107, 493–510. <https://doi.org/10.1111/tpj.15304>

- Bergougnoux V., Caissard J.-C., Jullien F., Magnard J.L., Scalliet G., Cock J.M., Huguency P., Baudino S., **2007**. Both the adaxial and abaxial epidermal layers of the rose petal emit volatile scent compounds. *Planta* 226, 853-866. <https://doi.org/10.1007/s00425-007-0531-1>
- Bessman M.J., **2019**. A cryptic activity in the Nudix hydrolase superfamily. *Protein Sci.* 28, 1494-1500. <https://doi.org/10.1002/pro.3666>
- Bessman M.J., Frick D.N., O'Handley S.F., **1996**. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.* 271, 25059-25062. <https://doi.org/10.1074/jbc.271.41.25059>
- Bick J.A., Lange B.M., **2003**. Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Archives of Biochemistry and Biophysics* 415, 146-154. [https://doi.org/10.1016/S0003-9861\(03\)00233-9](https://doi.org/10.1016/S0003-9861(03)00233-9)
- Blerot B., Martinelli L., Prunier C., Saint-Marcoux D., Legrand S., Bony A., Sarrabère L., Gros F., Boyer N., Caissard J.-C., Baudino S., Jullien F., **2018**. Functional analysis of four terpene synthases in rose-scented *Pelargonium* cultivars (*Pelargonium* × *hybridum*) and Evolution of Scent in the *Pelargonium* Genus. *Front. Plant Sci.* 9, 1435. <https://doi.org/10.3389/fpls.2018.01435>
- Boachon B., Burdloff Y., Ruan J.X., Rojo R., Junker R.R., Vincent B., Nicolè F., Bringel F., Lesot A., Henry L., Bassard J.-E., Mathieu S., Allouche L., Kaplan I., Dudareva N., Vuilleumier S., Miesch L., André F., Navrot N., Chen X.-Y., Werck-Reichhart D., **2019**. A promiscuous CYP706A3 reduces terpene volatile emission from *Arabidopsis* flowers, affecting florivores and the floral microbiome. *Plant Cell* 31, 2947-2972. <https://doi.org/10.1105/tpc.19.00320>
- Boachon B., Junker R.R., Miesch L., Bassard J.-E., Höfer R., Caillieudeaux R., Seidel D.E., Lesot A., Heinrich C., Ginglinger J.-F., Allouche L., Vincent B., Wahyuni D.S.C., Paetz C., Beran F., Miesch M., Schneider B., Leiss K., Werck-Reichhart D., **2015**. CYP76C1 (Cytochrome P450)-mediated linalool metabolism and the formation of volatile and soluble linalool oxides in *Arabidopsis* flowers: A strategy for defense against floral antagonists. *Plant Cell* tpc.15.00399. <https://doi.org/10.1105/tpc.15.00399>
- Bohlmann J., Meyer-Gauen G., Croteau R., **1998**. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4126-4133. <https://doi.org/10.1073/pnas.95.8.4126>
- Bolger A.M., Lohse M., Usadel B., **2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bortolotti L., Costa C., **2014**. Chemical communication in the honey bee society. *Neurobiology of Chemical Communication* 247-210. <https://doi.org/10.1201/b16511-6>
- Boutanaev A.M., Osbourn A.E., **2018**. Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc. Natl Acad. Sci. USA* 115, E6650-E6658. <https://doi.org/10.1073/pnas.1721318115>
- Bouvier F., Suire C., d'Harlingue A., Backhaus R.A., Camara B., **2000**. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *Plant J* 24, 241-252. <https://doi.org/10.1046/j.1365-313x.2000.00875.x>
- Bradford M.M., **1976**. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry* 72, 248-254. [https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3)
- Burke C.C., Wildung M.R., Croteau R., **1999**. Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13062-13067. <https://doi.org/10.1073/pnas.96.23.13062>
- Burke C. and Croteau R., **2002**. Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate. *Journal of Biological Chemistry* 277, 3141-3149. <https://doi.org/10.1074/jbc.M105900200>
- Burlat V., Oudin A., Courtois M., Rideau M., St-Pierre B., **2004**. Co-expression of three MEP pathway genes and *geraniol 10-hydroxylase* in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *The Plant Journal* 38, 131-141. <https://doi.org/10.1111/j.1365-313X.2004.02030.x>
- Busso D., Delagoutte-Busso B., Moras D., **2005**. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal. Biochem.* 343, 313-321. <https://doi.org/10.1016/j.ab.2005.05.015>

- Buti M., Moretto M., Barghini E., Mascagni F., Natali L., Brilli M., Lomsadze A., Sonogo P., Giongo L., Alonge M., Velasco R., Varotto C., Šurbanovski N., Borodovsky M., Ward J.A., Engelen K., Cavallini A., Cestaro A., Sargent D.J., **2018**. The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry). *Gigascience*. 7, 1–14. <https://doi.org/10.1093/gigascience/giy010>
- Cairns T., **2003**. Classification. *Encyclopedia of Rose Science Amsterdam, The Netherlands: Elsevier*. pp 117–123.
- Cannon S.B., Mitra A., Baumgarten A., Young N.D., May G., **2004**. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol* 4, 10. <https://doi.org/10.1186/1471-2229-4-10>
- Cao L., Zhang P., Grant D.F., **2009**. An insect farnesyl phosphatase homologous to the N-terminal domain of soluble epoxide hydrolase. *Biochem. Biophys. Res. Commun.* 380, 188–192. <https://doi.org/10.1016/j.bbrc.2009.01.079>
- Carter M., Jemth A.S., Hagenkort A., Page B.D., Gustafsson R., Griese J.J., Gad H., Valerie N.C., Desroses M., Bostrom J., Warpman Berglund U., Helleday T., Stenmark P., **2015**. Crystal structure, biochemical and cellular activities demonstrate separate functions of MTH1 and MTH2. *Nat. Commun.* 6, 7871. <https://doi.org/10.1038/ncomms8871>
- Case D.A., Cheatham III T.E., Darden T., Gohlke H., Luo R., Merz Jr K.M., Onufriev A., Simmerling C., Wang B., Woods, R.J., **2005**. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. <https://doi.org/10.1002/jcc.20290>
- Castresana J., **2000**. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17, 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Cerbin S., Jiang N., **2018**. Duplication of host genes by transposable elements. *Curr Opin Genet Dev.* 49, 63–69. <https://doi.org/10.1016/j.gde.2018.03.005>
- Chang T.H., Hsieh F.L., Ko T.P., Teng K.H., Liang P.H., Wang A.H.J., **2010**. Structure of a heterotetrameric geranyl pyrophosphate synthase from mint (*Mentha piperita*) reveals intersubunit regulation. *The Plant Cell* 22, 454–467. <https://doi.org/10.1105/tpc.109.071738>
- Chaykin S., Law J., Phillips A.H., Tchen T.T., Bloch K., **1958**. Phosphorylated intermediates in the synthesis of squalene. *Proc. Natl. Acad. Sci. U.S.A.* 44, 998–1004. <https://doi.org/10.1073/pnas.44.10.998>
- Chen F., Tholl D., Bohlmann J., Pichersky E., **2011**. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom: Terpene synthase family. *The Plant Journal* 66, 212–229. <https://doi.org/10.1111/j.1365-313X.2011.04520.x>
- Chen Q., Fan D., Wang G., **2015**. Heteromeric Geranyl(geranyl) Diphosphate Synthase Is Involved in Monoterpene Biosynthesis in *Arabidopsis* Flowers. *Molecular Plant* 8, 1434–1437. <https://doi.org/10.1016/j.molp.2015.05.001>
- Chen W., Viljoen A.M., **2010**. Geraniol - A review of a commercially important fragrance material. *S. Afr. J. Bot.* 76, 643–651. <https://doi.org/10.1016/j.sajb.2010.05.008>
- Chen X., Wang Y., Sun J., Wang J., Xun H., Tang F., **2016**. Cloning, expression and functional characterization of two sesquiterpene synthase genes from moso bamboo (*Phyllostachys edulis*). *Protein Expr. Purif.* 120, 1–6. <https://doi.org/10.1016/j.pep.2015.11.019>
- Chen X.-M., Kobayashi H., Sakai M., Hirata H., Asai T., Ohnishi T., Baldermann S., Watanabe N., **2011**. Functional characterization of rose phenylacetaldehyde reductase (PAR), an enzyme involved in the biosynthesis of the scent compound 2-phenylethanol. *Journal of Plant Physiology* 168, 88–95. <https://doi.org/10.1016/j.jplph.2010.06.011>
- Cheng A.X., Xiang C.Y., Li J.X., Yang C.Q., Hu W.L., Wang L.J., Lou Y.G., Chen X.Y., **2007**. The rice (E)-beta-caryophyllene synthase (OsTPS3) accounts for the major inducible volatile sesquiterpenes. *Phytochemistry* 68, 1632–1641. <https://doi.org/10.1016/j.phytochem.2007.04.008>
- Churchill G.A., Doerge R.W., **1994**. Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. <https://doi.org/10.1093/genetics/138.3.963>
- Closa M., Vranová E., Bortolotti C., Bigler L., Arró M., Ferrer A., Grissem W., **2010**. The *Arabidopsis thaliana* FPP synthase isozymes have overlapping and specific functions in isoprenoid biosynthesis, and complete loss of FPP synthase activity causes early developmental arrest: *Arabidopsis thaliana* FPP synthase mutants. *The Plant Journal* 63, 512–525. <https://doi.org/10.1111/j.1365-313X.2010.04253.x>

Bibliography

- Conart C., Saclier N., Foucher F., Goubert C., Rius-Bony A., Paramita S.N., Moja S., Thouroude T., Douady C., Sun P., Nairaud B., Saint-Marcoux D., Bahut M., Jeauffre J., Hibrand Saint-Oyant L., Schuurink R.C., Magnard J.-L., Boachon B., Dudareva N., Baudino S., Caissard J.-C., **2022**. Duplication and Specialization of NUDX1 in Rosaceae Led to Geraniol Production in Rose Petals. *Mol Biol Evol* 39, <https://doi.org/10.1093/molbev/msac002>
- Connolly J.D., Hill R.A., **1991**. Triterpenoids, in: *Second Supplements to the 2nd Edition of Rodd's Chemistry of Carbon Compounds*. Elsevier, pp. 555–596. <https://doi.org/10.1016/B978-044453347-0.50085-9>
- Croteau R., **1987**. Biosynthesis and catabolism of monoterpenoids. *chemical review* 87, 26.
- Cuchet A., Anchisi A., Schiets F., Clément Y., Lantéri P., Bonnefoy C., Jame P., Carénini E., Casabianca H. **2021**. Determination of enantiomeric and stable isotope ratio fingerprints of active secondary metabolites in neroli (*Citrus aurantium L.*) essential oils for authentication by multidimensional gas chromatography and GC-C/P-IRMS. *J Chromatogr B Analyt Technol Biomed Life Sci* 1185, 123003. <https://doi.org/10.1016/j.jchromb.2021.123003>
- Cunillera N., Arró M., Delourme D., Karst F., Boronat A., Ferrer A., **1996**. *Arabidopsis thaliana* contains two differentially expressed farnesyl-diphosphate synthase genes. *Journal of Biological Chemistry* 271, 7774–7780. <https://doi.org/10.1074/jbc.271.13.7774>
- Cunillera N., Boronat A., Ferrer A., **1997**. The *Arabidopsis thaliana* FPS1 Gene Generates a novel mRNA that encodes a mitochondrial farnesyl-diphosphate synthase isoform. *Journal of Biological Chemistry* 272, 15381–15388. <https://doi.org/10.1074/jbc.272.24.15381>
- D'Andrea L., Amenós M., Rodríguez-Concepción M., **2014**. Confocal laser scanning microscopy detection of chlorophylls and carotenoids in chloroplasts and chromoplasts of tomato fruit. *Methods in Molecular Biology* 1153, 227–232. https://doi.org/10.1007/978-1-4939-0606-2_16
- Daccord N., Celton J.-M., Linsmith G., Becker C., Choisne N., Schijlen E., van de Geest H., Bianco L., Micheletti D., Velasco R., et al., **2017**. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet.* 49, 1099–1106. <https://doi.org/10.1038/ng.3886>
- Davidovich-Rikanati R., Lewinsohn E., Bar E., Iijima Y., Pichersky E., Sitrit Y., **2008**. Overexpression of the lemon basil α -zingiberene synthase gene increases both mono- and sesquiterpene contents in tomato fruit: *Overexpression of ZIS in tomato fruit*. *The Plant Journal* 56, 228–238. <https://doi.org/10.1111/j.1365-313X.2008.03599.x>
- DeBolt S., **2010**. Copy Number Variation shapes genome diversity in *Arabidopsis* over immediate family generational scales. *Genome Biol Evol.* 2, 441–453. <https://doi.org/10.1093/gbe/evq033>
- Debray K., Marie-Magdelaine J., Ruttink T., Clotault J., Foucher F., Malécot V., **2019**. Identification and assessment of variable single-copy orthologous (SCO) nuclear loci for low-level phylogenomics: a case study in the genus *Rosa* (Rosaceae). *BMC Evol Biol.* 19, 152. <https://doi.org/10.1186/s12862-019-1479-z>
- de Bruijn W.J.C., Levisson M., Beekwilder J., van Berkel W.J.H., Vincken J.-P., **2020**. Plant aromatic prenyltransferases: tools for microbial cell factories. *Trends in Biotechnology* 38, 917–934. <https://doi.org/10.1016/j.tibtech.2020.02.006>
- Degenhardt J., Köllner T.G., Gershenzon J., **2009**. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637. <https://doi.org/10.1016/j.phytochem.2009.07.030>
- De Loof A., Schoofs L., **2019**. Mode of action of Farnesol, the “noble unknown” in particular in Ca^{2+} homeostasis, and its juvenile hormone-esters in evolutionary retrospect. *Front. Neurosci.* 13, 141. <https://doi.org/10.3389/fnins.2019.00141>
- Demissie Z.A., Erland Lauren A.E., Rheault M.R., Mahmoud S.S., **2013**. The biosynthetic origin of irregular monoterpenes in *Lavandula*. *Journal of Biological Chemistry* 288, 6333–6341. <https://doi.org/10.1074/jbc.M112.431171>
- Despinasse Y., Fiorucci S., Antonczak S., Moja S., Bony A., Nicolè F., Baudino S., Magnard J.-L., Jullien F., **2017**. Bornyl-diphosphate synthase from *Lavandula angustifolia*: A major monoterpene synthase involved in essential oil quality. *Phytochemistry* 137, 24–33. <https://doi.org/10.1016/j.phytochem.2017.01.015>
- Disch A., Schwender J., Müller C., Lichtenthaler H.K., Rohmer M., **1998**. Distribution of the mevalonate and glyceraldehyde phosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium *Synechocystis* PCC 6714. *Biochemical Journal* 333, 381–388. <https://doi.org/10.1042/bj3330381>
- Dobrzanska M., Szurmak B., Wyslouch-Cieszynska A., Kraszewska E., **2002**. Cloning and characterization of the first member of the Nudix family from *Arabidopsis thaliana*. *J. Biol. Chem.* 277, 50482–50486. <https://doi.org/10.1074/jbc.M205207200>

- Dobson H.E.M., Danielson E.M., Wesep I.D.V., **1999**. Pollen odor chemicals as modulators of bumble bee foraging on *Rosa rugosa* Thunb. (*Rosaceae*). *Plant Species Biol* 14, 153–166. <https://doi.org/10.1046/j.1442-1984.1999.00020.x>
- Dogbo O., Camara B., **1987**. Purification of isopentenyl pyrophosphate isomerase and geranylgeranyl pyrophosphate synthase from *Capsicum* chromoplasts by affinity chromatography. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 920, 140–148. [https://doi.org/10.1016/0005-2760\(87\)90253-0](https://doi.org/10.1016/0005-2760(87)90253-0)
- Dong L., Miettinen K., Goedbloed M., Verstappen F.W.A., Voster A., Jongsma M.A., Memelink J., Krol S. van der, Bouwmeester H.J., **2013**. Characterization of two geraniol synthases from *Valeriana officinalis* and *Lippia dulcis*: Similar activity but difference in subcellular localization. *Metabolic Engineering* 20, 198–211. <https://doi.org/10.1016/j.ymben.2013.09.002>
- Dubois A., Carrere S., Raymond O., Pouvreau B., Cottret L., Rocchia A., Onesto J.P., Sakr S., Atanassova R., Baudino S., Foucher F., Le Bris M., Gouzy J., Bendahmane M., **2012**. Transcriptome database resource and gene expression atlas for the rose. *BMC Genomics*. 13, 638–648. <https://doi.org/10.1186/1471-2164-13-638>
- Ducluzeau A., Wamboldt Y., Elowsky C.G., Mackenzie S.A., Schuurink R.C., Basset G.J.C., **2012**. Gene network reconstruction identifies the authentic *trans* -prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in *Arabidopsis*. *The Plant Journal* 69, 366–375. <https://doi.org/10.1111/j.1365-313X.2011.04796.x>
- Dudareva N., Klempien A., Muhlemann J.K., Kaplan I., **2013**. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol* 198, 16–32. <https://doi.org/10.1111/nph.12145>
- Edger P.P., VanBuren R., Colle M., Poorten T.J., Wai C.M., Niederhuth C.E., Alger E.I., Ou S., Acharya C.B., Wang J., et al., **2018**. Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (*Fragaria vesca*) with chromosome-scale contiguity. *Gigascience*. 7, 1–7. <https://doi.org/10.1093/gigascience/gix124>
- Emsley P., Lohkamp B., Scott W.G., Cowtan K., **2010**. Features and development of *Coot*. *Acta Cryst. D* 66, 486–501. <https://doi.org/10.1107/S0907444910007493>
- Epifano F., Taddeo V.A., Genovese S., Preziuso F., Fraternali D., **2018**. Modulation of the prenylation step in *Anethum graveolens* cultured calli. Part II. The role of p-cumaric acid and boropinic acid. *Industrial Crops and Products* 124, 209–212. <https://doi.org/10.1016/j.indcrop.2018.07.085>
- Farhi M., Lavie O., Masci T., Hendel-Rahman K., Weiss D., Abeliovich H., Vainstein A., **2010**. Identification of rose phenylacetaldehyde synthase by functional complementation in yeast. *Plant Mol Biol* 72, 235–245. <https://doi.org/10.1007/s11103-009-9564-0>
- Ferguson A.A., Zhao D., Jiang N., **2013**. Selective acquisition and retention of genomic sequences by *Pack-Mutator* -Like elements based on Guanine-Cytosine content and the breadth of expression. *Plant Physiol.* 163, 1419–1432. <https://doi.org/10.1104/pp.113.223271>
- Fernandez-Pozo N., Rosli H.G., Martin G.B., Mueller L.A., **2015**. The SGN VIGS tool: User-friendly software to design virus-induced gene silencing (VIGS) Constructs for functional genomics. *Mol Plant* 8, 486–488. <https://doi.org/10.1016/j.molp.2014.11.024>
- Flügge U. -I., Gao W., **2005**. Transport of isoprenoid intermediates across chloroplast envelope membranes. *Plant Biology* 7, 91–97. <https://doi.org/10.1055/s-2004-830446>
- Fougère-Danezan M., Joly S., Bruneau A., Gao X.-F., Zhang L.-B., **2015**. Phylogeny and biogeography of wild roses with specific attention to polyploids. *Annals of Botany* 115, 275–291. <https://doi.org/10.1093/aob/mcu245>
- Francis M.J., O'connell M., **1969**. The incorporation of mevalonic acid into rose petal monoterpenes. *Phytochemistry* 8, 1705-1708. [https://doi.org/10.1016/S0031-9422\(00\)85957-9](https://doi.org/10.1016/S0031-9422(00)85957-9)
- Gabelli S.B., McLellan J.S., Montalvetti A., Oldfield E., Docampo R., Amzel L.M., **2005**. Structure and mechanism of the farnesyl diphosphate synthase from *Trypanosoma cruzi*: Implications for drug design. *Proteins* 62, 80–88. <https://doi.org/10.1002/prot.20754>
- Gad H., Koolmeister T., Jemth A.-S., Eshtad S., Jacques S.A., Ström C.E., Svensson L.M., Schultz N., Lundbäck T., Einarsdóttir B.O., Saleh A., Göktürk C., Baranczewski P., Svensson R., Berntsson R.P.-A., Gustafsson R., Strömberg K., Sanjiv K., Jacques-Cordonnier M.-C., Desroses M., Gustavsson A.-L., Olofsson R., Johansson F., Homan E.J., Loseva O., Bräutigam L., Johansson L., Höglund A., Hagenkört A., Pham T., Altun M., Gaugaz F.Z., Vikingsson S., Evers B., Henriksson M., Vallin K.S.A., Wallner O.A., Hammarström L.G.J., Wiita E., Almlöf I., Kalderén C., Axelsson H., Djureinovic T., Puigvert J.C., Häggblad M., Jeppsson F., Martens U., Lundin C., Lundgren B., Granelli I., Jensen A.J., Artursson P., Nilsson J.A., Stenmark P., Scobie M., Berglund U.W., Helleday T., **2014**. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* 508, 215–221. <https://doi.org/10.1038/nature13181>

Bibliography

- Galindo-González L., Mhiri C., Deyholos M.K., Grandbastien M.A., **2017**. LTR-retrotransposons in plants: Engines of evolution. *Gene*. 626, 14–25. <https://doi.org/10.1016/j.gene.2017.04.051>
- Gerber E., Hemmerlin A., Hartmann M., Heintz D., Hartmann M.-A., Mutterer J., Rodríguez-Concepción M., Boronat A., Van Dorselaer A., Rohmer M., Crowell D.N., Bach T.J., **2009**. The Plastidial 2- C -Methyl- D - Erythritol 4-Phosphate pathway provides the isoprenyl moiety for protein geranylgeranylation in tobacco BY-2 Cells. *The Plant Cell* 21, 285–300. <https://doi.org/10.1105/tpc.108.063248>
- Gibson R.W., Pickett J.A., **1983**. Wild potato repels aphids by release of aphid alarm pheromone. *Nature* 302, 608–609. <https://doi.org/10.1038/302608a0>
- Gonda I., Bar E., Portnoy V., Lev S., Burger J., Schaffer A.A., Tadmor Y., Gepstein S., Giovannoni J.J., Katzir N., Lewinsohn E., **2010**. Branched-chain and aromatic amino acid catabolism into aroma volatiles in *Cucumis melo* L. fruit. *Journal of Experimental Botany* 61, 1111–1123. <https://doi.org/10.1093/jxb/erp390>
- Goodstein D.M., Shuu S., Howson R., Neupane R., Hayes R.D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar D.S., **2012**. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res* 40, 11798–1186. <https://doi.org/10.1093/nar/gkr944>
- Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. **2011**. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Grandbastien M.-A., **2015**. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Bioch. Bioph. Acta* 1849, 403–416. <https://doi.org/10.1016/j.bbagr.2014.07.017>
- Green S.A., Chen X., Nieuwenhuizen N.J., Matich A.J., Wang M.Y., Bunn B.J., Yauk Y.-K., Atkinson R.G., **2012**. Identification, functional characterization, and regulation of the enzyme responsible for floral (*E*)-nerolidol biosynthesis in kiwifruit (*Actinidia chinensis*). *J. Exp. Bot.* 63, 1951–1967. <https://doi.org/10.1093/jxb/err393>
- Gu T., Han Y., Huang R., McAvoy R.J., Li Y., **2016**. Identification and characterization of histone lysine methylation modifiers in *Fragaria vesca*. *Sci Rep.* 6, 23581. <https://doi.org/10.1038/srep23581>
- Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O., **2010**. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59, 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Guirmand G., Guihur A., Perello C., Phillips M., Mahroug S., Oudin A., Dugé de Bernonville T., Besseau S., Lanoue A., Giglioli-Guivarc'h N., Papon N., St-Pierre B., Rodríguez-Concepción M., Burlat V., Courdavault V., **2020**. Cellular and subcellular compartmentation of the 2C-Methyl-D-Erythritol 4-Phosphate pathway in the Madagascar periwinkle. *Plants* 9, 462. <https://doi.org/10.3390/plants9040462>
- Gunawardana D., Cheng H.-C., Gayler K.R., **2007**. Identification of functional domains in *Arabidopsis thaliana* mRNA decapping enzyme (AtDcp2). *Nucleic Acids Research* 36, 203–216. <https://doi.org/10.1093/nar/gkm1002>
- Gunawardana D., Likic V.A., Gayler K.R., **2009**. A comprehensive bioinformatics analysis of the Nudix superfamily in *Arabidopsis thaliana*. *Comp. Funct. Genomics* 2009, 1–13. <https://doi.org/10.1155/2009/820381>
- Günther J., Lackus N.D., Schmidt A., Huber M., Stödtler H.-J., Reichelt M., Gershenzon J., Köllner T.G., **2019**. Separate pathways contribute to the herbivore-induced formation of 2-Phenylethanol in poplar. *Plant Physiol.* 180, 767–782. <https://doi.org/10.1104/pp.19.00059>
- Gupta P., Akhtar N., Tewari S.K., Sangwan R.S., Trivedi P.K., **2011**. Differential expression of farnesyl diphosphate synthase gene from *Withania somnifera* in different chemotypes and in response to elicitors. *Plant Growth Regul* 65, 93–100. <https://doi.org/10.1007/s10725-011-9578-x>
- Gutensohn M., Klempien A., Kaminaga Y., Nagegowda D.A., Negre-Zakharov F., Huh J.-H., Luo H., Weizbauer R., Mengiste T., Tholl D., Dudareva N., **2011**. Role of aromatic aldehyde synthase in wounding/herbivory response and flower scent production in different *Arabidopsis* ecotypes: Phenylacetaldehyde biosynthesis in *A. thaliana*. *The Plant Journal* 66, 591–602. <https://doi.org/10.1111/j.1365-313X.2011.04515.x>
- Gutensohn M., Orlova I., Nguyen T.T.H., Davidovich-Rikanati R., Ferruzzi M.G., Sitrit Y., Lewinsohn E., Pichersky E., Dudareva N., **2013**. Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J* 75, 351–363. <https://doi.org/10.1111/tbj.12212>
- Guterman I., Shalit M., Menda N., Piestun D., Dafny-Yelin M., Shalev G., Bar E., Davydov O., Ovadis M., Emanuel M., Wang J., Adam Z., Pichersky E., Lewinsohn E., Zamir D., Vainstein A., Weiss D., **2002**. Rose Scent. *The Plant Cell* 14, 2325–2338. <https://doi.org/10.1105/tpc.005207>

- Guterman I., Masci T., Chen X., Negre F., Pichersky E., Dudareva N., Weiss D., Vainstein A., **2006**. Generation of phenylpropanoid pathway-derived volatiles in transgenic plants: tose alcohol acetyltransferase produces phenylethyl acetate and benzyl acetate in *Petunia* flowers. *Plant Mol Biol* 60, 555–563. <https://doi.org/10.1007/s11103-005-4924-x>
- Hahn M.W., **2009**. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered.* 100, 605–617. <https://doi.org/10.1093/jhered/esp047>
- Hampel, D., Mosandl, A., and Wüst, M. **2006**. Biosynthesis of mono- and sesquiterpenes in strawberry fruits and foliage: 2H labelling studies. *J Agric Food Chem* 54, 1473–1478. <https://doi.org/10.1021/jf0523972>
- Hampel D., Swatski A., Mosandl A., Wüst M., **2007**. Biosynthesis of monoterpenes and norisoprenoids in raspberry Fruits (*Rubus idaeus* L.): The role of cytosolic mevalonate and plastidial methylerythritol phosphate pathway. *J. Agric. Food Chem.* 55, 9296–9304. <https://doi.org/10.1021/jf071311x>
- Han Y., Gasic K., Korban S.S., **2007**. Multiple-Copy Cluster-Type organization and evolution of genes Encoding O-Methyltransferases in the Apple. *Genetics.* 176, 2625–2635. <https://doi.org/10.1534/genetics.107.073650>
- Han Y., Korban S.S., **2007**. *Spring*: A novel family of miniature inverted-repeat transposable elements is associated with genes in apple. *Genomics.* 90, 195–200. <https://doi.org/10.1016/j.ygeno.2007.04.005>
- Harrell F., Dupont C., **2020**. Hmisc: Harrell miscellaneous R package version 4.4.2.
- Harris R.S., **2007**. Improved pairwise alignment of genomic DNA. [Ph.D. thesis]: Pennsylvania State University.
- Hemmerlin A., Harwood J.L., Bach T.J., **2012**. A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Progress in Lipid Research* 51, 95–148. <https://doi.org/10.1016/j.plipres.2011.12.001>
- Henry L.K., Gutensohn M., Thomas S.T., Noel J.P., Dudareva N., **2015**. Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10050–10055. <https://doi.org/10.1073/pnas.1504798112>
- Henry L.K., Thomas S.T., Widhalm J.R., Lynch J.H., Davis T.C., Kessler S.A., Bohlmann J., Noel J.P., Dudareva N., **2018**. Contribution of isopentenyl phosphate to plant terpenoid metabolism. *Nature Plants* 4, 721–729. <https://doi.org/10.1038/s41477-018-0220-z>
- Hibrand Saint-Oyant L., Ruttink T., Hamama L., Kirov I., Lakhwani D., Zhou N.N., Bourke P.M., Daccord N., Leus L., Schulz D., Van de Geest H., Hesselink T., Van Laere K., Debray K., Balzergue S., Thouroude T., Chastellier A., Jeauffre J., Voisine L., Gaillard S., Borm T.J.A., Arens P., Voorrips R.E., Maliepaard C., Neu E., Linde M., Le Paslier M.C., Bérard A., Bounon R., Clotault J., Choisne N., Quesneville H., Kawamura K., Aubourg S., Sakr S., Smulders M.J.M., Schijlen E., Bucher E., Debener T., De Riek J., Foucher F., **2018**. A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nature Plants* 4, 473–484. <https://doi.org/10.1038/s41477-018-0166-1>
- Hirata H., Ohnishi T., Ishida H., Tomida K., Sakai M., Hara M., Watanabe N., **2012**. Functional characterization of aromatic amino acid aminotransferase involved in 2-phenylethanol biosynthesis in isolated rose petal protoplasts. *Journal of Plant Physiology* 169, 444–451. <https://doi.org/10.1016/j.jplph.2011.12.005>
- Hirata H., Ohnishi T., Watanabe N., **2016**. Biosynthesis of floral scent 2-phenylethanol in rose flowers. *Bioscience, Biotechnology, and Biochemistry* 80, 1865–1873. <https://doi.org/10.1080/09168451.2016.1191333>
- Hivert G., Davidovich-Rikanati R., Bar E., Sitrit Y., Schaffer A., Dudareva N., Lewinsohn E., **2020**. Prenyltransferases catalyzing geranyldiphosphate formation in tomato fruit. *Plant Science* 296, 110504. <https://doi.org/10.1016/j.plantsci.2020.110504>
- Höfer R., Dong L., André F., Ginglinger J.-F., Lugan R., Gavira C., Grec S., Lang G., Memelink J., Van Der Krol S., Bouwmeester H., Werck-Reichhart D., **2013**. Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco)iridoid pathway. *Metabolic Engineering* 20, 221–232. <https://doi.org/10.1016/j.ymben.2013.08.001>
- Hosfield D.J., Zhang Y., Dougan D.R., Broun A., Tari L.W., Swanson R.V., Finn J., **2004**. Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *Journal of Biological Chemistry* 279, 8526–8529. <https://doi.org/10.1074/jbc.C300511200>
- Hsiao Y.-Y., Jeng M.-F., Tsai W.-C., Chuang Y.-C., Li C.-Y., Wu T.-S., Kuoh C.-S., Chen W.-H., Chen H.-H., **2008**. A novel homodimeric geranyl diphosphate synthase from the orchid *Phalaenopsis bellina* lacking a DD(X)₂₋₄ D motif. *The Plant Journal* 55, 719–733. <https://doi.org/10.1111/j.1365-313X.2008.03547.x>
- Hsieh F.-L., Chang T.-H., Ko T.-P., Wang A.H.-J., **2011**. Structure and mechanism of an *Arabidopsis* Medium/Long-Chain-Length prenyl pyrophosphate synthase. *Plant Physiology* 155, 1079–1090. <https://doi.org/10.1104/pp.110.168799>

- Hsieh F.L., Chang T.H., Ko T.P., Wang A.H.J., **2010**. Enhanced specificity of mint geranyl pyrophosphate synthase by modifying the R-Loop interactions. *Journal of Molecular Biology* 404, 859–873. <https://doi.org/10.1016/j.jmb.2010.10.011>
- Hsieh P., Vollger M.R., Dang V., Porubsky D., Baker C., Cantsilieris S., Hoekzema K., Lewis A.P., Munson K.M., Sorensen M., Kronenberg Z.N., Murali S., Nelson B.J., Chiatante G., Maggiolini F.A.M., Blanché H., Underwood J.G., Antonacci F., Deleuze J.-F., Eichler E.E., **2019**. Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes. *Science* 366, eaax2083. <https://doi.org/10.1126/science.aax2083>
- Huang F.C., Horváth G., Molnár P., Turcsi E., Deli J., Schrader J., Sandmann G., Schmidt H., Schwab W., **2009a**. Substrate promiscuity of RdCCD1, a carotenoid cleavage oxygenase from *Rosa damascena*. *Phytochemistry* 70, 457–464. <https://doi.org/10.1016/j.phytochem.2009.01.020>
- Huang F.C., Molnár P., Schwab W., **2009b**. Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. *Journal of Experimental Botany* 60, 3011–3022. <https://doi.org/10.1093/jxb/erp137>
- Huang S., Lee C.P., Harvey Millar A., **2015**. Activity assay for plant mitochondrial enzymes In: Whelan, J., Murcha, M. (eds) *Plant Mitochondria. Methods in Molecular Biology*, vol 1305. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-2639-8_10
- Ignea C., Pontini M., Maffei M.E., Makris A.M., Kampranis S.C., **2014**. Engineering monoterpene production in yeast using a synthetic dominant negative geranyl diphosphate synthase. *ACS Synth. Biol.* 3, 298–306. <https://doi.org/10.1021/sb400115e>
- Iijima Y., Gang D.R., Fridman E., Lewinsohn E., Pichersky E., **2004**. Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiology* 134, 370–379. <https://doi.org/10.1104/pp.103.032946>
- Iijima Y., Wang G., Fridman E., Pichersky E., **2006**. Analysis of the enzymatic formation of citral in the glands of sweet basil. *Archives of Biochemistry and Biophysics* 448, 141–149. <https://doi.org/10.1016/j.abb.2005.07.026>
- Iijima Y., Koeduka T., Suzuki H., Kubota K., **2014**. Biosynthesis of geranial, a potent aroma compound in ginger rhizome (*Zingiber officinale*): Molecular cloning and characterization of geraniol dehydrogenase. *Plant Biotechnology* 31, 525–534. <https://doi.org/10.5511/plantbiotechnology.14.1020a>
- Ilg A., Bruno M., Beyer P., Al-Babili S., **2014**. Tomato carotenoid cleavage dioxygenases 1A and 1B: Relaxed double bond specificity leads to a plenitude of dialdehydes, mono-apocarotenoids and isoprenoid volatiles. *FEBS Open Bio* 4, 584–593. <https://doi.org/10.1016/j.fob.2014.06.005>
- Inouye H., Ueda S., Inoue K., Matsumura H., **1979**. Biosynthesis of shikonin in callus cultures of *Lithospermum erythrorhizon*. *Phytochemistry* 18, 1301–1308. [https://doi.org/10.1016/0031-9422\(79\)83012-5](https://doi.org/10.1016/0031-9422(79)83012-5)
- Irmisch S., Jiang Y., Chen F., Gershenzon J., Köllner T.G., **2014**. Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*). *BMC Plant Biol* 14, 270. <https://doi.org/10.1186/s12870-014-0270-y>
- Ito M., Honda G., **2007**. Geraniol synthases from perilla and their taxonomical significance. *Phytochemistry* 68, 446–453. <https://doi.org/10.1016/j.phytochem.2006.11.006>
- Iwata H., Gaston A., Remay A., Thouroude T., Jeauffre J., Kawamura K., Oyant L.H.-S., Araki T., Denoyes B., Foucher F., **2012**. The *TFL1* homologue *KSN* is a regulator of continuous flowering in rose and strawberry. *Plant J.* 69, 116–125. <https://doi.org/10.1111/j.1365-313X.2011.04776.x>
- Iwasaki S., Takeda A., Motose H., Watanabe Y., **2007**. Characterization of *Arabidopsis* decapping proteins AtDCP1 and AtDCP2, which are essential for post-embryonic development. *FEBS Letters* 581, 2455–2459. <https://doi.org/10.1016/j.febslet.2007.04.051>
- Jemth A.-S., Scaletti E., Carter M., Helleday T., Stenmark P., **2019**. Crystal structure and substrate specificity of the 8-oxo-dGTP hydrolase NUDT1 from *Arabidopsis thaliana*. *Biochemistry* 58, 887–899. <https://doi.org/10.1021/acs.biochem.8b00950>
- Jiang N., Bao Z., Zhang X., Eddy S.R., Wessler S.R., **2004**. Pack-MULE transposable elements mediate gene evolution in plants. *Nature*. 431, 569–573. <https://doi.org/10.1038/nature02953>
- Jiang S., Wang X., Shi C., Luo J., **2019**. Genome-Wide identification and analysis of High-Copy-Number LTR retrotransposons in asian pears. *Genes* 10, 156. <https://doi.org/10.3390/genes10020156>
- Jung S., Cestaro A., Troggio M., Main D., Zheng P., Cho I., Folta K.M., Sosinski B., Abbott A., Celton J.-M., Arús P., Shulaev V., Verde I., Morgante M., Rokhsar D., Velasco R., Sargent D., **2012**. Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceous subfamilies. *BMC Genomics* 13, 129. <https://doi.org/10.1186/1471-2164-13-129>

Bibliography

- Jung S., Lee T., Cheng C.-H., Buble K., Zheng P., Yu J., Humann J., Ficklin S.P., Gasic K., Scott K., Frank M., Ru S., Hough H., Evans K., Peace C., Olmstead M., DeVetter L.W., McFerson J., Coe M., Wegrzyn J.L., Staton M.E., Abbott A.G., Main D., **2019**. 15 years of GDR: New data and functionality in the Genome Database for *Rosaceae*. *Nucleic Acids Research* 47, D1137–D1145. <https://doi.org/10.1093/nar/gky1000>
- Junker R.R., Blüthgen N., **2010**. Floral scents repel facultative flower visitors, but attract obligate ones. *Annals of Botany* 105, 777–782. <https://doi.org/10.1093/aob/mcq045>
- Jux A., Gleixner G., Boland W., **2001**. Classification of Terpenoids according to the Methylerythritolphosphate or the Mevalonate Pathway with Natural ¹²C/¹³C Isotope Ratios: Dynamic Allocation of Resources in Induced Plants. *Angewandte Chemie - International Edition* 40, 2091–2094. [https://doi.org/10.1002/1521-3773\(20010601\)40:11<2091::AID-ANIE2091>3.0.CO;2-5](https://doi.org/10.1002/1521-3773(20010601)40:11<2091::AID-ANIE2091>3.0.CO;2-5)
- Kaminaga Y., Schnepf J., Peel G., Kish C.M., Ben-Nissan G., Weiss D., Orlova I., Lavie O., Rhodes D., Wood K., Porterfield D.M., Cooper A.J.L., Schloss J.V., Pichersky E., Vainstein A., Dudareva N., **2006**. Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation. *Journal of Biological Chemistry* 281, 23357–23366. <https://doi.org/10.1074/jbc.M602708200>
- Kamran H.M., Hussain S.B., Junzhong S., Xiang L., Chen L.-Q., **2020**. Identification and Molecular Characterization of Geranyl Diphosphate Synthase (GPPS) Genes in Wintersweet Flower. *Plants* 9, 666. <https://doi.org/10.3390/plants9050666>
- Karimi M., Depicker A., Hilson P., **2007**. Recombinational cloning with plant gateway vectors. *Plant Physiol* 145, 1144–1154. <https://doi.org/10.1104/pp.107.106989>
- Kavanagh K.L., Guo K., Dunford J.E., Wu X., Knapp S., Ebetino F.H., Rogers M.J., Russell R.G.G., Oppermann U., **2006**. The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs. *PNAS* 103, 5. <https://doi.org/10.1073/pnas.0601643103>
- Keim V., Manzano D., Fernández F.J., Closa M., Andrade P., Caudepón D., Bortolotti C., Vega M.C., Arró M., Ferrer A., **2012**. Characterization of Arabidopsis FPS Isozymes and FPS Gene Expression Analysis Provide Insight into the Biosynthesis of Isoprenoid Precursors in Seeds. *PLoS ONE* 7, e49109. <https://doi.org/10.1371/journal.pone.0049109>
- Kellogg B.A., Poulter C.D., **1997**. Chain elongation in the isoprenoid biosynthetic pathway. *Current Opinion in Chemical Biology* 1, 570–578. [https://doi.org/10.1016/S1367-5931\(97\)80054-3](https://doi.org/10.1016/S1367-5931(97)80054-3)
- Klaus S.M.J., Wegkamp A., Sybesma W., Hugenholtz J., Gregory J.F., Hanson, A.D., **2005**. A Nudix enzyme removes pyrophosphate from dihydroneopterin triphosphate in the folate synthesis pathway of bacteria and plants. *J. Biol. Chem.* 280, 5274–5280. <https://doi.org/10.1074/jbc.M413759200>
- Knudsen J.T., Eriksson R., Gershenzon J., Ståhl B., **2006**. Diversity and distribution of floral scent. *Bot Rev.* 72:1–120. [https://doi.org/10.1663/0006-8101\(2006\)72\[1:DADDFS\]2.0.CO;2](https://doi.org/10.1663/0006-8101(2006)72[1:DADDFS]2.0.CO;2)
- Kobayashi S., Goto-Yamamoto N., Hirochika H., **2004**. Retrotransposon-induced mutations in grape skin color. *Science* 304: 982. <https://doi.org/10.1126/science.1095011>
- Krasileva K.V., **2019**. The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. *Curr Opin Plant Biol.* 48, 18–25. <https://doi.org/10.1016/j.pbi.2019.01.004>
- Kraszewska E., **2008**. The plant Nudix hydrolase family. *Acta Biochim Pol* 55, 663–671. https://doi.org/10.18388/abp.2008_3025
- Kreuz K., Kleinig H., **1981**. On the compartmentation of isopentenyl diphosphate synthesis and utilization in plant cells. *Planta* 153, 578–281. <https://doi.org/10.1007/BF00385544>
- Krissinel E., Henrick K., **2007**. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Krüssmann G. **1981**, The complete book of Roses. *Timber Press*, 436 pp
- Kuhn M., Jackson s., Cimentada J., **2020**. Corrr: correlations in R. R package version 0.4.3.
- Kumar S., Stecher G., Li M., Knyaz C., Tamura K., **2018**. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kumari P., Gahlaut V., Kaur E., Singh S., Kumar S., Jaiswal V., **2021**. Genome-wide characterization of GRAS transcription factors and their potential roles in development and drought resilience in iose (*Rosa chinensis*) (preprint). In Review. <https://doi.org/10.3390/horticulturae8050429>

- Kupke T., Caparrós-Martín J.A., Malquichagua Salazar K.J., Culiáñez-Macià F.A. **2009**. Biochemical and physiological characterization of *Arabidopsis thaliana* AtCoAse: a Nudix CoA hydrolyzing protein that improves plant development. *Physiol. Plant.* 135, 365-378. <https://doi.org/10.1111/j.1399-3054.2009.01205.x>
- Kurkcuoglu Z., Koukos P.I., Citro N., Trellet M.E., Rodrigues J., Moreira I.S., Roel-Touris J., Melquiond A.S.J., Geng C., Schaarschmidt J., Xue L.C., Vangone A., Bonvin A., **2018**. Performance of HADDOCK and a simple contact-based protein-ligand binding affinity predictor in the D3R grand challenge 2. *J. Comput. Aided Mol. Des.* 32, 175-185. <https://doi.org/10.1007/s10822-017-0049-y>
- Lackus N.D., Petersen N.P., Nagel R., Schmidt A., Irmisch S., Gershenzon J., Köllner T.G., **2019**. Identification and characterization of trans-isopentenyl diphosphate synthases involved in herbivory-induced volatile terpene formation in *Populus trichocarpa*. *Molecules* 24, 2408. <https://doi.org/10.3390/molecules24132408>
- Lange B.M., Wildung M.R., McCaskill D., Croteau R., **1998**. A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. *Proc. Natl. Acad. Sci. U.S.A.* 95, 2100–2104. <https://doi.org/10.1073/pnas.95.5.2100>
- Lapczynski, A., Bhatia, S.P., Letizia, C.S. and Api, A.M. **2008**. Fragrance material review on farnesol. *Food Chem. Toxicol.* 46, S149-S156. <https://doi.org/10.1016/j.fct.2008.06.046>
- Larson N.R., O'Neal S.T., Bernier U.R., Bloomquist J.R., Anderson T.D., **2020**. Terpenoid-induced feeding deterrence and antennal response of honey bees. *Insects* 11, 83. <https://doi.org/10.3390/insects11020083>
- Lee J., Cheng X., Swails J.M., Yeom M.S., Eastman P.K., Lemkul J.A., Wei S., Buckner J., Jeong J.C., Qi Y., Jo S., Pande V.S., Case D.A., Brooks C.L., MacKerell A.D., Klauda J.B., Im, W., **2016**. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* 12, 405-413. <https://doi.org/10.1021/acs.jctc.5b00935>
- Leivar P., Antolín-Llovera M., Ferrero S., Closa M., Arró M., Ferrer A., Boronat A., Campos N., **2011**. Multilevel control of *Arabidopsis* 3-hydroxy-3-methylglutaryl Coenzyme A reductase by protein phosphatase 2A. *The Plant Cell* 23, 1494–1511. <https://doi.org/10.1105/tpc.110.074278>
- Lescot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., Rouzé P., Rombauts S., **2002**. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. <https://doi.org/10.1093/nar/30.1.325>
- Li B., Dewey C.N., **2011**. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li R., Tee C.S., Jiang Y.L., Jiang X.Y., Venkatesh P.N., Sarojam R., Ye J., **2015**. A terpenoid phytoalexin plays a role in basal defense of *Nicotiana benthamiana* against Potato virus X. *Sci Rep* 5, 9682. <https://doi.org/10.1038/srep09682>
- Li W., Lybrand D.B., Xu H., Zhou F., Last R.L., Pichersky E., **2020**. A trichome-specific, plastid-localized *Tanacetum cinerariifolium* Nudix protein hydrolyzes the natural pyrethrin pesticide biosynthetic intermediate trans-chrysanthemyl diphosphate. *Front. Plant Sci.* 11, 482. <https://doi.org/10.3389/fpls.2020.00482>
- Li X., Xu Y., Shen S., Yin X., Klee H., Zhang B., Chen K., **2017**. Transcription factor CitERF71 activates the terpene synthase gene CitTPS16 involved in the synthesis of E-geraniol in sweet orange fruit. *Journal of Experimental Botany* 68, 4929–4938. <https://doi.org/10.1093/jxb/erx316>
- Lichtenthaler H.K., **2000**. Discovery of the Two Parallel Pathways for Isoprenoid Biosynthesis in Plants, *in: Discoveries In Plant Biology. WORLD SCIENTIFIC*, pp. 141–161. https://doi.org/10.1142/9789812813503_0007
- Lichtenthaler H.K., **1999**. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 50, 47–65. <https://doi.org/10.1146/annurev.arplant.50.1.47>
- Lim K.Y., Werlemark G., Matyasek R., Bringloe J.B., Sieber V., El Mokadem H., Meynet J., Hemming J., Leitch A.R., Roberts A.V., **2005**. Evolutionary implications of permanent odd polyploidy in the stable sexual, pentaploid of *Rosa canina* L. *Heredity* 94, 501–506. <https://doi.org/10.1038/sj.hdy.6800648>
- Liorzou M., Pernet A., Li S., Chastellier A., Thouroude T., Michel G., Malécot V., Gaillard S., Briée C., Foucher F., Oghina-Pavie C., Clotault J., Grapin A., **2016**. Nineteenth century French rose (*Rosa* sp.) germplasm shows a shift over time from a European to an Asian genetic background. *EXBOTJ* 67, 4711–4725. <https://doi.org/10.1093/jxb/erw269>
- Liu J., Guan Z., Liu H., Qi L., Zhang D., Zou T., Yin P., **2018**. Structural insights into the substrate recognition mechanism of *Arabidopsis* GPP-bound NUDX1 for noncanonical monoterpene biosynthesis. *Molecular Plant* 11, 218–221. <https://doi.org/10.1016/j.molp.2017.10.006>

Bibliography

- Lu H., Giordano F., Ning Z., **2016**. Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14, 265–279. <https://doi.org/10.1016/j.gpb.2016.05.004>
- Lutke-Brinkhaus F., Liedvogel B., Kleinig H., **1984**. On the biosynthesis of ubiquinones in plant mitochondria. *Eur J Biochem* 141, 537–541. <https://doi.org/10.1111/j.1432-1033.1984.tb08226.x>
- Lye Z.N., Purugganan M.D., **2019**. Copy number variation in domestication. *Trends Plant Sci.* 24, 352–365. <https://doi.org/10.1016/j.tplants.2019.01.003>
- Lynen F., Eggerer H., Henning U., Kessel I., **1958**. Farnesyl-pyrophosphat und 3-Methyl- Δ^3 -butenyl-1-pyrophosphat, die biologischen Vorstufen des Squalens. Zur Biosynthese der Terpene, III. *Angew. Chem.* 70, 738–742. <https://doi.org/10.1002/ange.19580702403>
- Lynen F., Reichert E., Rueff L., **1951**. Zum biologischen Abbau der Essigsäure VI1). “Aktivierte Essigsäure”, ihre Isolierung aus Hefe und ihre chemische Natur2). *Justus Liebigs Ann. Chem.* 574, 1–32. <https://doi.org/10.1002/jlac.19515740102>
- Magnard J.-L., Roccia A., Caissard J.-C., Vergne P., Sun P., Hecquet R., Dubois A., Hibrand-Saint Oyant L., Jullien F., Nicolè F., Raymond O., Huguet S., Baltenweck R., Meyer S., Claudel P., Jeauffre J., Rohmer M., Foucher F., Huguenev P., Bendahmane M., Baudino S., **2015**. Biosynthesis of monoterpene scent compounds in roses. *Science.* 349, 81–83. <https://doi.org/10.1126/science.aab0696>
- Magnard J.-L., Bony A.R., Bettini F., Campanaro A., Blerot B., Baudino S., Jullien F., **2018**. Linalool and linalool nerolidol synthases in roses, several genes for little scent. *Plant Physiology and Biochemistry* 127, 74–87. <https://doi.org/10.1016/j.plaphy.2018.03.009>
- Manzano D., Busquets A., Closa M., Hoyerová K., Schaller H., Kamínek M., Arró M., Ferrer A., **2006**. Overexpression of Farnesyl Diphosphate Synthase in Arabidopsis Mitochondria Triggers Light-dependent Lesion Formation and Alters Cytokinin Homeostasis. *Plant Mol Biol* 61, 195–213. <https://doi.org/10.1007/s11103-006-6263-y>
- Maron L.G., Guimarães C.T., Kirst M., Albert P.S., Birchler J.A., Bradbury P.J., Buckler E.S., Coluccio A.E., Danilova T.V., Kudrna D., Magalhaes J.V., Pineros M.A., Schatz M.C., Wing R.A., Kochian L.V., **2013**. Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. *Proc Natl Acad Sci USA.* 110, 5241–5246. <https://doi.org/10.1073/pnas.1220766110>
- Masure P., **2013**. Guide des rosiers sauvages. Paris, France: Delachaux et Niestlé.
- Masumoto N., Korin M., Ito M., **2010**. Geraniol and linalool synthases from wild species of perilla. *Phytochemistry* 71, 1068–1075. <https://doi.org/10.1016/j.phytochem.2010.04.006>
- McCoy A.J., Grosse-Kunstleve R.W., Adams, P.D., Winn M.D., Storoni L.C., Read R.J., **2007**. Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674. <https://doi.org/10.1107/S0021889807021206>
- McLennan A., **2006**. The Nudix hydrolase superfamily. *Cell. Mol. Life Sci.* 63, 123–143. <https://doi.org/10.1007/s00018-005-5386-7>
- McLennan A., **2013**. Substrate ambiguity among the nudix hydrolases: biologically significant, evolutionary remnant, or both? *Cell Mol Life Sci.* 70, 373–385. <https://doi.org/10.1007/s00018-012-1210-3>
- Meyer O., Hoeffler J.F., Grosdemange-Billiard C., Rohmer M., **2004**. Practical synthesis of 1-deoxy-D-xylulose and 1-deoxy-D-xylulose 5-phosphate allowing deuterium labelling. *Tetrahedron* 60, 12153–12162. <https://doi.org/10.1016/j.tet.2004.10.016>
- Mildvan A.S., Xia Z., Azurmendi H.F., Saraswat V., Legler P.M., Massiah M.A., Gabelli S.B., Bianchet M.A., Kang L.W., Amzel L.M., **2005**. Structures and mechanisms of Nudix hydrolases. *Arch. Biochem. Biophys.* 433, 129–143. <https://doi.org/10.1016/j.abb.2004.08.017>
- Morata J., Marin F., Payet J., Casacuberta J.M., **2018**. Plant lineage-specific amplification of Transcription Factor Binding Motifs by Miniature Inverted-Repeat Transposable Elements (MITES). *Genome Biol Evol.* 10, 1210–1220. <https://doi.org/10.1093/gbe/evy073>
- Nakamura T., Meshitsuka S., Kitagawa S., Abe N., Yamada J., Ishino T., Nakano H., Tsuzuki T., Doi T., Kobayashi Y., Fujii S., Sekiguchi M., Yamagata Y., **2010**. Structural and dynamic features of the MutT protein in the recognition of nucleotides with the mutagenic 8-oxoguanine base. *J. Biol. Chem.* 285, 444–452. <https://doi.org/10.1074/jbc.M109.066373>
- Nakamura N., Hirakawa H., Sato S., Otagaki S., Matsumoto S., Tabata S., Tanaka Y., **2017**. Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. *DNA Res.* 25, 113–121. <https://doi.org/10.1093/dnares/dsx042>

Bibliography

- Muhlemann J.K., Klempien A., Dudareva N., **2014**. Floral volatiles: from biosynthesis to function: Floral volatiles. *Plant Cell Environ* 37, 1936–1949. <https://doi.org/10.1111/pce.12314>
- Nagegowda D.A., Gupta P., **2020**. Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. *Plant Science* 294, 110457. <https://doi.org/10.1016/j.plantsci.2020.110457>
- Nagegowda D.A., Gutensohn M., Wilkerson C.G., Dudareva N., **2008**. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *The Plant Journal* 55, 224–239. <https://doi.org/10.1111/j.1365-313X.2008.03496.x>
- Narita K., Ohnuma S. -i., Nishino T., **1999**. Protein Design of Geranyl Diphosphate Synthase. Structural Features That Define the Product Specificities of Prenyltransferases. *Journal of Biochemistry* 126, 566–571. <https://doi.org/10.1093/oxfordjournals.jbchem.a022487>
- Nelson B.K., Cai X., Nebenführ A., **2007**. A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant Journal* 51, 1126–1136. <https://doi.org/10.1111/j.1365-313X.2007.03212.x>
- Nyati P., Nouzova M., Rivera-Perez C., Clifton M.E., Mayoral J.G., Noriega F.G., **2013**. Farnesyl phosphatase, a *Corpora allata* enzyme involved in juvenile hormone biosynthesis in *Aedes aegypti*. *PLOS ONE* 8, e71967. <https://doi.org/10.1371/journal.pone.0071967>
- Ogawa T., Ueda Y., Yoshimura K., Shigeoka S., **2005**. Comprehensive analysis of cytosolic Nudix hydrolases in *Arabidopsis thaliana*. *J. Biol. Chem.* 280, 25277–25283. <https://doi.org/10.1074/jbc.M503536200>
- Ogawa T., Yoshimura K., Miyake H., Ishikawa K., Ito D., Tanabe N., Shigeoka S., **2008**. Molecular characterization of organelle-type Nudix hydrolases in Arabidopsis. *Plant Physiol.* 148, 1412–1424. <https://doi.org/10.1104/pp.108.128413>
- Ohnuma S., Narita K., Nakazawa T., Ishida C., Takeuchi Y., Ohto C., Nishino T., **1996**. A Role of the Amino Acid Residue Located on the Fifth Position before the First Aspartate-rich Motif of Farnesyl Diphosphate Synthase on Determination of the Final Product. *Journal of Biological Chemistry* 271, 30748–30754. <https://doi.org/10.1074/jbc.271.48.30748>
- Okada Y., Sugimoto M., Ito K., **2001**. Molecular cloning and expression of farnesyl pyrophosphate synthase gene responsible for essential oil biosynthesis in hop (*Humulus lupulus*). *Journal of Plant Physiology* 158, 1183–1188. [https://doi.org/10.1078/S0176-1617\(04\)70145-5](https://doi.org/10.1078/S0176-1617(04)70145-5)
- Oghina-Pavie C. **2015**. Rose and Pear Breeding in Nineteenth-Century France: The Practice and Science of Diversity. In *New Perspectives on the History of Life Sciences and Agriculture*, Springer pp. 55–77. https://doi.org/10.1007/978-3-319-12185-7_4
- Oliva M., Bar E., Ovadia R., Perl A., Galili G., Lewinsohn E., Oren-Shamir M., **2017**. Phenylpyruvate Contributes to the Synthesis of Fragrant Benzenoid–Phenylpropanoids in *Petunia × hybrida* Flowers. *Front. Plant Sci.* 8, 769. <https://doi.org/10.3389/fpls.2017.00769>
- Önder D. G., önder S., Karakurt Y., **2021**. Identification and molecular characterization of geraniol and linalool synthase genes related to monoterpene biosynthesis in damask rose (*Rosa damascena* Mill.), several genes for little scent. *Journal of Essential Oil Bearing Plants* 24, 920–924. <https://doi.org/10.1080/0972060X.2021.1977721>
- Ono K., Akagi T., Morimoto T., Wunsch A., Tao R., **2018**. Genome re-sequencing of diverse sweet cherry (*Prunus avium*) individuals reveals a modifier gene mutation conferring pollen-part self-compatibility. *Plant Cell Physiol.* 59, 1265–1275. <https://doi.org/10.1093/pcp/pcy068>
- Orlova I., Nagegowda D.A., Kish C.M., Gutensohn M., Maeda H., Varbanova M., Fridman E., Yamaguchi S., Hanada A., Kamiya Y., Krichevsky A., Citovsky V., Pichersky E., Dudareva N., **2010**. The Small Subunit of Snapdragon Geranyl Diphosphate Synthase Modifies the Chain Length Specificity of Tobacco Geranylgeranyl Diphosphate Synthase in Planta. *The Plant Cell* 21, 4002–4017. <https://doi.org/10.1105/tpc.109.071282>
- Park J., Zielinski M., Magder A., Tsantrizos Y.S., Berghuis A.M., **2017a**. Human farnesyl pyrophosphate synthase is allosterically inhibited by its own product. *Nat Commun* 8, 14132. <https://doi.org/10.1038/ncomms14132>
- Park J., Zielinski M., Magder A., Tsantrizos Y.S., Berghuis A.M., **2017b**. Human farnesyl pyrophosphate synthase is allosterically inhibited by its own product. *Nat Commun* 8, 14132. <https://doi.org/10.1038/ncomms14132>
- Parveen I., Wang M., Zhao J., Chittiboyina A.G., Tabanca N., Ali A., Baerson S.R., Techen N., Chappell J., Khan I.A., Pan, Z., **2015**. Investigating sesquiterpene biosynthesis in *Ginkgo biloba*: molecular cloning and functional characterization of (*E,E*)-farnesol and α -bisabolene synthases. *Plant Mol. Biol.* 89, 451–462. <https://doi.org/10.1007/s11103-015-0381-3>

- Perry G.H., Dominy N.J., Claw K.G., Lee A.S., Fiegler H., Redon R., Werner J., Villanea F.A., Mountain J.L., Misra R., Carter N.P., Lee C., Stone A.C., **2007**. Diet and the evolution of human amylase gene copy number variation. *Nat Genet.* 39, 1256–1260. <https://doi.org/https://doi.org/10.1038/ng2123>
- Pfaffl M.W., **2001**. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, e45. <https://doi.org/10.1093/nar/29.9.e45>
- Phillips M.A., D'Auria J.C., Gershenzon J., Pichersky E., **2008**. The *Arabidopsis thaliana* Type I Isopentenyl Diphosphate Isomerases Are Targeted to Multiple Subcellular Compartments and Have Overlapping Functions in Isoprenoid Biosynthesis. *Plant Cell* 20, 677–696. <https://doi.org/10.1105/tpc.107.053926>
- Pichersky E., Lewinsohn E., Croteau R., **1995**. Purification and Characterization of S-Linalool Synthase, an Enzyme Involved in the Production of Floral Scent in *Clarkia breweri*. *Archives of Biochemistry and Biophysics* 316, 803–807. <https://doi.org/10.1006/abbi.1995.1107>
- Pickett J.A., Williams I.H., Martin A.P., Smith M.C., **1980**. Nasonov pheromone of the honey bee, *Apis mellifera* L. (HYMENOPTERA: APIDAE). *Journal of chemical ecology* 6, 425–434.
- Polichuk D.R., Zhang Y., Reed D.W., Schmidt J.F., Covello P.S., **2010**. A glandular trichome-specific monoterpene alcohol dehydrogenase from *Artemisia annua*. *Phytochemistry* 71, 1264–1269. <https://doi.org/10.1016/j.phytochem.2010.04.026>
- Prunier J., Caron S., MacKay J., **2017**. CNVs into the wild: screening the genomes of conifer trees (*Picea spp.*) reveals fewer gene copy number variations in hybrids and links to adaptation. *BMC Genomics.* 18, 97. <https://doi.org/10.1186/s12864-016-3458-8>
- R Core Team, **2015**. R: A language and environment for statistical computing. Vienna, Austria.
- Raguso R.A., **2004**. Why do flowers smell? The chemical ecology of fragrance-driven pollination. *Advances in Insect Chemical Ecology*. Cambridge: Cambridge University Press. pp 151–178.
- Rai A., Smita S.S., Singh A.K., Shanker K., Nagegowda D.A., **2013**. Heteromeric and Homomeric Geranyl Diphosphate Synthases from *Catharanthus roseus* and Their Role in Monoterpene Indole Alkaloid Biosynthesis. *Molecular Plant* 6, 1531–1549. <https://doi.org/10.1093/mp/sst058>
- Rao S., Meng X., Liao Y., Yu T., Cao J., Tan J., Xu F., Cheng S., **2019**. Characterization and functional analysis of two novel 3-hydroxy-3-methylglutaryl-coenzyme A reductase genes (GbHMGR2 and GbHMGR3) from *Ginkgo biloba*. *Sci Rep* 9, 14109. <https://doi.org/10.1038/s41598-019-50629-8>
- Raymond O., Gouzy J., Just J., Badouin H., Verdenaud M., Lemainque A., Vergne P., Moja S., Choisne N., Pont C., Carrère S., Caissard J.-C., Couloux A., Cottret L., Aury J.-M., Szécsi J., Latrasse D., Madoui M.-A., François L., Fu X., Yang S.-H., Dubois A., Piola F., Larrieu A., Perez M., Labadie K., Perrier L., Govetto B., Labrousse Y., Villand P., Bardoux C., Boltz V., Lopez-Roques C., Heitzler P., Vernoux T., Vandenbussche M., Quesneville H., Boualem A., Bendahmane A., Liu C., Le Bris M., Salse J., Baudino S., Benhamed M., Wincker P., Bendahmane M., **2018**. The *Rosa* genome provides new insights into the domestication of modern roses. *Nat Genet* 50, 772–777. <https://doi.org/10.1038/s41588-018-0110-3>
- Rehder A. **1940** Manual of cultivated Trees and Shrubs Hardy in North America. 2nd edition. 996 pp
- Richter A., Seidl-Adams I., Köllner T.G., Schaff C., Tumlinson J.H., Degenhardt J., **2015**. A small, differentially regulated family of farnesyl diphosphate synthases in maize (*Zea mays*) provides farnesyl diphosphate for the biosynthesis of herbivore-induced sesquiterpenes. *Planta* 241, 1351–1361. <https://doi.org/10.1007/s00425-015-2254-z>
- Ritz C.M., Schmutz H., Wissemann V., **2005**. Evolution by Reticulation: European Dogroses Originated by Multiple Hybridization Across the Genus *Rosa*. *Journal of Heredity* 96, 4–14. <https://doi.org/10.1093/jhered/esi011>
- Rivera S.B., Swedlund B.D., King G.J., Bell R.N., Hussey C.E., Shattuck-Eidens D.M., Wrobel W.M., Peiser G.D., Poulter C.D., **2001**. Chrysanthemyl diphosphate synthase: Isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4373–4378. <https://doi.org/10.1073/pnas.071543598>
- Robinson M.D., McCarthy D.J., Smyth G.K., **2010**. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Robinson M.D., Oshlack, A., **2010**. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- Roccia A., Hibrand-Saint Oyant L., Cavel E., Caissard J.-C., Machenaud J., Thouroude T., Jeauffre J., Bony A., Dubois A., Vergne P., Szécsi J., Foucher F., Bendahmane M., Baudino S., **2019**. Biosynthesis of 2-Phenylethanol

- in Rose Petals Is Linked to the Expression of One Allele of *RhPAAS*. *Plant Physiol.* 179, 1064–1079. <https://doi.org/10.1104/pp.18.01468>
- Rodríguez-Concepción M., Boronat A., **2002**. Elucidation of the Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis in Bacteria and Plastids. A Metabolic Milestone Achieved through Genomics. *Plant Physiology* 130, 1079–1089. <https://doi.org/10.1104/pp.007138>
- Rohdich F., Hecht S., Gärtner K., Adam P., Krieger C., Amslinger S., Arigoni D., Bacher A., Eisenreich W., **2002**. Studies on the nonmevalonate terpene biosynthetic pathway: Metabolic role of IspH (LytB) protein. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1158–1163. <https://doi.org/10.1073/pnas.032658999>
- Rohmer M., **1999**. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants†. *Nat. Prod. Rep.* 16, 565–574. <https://doi.org/10.1039/a709175c>
- Rondeau J.-M., Bitsch F., Bourgier E., Geiser M., Hemmig R., Kroemer M., Lehmann S., Ramage P., Rieffel S., Strauss A., Green J.R., Jahnke W., **2006**. Structural Basis for the Exceptional in vivo
- Ruan J.-X., Li J.-X., Fang X., Wang L.-J., Hu W.-L., Chen X.-Y., Yang C.-Q., **2016**. Isolation and characterization of three new monoterpene synthases from *Artemisia annua*. *Front. Plant Sci.* 7, 1-10. <https://doi.org/10.3389/fpls.2016.00638>
- Efficacy of Bisphosphonate Drugs. *ChemMedChem* 1, 267–273. <https://doi.org/10.1002/cmdc.200500059>
- Rusdi N.A., Goh H.H., Sabri S., Ramzi A.B., Mohd Noor N., Baharum, S.N., **2018**. Functional characterisation of new sesquiterpene synthase from the malaysian herbal plant, *Polygonum Minus*. *Molecules* 23, 1370. <https://doi.org/10.3390/molecules23061370>
- Sakai M., Hirata H., Sayama H., Sekiguchi K., Itano H., Asai T., Dohra H., Hara M., Watanabe N., **2007**. Production of 2-Phenylethanol in Roses as the Dominant Floral Scent Compound from L -Phenylalanine by Two Key Enzymes, a PLP-Dependent Decarboxylase and a Phenylacetaldehyde Reductase. *Bioscience, Biotechnology, and Biochemistry* 71, 2408–2419. <https://doi.org/10.1271/bbb.70090>
- Sakumi K., Furuichi M., Tsuzuki T., Kakuma T., Kawabata S., Maki H., Sekiguchi M., **1993**. Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA synthesis. *Journal of Biological Chemistry* 268, 23524–23530. [https://doi.org/10.1016/S0021-9258\(19\)49494-5](https://doi.org/10.1016/S0021-9258(19)49494-5)
- Šali, A., Blundell, T.L., **1993**. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815. <https://doi.org/10.1006/jmbi.1993.1626>
- Sallaud C., Rontein D., Onillon S., Jabès F., Duffé P., Giacalone C., Thoraval S., Escoffier C., Herbette G., Leonhardt N., Causse M., Tissier A., **2009**. A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z -Farnesyl Pyrophosphate in the Wild Tomato *Solanum habrochaites*. *The Plant Cell* 21, 301–317. <https://doi.org/10.1105/tpc.107.057885>
- Sapir-Mir M., Mett A., Belausov E., Tal-Meshulam S., Frydman A., Gidoni D., Eyal Y., **2008**. Peroxisomal Localization of Arabidopsis Isopentenyl Diphosphate Isomerases Suggests That Part of the Plant Isoprenoid Mevalonic Acid Pathway Is Compartmentalized to Peroxisomes. *Plant Physiology* 148, 1219–1228. <https://doi.org/10.1104/pp.108.127951>
- Sato-Masumoto N., Ito M., **2014**. Two types of alcohol dehydrogenase from *Perilla* can form citral and perillaldehyde. *Phytochemistry* 104, 12–20. <https://doi.org/10.1016/j.phytochem.2014.04.019>
- Scalliet G., Journot N., Jullien F., Baudino S., Magnard J.-L., Channelière S., Vergne P., Dumas C., Bendahmane M., Cock J.M., Hugueney P., **2002**. Biosynthesis of the major scent components 3,5-dimethoxytoluene and 1,3,5-trimethoxybenzene by novel rose O -methyltransferases. *FEBS Letters* 523, 113–118. [https://doi.org/10.1016/S0014-5793\(02\)02956-3](https://doi.org/10.1016/S0014-5793(02)02956-3)
- Scalliet G., Lionnet C., Le Behec M., Dutron L., Magnard J.-L., Baudino S., Bergougnot V., Jullien F., Chambrier P., Vergne P., Dumas C., Cock J.M., Hugueney P., **2006**. Role of Petal-Specific Orcinol O -Methyltransferases in the Evolution of Rose Scent. *Plant Physiology* 140, 18–29. <https://doi.org/10.1104/pp.105.070961>
- Scalliet G., Piola F., Douady C.J., Réty S., Raymond O., Baudino S., Bordji K., Bendahmane M., Dumas C., Cock J.M., Hugueney P., **2008**. Scent evolution in Chinese roses. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5927–5932. <https://doi.org/10.1073/pnas.0711551105>
- Scannell D.R., Wolfe K.H., **2008**. A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res.* 18, 137–147. <https://doi.org/10.1101/gr.6341207>
- Schillmiller A.L., Schauvinhold I., Larson M., Xu R., Charbonneau A.L., Schmidt A., Wilkerson C., Last R.L., Pichersky E., **2009**. Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate

- precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10865–10870. <https://doi.org/10.1073/pnas.0904113106>
- Schindler S., Bach T.J., Lichtenthaler H.K., **1985**. Differential Inhibition by Mevinolin of Prenyllipid Accumulation in Radish Seedlings. *Zeitschrift für Naturforschung C* 40, 208–214. <https://doi.org/10.1515/znc-1985-3-412>
- Schmidt A., Gershenzon J., **2008**. Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). *Phytochemistry* 69, 49–57. <https://doi.org/10.1016/j.phytochem.2007.06.022>
- Schmidt A., Waechter B., Temp U., Krekling T., Seğuin A., Gershenzon J., **2010**. A Bifunctional Geranyl and Geranylgeranyl Diphosphate Synthase Is Involved in Terpene Oleoresin Formation in *Picea abies*. *Plant Physiology* 152, 639–655. <https://doi.org/10.1104/pp.109.144691>
- Schnee C., Kollner T.G., Gershenzon J., Degenhardt J., **2002**. The maize gene *terpene synthase 1* encodes a sesquiterpene synthase catalyzing the formation of (*E*)- β -farnesene, (*E*)-nerolidol, and (*E,E*)-farnesol after herbivore damage. *Plant Physiol.* 130, 2049–2060. <https://doi.org/10.1104/pp.008326>
- Schorr P., Young M.A., **2007**. Modern Roses 12: The Comprehensive List of Roses in Cultivation Or of Historical Or Botanical Importance: American Rose Society.
- Schrödinger L.L.C., **2015**. The PyMOL Molecular Graphics System, Version 1.8.
- Schwab W., Davidovich-Rikanati R., Lewinsohn E., **2008**. Biosynthesis of plant-derived flavor compounds. *Plant J.* 54, 712–732. <https://doi.org/10.1111/j.1365-313X.2008.03446.x>
- Setoyama D., Ito R., Takagi Y., Sekiguchi M., **2011**. Molecular actions of *Escherichia coli* MutT for control of spontaneous mutagenesis. *Mutat. Res.* 707, 9–14. <https://doi.org/10.1016/j.mrfmmm.2010.12.001>
- Shalit M., Guterman I., Volpin H., Bar E., Tamari T., Menda N., Adam Z., Zamir D., Vainstein A., Weiss D., Pichersky E., Lewinsohn E., **2003**. Volatile ester formation in roses identification of an Acetyl-Coenzyme A. geraniol/citronellol acetyltransferase in developing rose petals. *Plant Physiology* 131, 1868–1876. <https://doi.org/10.1104/pp.102.018572>
- Shalit M., Shafir S., Larkov O., Bar E., Kaslassi D., Adam Z., Zamir D., Vainstein A., Weiss D., Ravid U., Lewinsohn E., **2004**. Volatile compounds emitted by rose cultivars: fragrance perception by man and honeybees. *Israel J. Plant Sci.* 52, 245–255. <https://doi.org/10.1560/P7G3-FT41-XJCP-1XFM>
- Shen M.Y., Šali A., **2006**. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15, 2507–2524. <https://doi.org/10.1110/ps.062416606>
- Sheng L., Zeng Y., Wei T., Zhu M., Fang X., Yuan X., Luo Y., Feng L., **2018**. Cloning and Functional Verification of Genes Related to 2-Phenylethanol Biosynthesis in *Rosa rugosa*. *Genes* 9, 576. <https://doi.org/10.3390/genes9120576>
- Shirai K., Hanada K., **2019**. Contribution of functional divergence through copy number variations to the inter-species and intra-species diversity in specialized metabolites. *Front Plant Sci.* 10, 1567. <https://doi.org/10.3389/fpls.2019.01567>
- Simkin A.J., Guirimand G., Papon N., Courdavault V., Thabet I., Ginis O., Bouzid S., Giglioli-Guivarc'h N., Clastre M., **2011**. Peroxisomal localisation of the final steps of the mevalonic acid pathway in planta. *Planta* 234, 903–914. <https://doi.org/10.1007/s00425-011-1444-6>
- Simkin A.J., Miettinen K., Claudel P., Burlat V., Guirimand G., Courdavault V., Papon N., Meyer S., Godet S., St-Pierre B., Giglioli-Guivarc'h N., Fischer M.J.C., Memelink J., Clastre M., **2013**. Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* 85, 36–43. <https://doi.org/10.1016/j.phytochem.2012.09.014>
- Soler E., Clastre M., Bantignies B., Marigo G., Ambid C., **1993**. Uptake of isopentenyl diphosphate by plastids isolated from *Vitis vinifera* L. cell suspensions. *Planta* 191. <https://doi.org/10.1007/BF00195689>
- Sommer S., Severin K., Camara, B., Heide L., **1995**. Intracellular localization of geranyl pyrophosphate synthase from cell cultures of *Lithospermum erythrorhizon*. *Phytochemistry* 38, 623–627. [https://doi.org/10.1016/0031-9422\(94\)00684-L](https://doi.org/10.1016/0031-9422(94)00684-L)
- Song L., Poulter C.D., **1994**. Yeast farnesyl-diphosphate synthase: site-directed mutagenesis of residues in highly conserved prenyltransferase domains I and II.. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3044–3048. <https://doi.org/10.1073/pnas.91.8.3044>
- Srouji J.R., Xu A., Park A., Kirsch J.F., Brenner S.E., **2017**. The evolution of function within the Nudix homology clan. *Proteins.* 85, 775–811. <https://doi.org/10.1002/prot.25223>

Bibliography

- Stanley Fernandez S.M., Kellogg B.A., Poulter C.D., **2000**. Farnesyl Diphosphate Synthase. Altering the Catalytic Site To Select for Geranyl Diphosphate Activity. *Biochemistry* 39, 15316–15321. <https://doi.org/10.1021/bi0014305>
- Street I.P., Coffman H.R., Poulter C.D., Baker J.A., **1994**. Identification of Cys139 and Glu207 As Catalytically Important Groups in the Active Site of Isopentenyl Diphosphate:Dimethylallyl Diphosphate Isomerase. *Biochemistry* 33, 4212–4217. <https://doi.org/10.1021/bi00180a014>
- Sun P., Schuurink R.C., Caissard J.C., Huguene P., Baudino S., **2016**. My Way: Noncanonical biosynthesis pathways for plant volatiles. *Trends Plant Sci.* 21, 884–894. <https://doi.org/10.1016/j.tplants.2016.07.007>
- Sun P., Dégut C., Réty S., Caissard J., Hibrand-Saint Oyant L., Bony A., Paramita S.N., Conart C., Magnard J., Jeauffre J., Abd-El-Halim A.M., Marie-Magdelaine J., Thouroude T., Baltenweck R., Tisné C., Foucher F., Haring M., Huguene P., Schuurink R.C., Baudino S., **2020**. Functional diversification in the *Nudix hydrolase* gene family drives sesquiterpene biosynthesis in *Rosa × wichurana*. *Plant J.* 104, 185–199. <https://doi.org/10.1111/tpj.14916>
- Suttiyut T., Auber R.P., Ghaste M., Kane C.N., McAdam S.A.M., Wisecaver J.H., Widhalm J.R., **2022**. Integrative analysis of the shikonin metabolic network identifies new gene connections and reveals evolutionary insight into shikonin biosynthesis. *Hortic Res* 9, uhab087. <https://doi.org/10.1093/hr/uhab087>
- Takahashi S., Zhao Y., O'maille P.E., Greenhagen B.T., Noel J.P., Coates R.M., Chappell J., **2005**. Kinetic and molecular analysis of 5-epiaristolochene 1,3-dihydroxylase, a cytochrome P450 enzyme catalyzing successive hydroxylations of sesquiterpenes. *J Biol Chem* 280, 3686–3696. <https://doi.org/10.1074/jbc.M411870200>
- Tarshis L.C., Proteau P.J., Kellogg B.A., Sacchetti J.C., Poulter C.D., **1996**. Regulation of product chain length by isoprenyl diphosphate synthases. *Proc. Natl. Acad. Sci. U.S.A.* 93, 15018–15023. <https://doi.org/10.1073/pnas.93.26.15018>
- Tarshis L.C., Yan M., Poulter C.D., Sacchetti J.C., **1994**. Crystal Structure of Recombinant Farnesyl Diphosphate Synthase at 2.6-Å Resolution. *Biochemistry* 33, 10871–10877. <https://doi.org/10.1021/bi00202a004>
- Tavormina P.A., Gibbs M. H., HUFF J. W., **1956**. The utilization of β-hydroxy-β-methyl-δ-valerolactone in cholesterol biosynthesis. *Journal of the American Chemical Society* 78, 4498–4499.
- Thabet I., Guirimand G., Courdavault V., Papon N., Godet S., Dutilleul C., Bouzid S., Giglioli-Guivarc'h N., Clastre M., Simkin A.J., **2011**. The subcellular localization of periwinkle farnesyl diphosphate synthase provides insight into the role of peroxisome in isoprenoid biosynthesis. *Journal of Plant Physiology* 168, 2110–2116. <https://doi.org/10.1016/j.jplph.2011.06.017>
- Tholl D., **2015**. Biosynthesis and biological functions of terpenoids in plants. *Adv Biochem Eng Biotechnol* 148, 63–106. https://doi.org/10.1007/10_2014_295
- Tholl D., Chen F., Petri J., Gershenzon J., Pichersky E., **2005**. Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers: Sesquiterpene synthases in Arabidopsis flower. *The Plant Journal* 42, 757–771. <https://doi.org/10.1111/j.1365-313X.2005.02417.x>
- Tholl D., Kish C.M., Orlova I., Sherman D., Gershenzon J., Pichersky E., Dudareva N., **2004**. Formation of Monoterpenes in *Antirrhinum majus* and *Clarkia breweri* Flowers Involves Heterodimeric Geranyl Diphosphate Synthases. *Plant Cell* 16, 977–992. <https://doi.org/10.1105/tpc.020156>
- Thompson J.D., Gibson T.J., Higgins D.G., **2002**. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. Chapter 2, Unit 2.3. <https://doi.org/10.1002/0471250953.bi0203s00>
- Tieman D., Taylor M., Schauer N., Fernie A.R., Hanson A.D., Klee H.J., **2006**. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8287–8292. <https://doi.org/10.1073/pnas.0602469103>
- Tieman D.M., Loucas H.M., Kim J.Y., Clark D.G., Klee H.J., **2007**. Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry* 68, 2660–2669. <https://doi.org/10.1016/j.phytochem.2007.06.005>
- Tong W.F. Lin S.W., **1988**. Purification and properties of rice alcohol dehydrogenase. *Bot. Bull. Academia Sinica* 29, 245–253.
- Torrens-Spence M.P., Liu P., Ding H., Harich K., Gillaspay G., Li J., **2013**. Biochemical Evaluation of the Decarboxylation and Decarboxylation-Deamination Activities of Plant Aromatic Amino Acid Decarboxylases. *Journal of Biological Chemistry* 288, 2376–2387. <https://doi.org/10.1074/jbc.M112.401752>

Bibliography

- Treharne K., Mercer E., Goodwin T., **1966**. Incorporation of [¹⁴C]carbon dioxide and [2-¹⁴C]mevalonic acid into terpenoids of higher plants during chloroplast development. *Biochemical Journal* 99, 239–245. <https://doi.org/10.1042/bj0990239>
- Trhlin M., Rajchard J., **2011**. Chemical communication in the honeybee (*Apis mellifera* L.): a review. *Veterinarni Medicina*. 56, 265–273.
- Tritsch D., Hemmerlin A., Bach T.J., Rohmer M., **2010**. Plant isoprenoid biosynthesis via the MEP pathway: In vivo IPP/DMAPP ratio produced by (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase in tobacco BY-2 cell cultures. *FEBS Letters* 584, 129–134. <https://doi.org/10.1016/j.febslet.2009.11.010>
- Ueoka H., Sasaki K., Miyawaki T., Ichino T., Tatsumi K., Suzuki S., Yamamoto H., Sakurai N., Suzuki H., Shibata D., Yazaki K., **2020**. A Cytosol-localized geranyl diphosphate synthase from *Lithospermum erythrorhizon* and its molecular evolution. *Plant Physiol.* 182, 1933–1945. <https://doi.org/10.1104/pp.19.00999>
- Van Schie C.C.N., Ament K., Schmidt A., Lange T., Haring M.A., Schuurink R.C., **2007**. Geranyl diphosphate synthase is required for biosynthesis of gibberellins: Gibberellin precursor biosynthesis. *The Plant Journal* 52, 752–762. <https://doi.org/10.1111/j.1365-3113X.2007.03273.x>
- Vandermoten S., Charlotheaux B., Santini S., Sen S.E., Béliveau C., Vandenberg M., Francis F., Brasseur R., Cusson M., Haubruge É., **2008**. Characterization of a novel aphid prenyltransferase displaying dual geranyl/farnesyl diphosphate synthase activity. *FEBS Lett* 582, 1928–1934. <https://doi.org/10.1016/j.febslet.2008.04.043>
- Vandermoten S., Santini S., Haubruge É., Heuze F., Francis F., Brasseur R., Cusson M., Charlotheaux B., **2009**. Structural features conferring dual Geranyl/Farnesyl diphosphate synthase activity to an aphid prenyltransferase. *Insect Biochemistry and Molecular Biology* 39, 707–716. <https://doi.org/10.1016/j.ibmb.2009.08.007>
- Verde I., Abbott A.G., Scalabrin S., Jung S., Shu S., Marroni F., Zhebentyayeva T., Dettori M.T., Grimwood J., Cattonaro F., et al., **2013**. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet.* 45, 487–494.
- Verde I., Jenkins J., Dondini L., Micali S., Pagliarani G., Vendramin E., Paris R., Aramini V., Gazza L., Rossini L., Bassi D., Troggio M., Shu S., Grimwood J., Tartarini S., Dettori M.T., Schmutz J., **2017**. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics.* 18, 225. <https://doi.org/10.1186/s12864-017-3606-9>
- Vilanova S., Sargent D.J., Arús P., Monfort A., **2008**. Synteny conservation between two distantly-related *Rosaceae* genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). *BMC Plant Biol* 8, 67. <https://doi.org/10.1186/1471-2229-8-67>
- Vranová E., Coman D., Gruissem W., **2012**. Structure and Dynamics of the Isoprenoid Pathway Network. *Molecular Plant* 5, 318–333. <https://doi.org/10.1093/mp/sss015>
- Wang A., Yamakake J., Kudo H., Wakasa Y., Hatsuyama Y., Igarashi M., Kasai A., Li T., Harada T., **2009**. Null mutation of the *MdACS3* gene, coding for a ripening-specific 1-aminocyclopropane-1-carboxylate synthase, leads to long shelf life in apple fruit. *Plant Physiol.* 151, 391–399. <https://doi.org/10.1104/pp.109.135822>
- Wang C., Chen Q., Fan D., Li J., Wang G., Zhang P., **2016**. Structural Analyses of Short-Chain Prenyltransferases Identify an Evolutionarily Conserved GFPPS Clade in Brassicaceae Plants. *Molecular Plant* 9, 195–204. <https://doi.org/10.1016/j.molp.2015.10.010>
- Wang G., Dixon R.A., **2009**. Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9914–9919. <https://doi.org/10.1073/pnas.0904069106>
- Wang L., Peng Q., Zhao J., Ren F., Zhou H., Wang W., Liao L., Owiti A., Jiang Q., Han Y., **2016**. Evolutionary origin of *Rosaceae*-specific active non-autonomous *hAT* elements and their contribution to gene regulation and genomic structural variation. *Plant Mol Biol.* 91, 179–191. <https://doi.org/10.1007/s11103-016-0454-y>
- Weber D.J., Bhatnagar S.K., Bullions L.C., Bessman M.J., Mildvan A.S., **1992**. NMR and isotopic exchange studies of the site of bond cleavage in the MutT reaction. *Journal of Biological Chemistry* 267, 16939–16942. [https://doi.org/10.1016/S0021-9258\(18\)41875-3](https://doi.org/10.1016/S0021-9258(18)41875-3)
- Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A.H., **2007**. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 8, 973–982. <https://doi.org/10.1038/nrg2165>
- Whittington D.A., Wise M.L., Urbansky M., Coates R.M., Croteau R.B., Christianson D.W., **2002**. Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15375–15380. <https://doi.org/10.1073/pnas.232591099>

- Widrechner M.P., **1981**. History and Utilization of *Rosa damascena* L. *ECONOMIC BOTANY* 35, 17.
- Williams D.C., McGarvey D.J., Katahira E.J., Croteau R., **1998**. Truncation of Limonene Synthase Preprotein Provides a Fully Active 'Pseudomature' Form of This Monoterpene Cyclase and Reveals the Function of the Amino-Terminal Arginine Pair. *Biochemistry* 37, 12213–12220. <https://doi.org/10.1021/bi980854k>
- Wissemann V., **2003**. Conventional Taxonomy (Wild Roses). *Encyclopedia of Rose Science*, pp 111-117. <https://doi.org/10.1016/B978-0-12-809633-8.05017-2>
- Wolf D. E., Hoffman C. H., Aldrich P. E., Skeggs H. R., Wright L. D., Folkers K., **1956**. Hydroxy-B-methyl-delta-dihydroxy-B-methylvaleric acid (divalonic acid), a new biological factor. *Journal of the American Chemical Society* 78, 4499–4499.
- Wu S., Watanabe N., Mita S., Dohra H., Ueda Y., Shibuya M., Ebizuka Y., **2004**. The Key Role of Phloroglucinol O-Methyltransferase in the Biosynthesis of *Rosa chinensis* Volatile 1,3,5-Trimethoxybenzene. *Plant Physiology* 135, 95–102. <https://doi.org/10.1104/pp.103.037051>
- Xiang Y., Huang C.-H., Hu Y., Wen J., Li S., Yi T., Chen H., Xiang J., Ma H., **2017**. Evolution of *Rosaceae* fruit types based on nuclear phylogeny in the context of geological times and genome Duplication. *Mol Biol Evol* msw242. <https://doi.org/10.1093/molbev/msw242>
- Xu H., Bohman B., Wong D.C.J., Rodriguez-Delgado C., Scaffidi A., Flematti G.R., Phillips R.D., Pichersky E., Peakall R., **2017**. Complex sexual deception in an orchid is achieved by Co-opting two independent biosynthetic pathways for pollinator attraction. *Current Biology* 27, 1867-1877.e5. <https://doi.org/10.1016/j.cub.2017.05.065>
- Yan H., Baudino S., Caissard J.-C., Florence Nicolè, Zhang H., Tang K., Li S., Lu S., **2018**. Functional characterization of the eugenol synthase gene (RcEGS1) in rose. *Plant Physiology and Biochemistry* 129, 21–26. <https://doi.org/10.1016/j.plaphy.2018.05.015>
- Yan Y., Li M., Zhang X., Kong W., Bendahmane M., Bao M., Fu X., **2022**. Tissue-Specific Expression of the Terpene Synthase Family Genes in *Rosa chinensis* and Effect of Abiotic Stress Conditions. *Genes* 13, 547. <https://doi.org/10.3390/genes13030547>
- Yang T., Li J., Wang H., Zeng Y., **2005**. A geraniol-synthase gene from *Cinnamomum tenuipilum*. *Phytochemistry* 66, 285–293. <https://doi.org/10.1016/j.phytochem.2004.12.004>
- Yoshimura K., Ogawa T., Ueda Y., Shigeoka S., **2007**. AtNUDX1, an 8-Oxo-7,8-Dihydro-2'-Deoxyguanosine 5'-Triphosphate Pyrophosphohydrolase, is Responsible for Eliminating Oxidized Nucleotides in Arabidopsis. *Plant and Cell Physiology* 48, 1438–1449. <https://doi.org/10.1093/pcp/pcm112>
- Yoshimura K., Shigeoka S., **2015**. Versatile physiological functions of the Nudix hydrolase family in Arabidopsis. *Bioscience, Biotechnology, and Biochemistry* 79, 354–366. <https://doi.org/10.1080/09168451.2014.987207>
- Yuan W. B., Wu Q. D., Wu X. R., **2011**. Analysis of the constituents of volatile oil from *Fructus auranti* immaturus by GC-MS. *Journal of Chinese Medicinal Materials* 34, 1067–1069.
- Zeidler J., Schwender J., Müller C., Wiesnerb J., **1998**. Inhibition of the Non-Mevalonate l-Deoxy-D-xylulose-5-phosphate Pathway of Plant Isoprenoid Biosynthesis by Fosmidomycin. *Zeitschrift für Naturforschung C* 53, 980–986. <https://doi.org/10.1515/znc-1998-11-1208>
- Zhang A., Xiong Y., Fang F., Jiang X., Wang T., Liu K., Peng H., Zhang X., **2022**. Diversity and functional evolution of terpene synthase in *Rosaceae*. *Plants* 11, 736. <https://doi.org/10.3390/plants11060736>
- Zhang L., Hu J., Han X., Li J., Gao Y., Richards C.M., Zhang C., Tian Y., Liu G., Gul H., Wang D., Tian Y., Yang C., Meng M., Yuan G., Kang G., Wu Y., Wang K., Zhang H., Wang D., Cong P., **2019**. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun.* 10, 1494. <https://doi.org/10.1038/s41467-019-09518-x>
- Zhang S.-D., Jin J.-J., Chen S.-Y., Chase M.W., Soltis D.E., Li H.-T., Yang J.-B., Li D.-Z., Yi T.-S., **2017**. Diversification of *Rosaceae* since the Late Cretaceous based on plastid phylogenomics. *New Phytol.* 214, 1355–1367. <https://doi.org/10.1111/nph.14461>
- Zhao D., Ferguson A.A., Jiang N., **2016**. What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochim Biophys Acta.* 1859, 366–380. <https://doi.org/10.1016/j.bbagr.2015.12.005>
- Zhou F., Pichersky E., **2020**. The complete functional characterisation of the terpene synthase family in tomato. *New Phytol* 226, 1341–1360. <https://doi.org/10.1111/nph.16431>
- Zhu Z.-M., Gao X.-F., Fougère-Danezan M., **2015**. Phylogeny of *Rosa* sections *Chinenses* and *Synstylae* (*Rosaceae*) based on chloroplast and nuclear markers. *Molecular Phylogenetics and Evolution* 87, 50–64. <https://doi.org/10.1016/j.ympev.2015.03.014>

Bibliography

Żmieńko A., Samelak A., Kozłowski P., Figlerowicz M., 2014. Copy number polymorphism in plant genomes. *Theor Appl Genet.* 127, 1–18. <https://doi.org/10.1007/s00122-013-2177-7>

Titre : Origine et évolution de la biosynthèse du géraniol dans les pétales de rose

Mots clés : rose, géraniol, biosynthèses, évolution, substrats

Résumé :

La fleur de rose (*Rosa sp*) émet de nombreux composés organiques volatils (COV) appartenant à des familles chimiques d'origines variées telles que des phenylpropanoïdes, des dérivés d'acides gras ou des terpènes. Parmi les terpènes produits et émis, le géraniol est un composé majoritaire chez les roses modernes caractérisées avec un parfum fort. Contrairement aux autres plantes qui utilisent des terpènes synthases pour produire le géraniol, il a été découvert qu'une Nudix hydrolase (NUDX1) était impliquée dans la biosynthèse du géraniol en déphosphorylant le geranyl diphosphate (GPP) en geranyl monophosphate (Magnard *et al.* 2015). L'objectif de cette thèse était de mieux comprendre l'implication des NUDX1 dans la biosynthèse des terpènes et tenter de retracer une histoire évolutive concernant l'apparition de cette voie de biosynthèse cytosolique ayant abouti à la production de COV chez la rose. Pour ce faire, un inventaire des *NUDX1* présent dans les génomes de référence publiés a été réalisé et a permis d'identifier et de caractériser une autre NUDX1 de rose impliquée dans la biosynthèse de (*E,E*)-farnésol. Grâce à un échantillonnage de roses sauvages il a également été possible de retracer une partie de l'histoire évolutive de *NUDX1-1a*, la copie de *NUDX1* de rose exprimée dans les pétales responsable de la première étape de biosynthèse du géraniol. Enfin, avec la découverte de l'origine de *NUDX1-1a* dans le genre *rosa*, la question concernant l'origine du GPP cytosolique a été abordée afin d'identifier le gène responsable. Il a été également possible de retracer l'origine chez les *Rosaceae* de cette fonction afin de compléter la connaissance des différents événements ayant permis l'apparition d'une voie de biosynthèse de terpènes non canonique chez la rose.

Title: Origin and evolution of geraniol biosynthesis in rose petals

Key words: rose, geraniol, biosynthesis, evolution, substrates

Summary:

Rose flower (*Rosa sp*) emits many volatile organic compounds (VOCs) belonging to chemical families of various origins such as phenylpropanoids, fatty acid derivatives or terpenes. Among the terpenes produced and emitted, geraniol is a major compound in modern roses characterized with a strong fragrance. Unlike other plants that use terpene synthases to produce geraniol, a Nudix hydrolase (NUDX1) was found to be involved in geraniol biosynthesis by dephosphorylating geranyl diphosphate (GPP) to geranyl monophosphate (Magnard *et al.* 2015). The objective of this thesis was to better understand the involvement of NUDX1 in the biosynthesis of terpenes and to attempt to trace an evolutionary history concerning the appearance of this cytosolic biosynthetic pathway that led to the production of VOCs in roses. To do this, an inventory of *NUDX1* present in the published reference genomes was carried out and made it possible to identify and characterize another rose *NUDX1* involved in the biosynthesis of (*E,E*)-farnesol. Thanks to a sampling of wild roses it was also possible to trace part of the evolutionary history of *NUDX1-1a*, the copy of rose *NUDX1* expressed in the petals responsible for the first step of geraniol biosynthesis. Finally, with the discovery of the origin of *NUDX1-1a* in the genus *rosa*, the question regarding the origin of cytosolic GPP was addressed in order to identify the responsible gene. It was also possible to trace the origin in the *Rosaceae* of this function in order to complete the knowledge of the different events that allowed the appearance of a non-canonical terpene biosynthetic pathway in the rose.