



**HAL**  
open science

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

Pulu Sun, Clément Dégut, Stéphane Réty, Jean Claude Caissard, Laurence Hibrand-Saint Oyant, Raymonde Baltenweck, Aurélie Bony, Saretta N Paramita, Corentin Conart, Jean-louis Magnard, et al.

### ► To cite this version:

Pulu Sun, Clément Dégut, Stéphane Réty, Jean Claude Caissard, Laurence Hibrand-Saint Oyant, et al.. Functional diversification in the Nudix hydrolase gene family drives sesquiterpene biosynthesis in *Rosa x wichurana*. *The Plant Journal*, 2020, 104 (1), pp.185-199. 10.1111/tpj.14916 . hal-02990985

**HAL Id: hal-02990985**










**<https://hal.science/hal-02990985v1>**

Submitted on 13 Nov 2020

**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.

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

Pulu Sun<sup>1,2</sup> , Clément Déгут<sup>3</sup> , Stéphane Réty<sup>4</sup> , Jean-Claude Caissard<sup>1</sup> , Laurence Hibrand-Saint Oyant<sup>5</sup> , Aurélie Bony<sup>1</sup>, Saretta N. Paramita<sup>1</sup>, Corentin Conart<sup>1</sup>, Jean-Louis Magnard<sup>1</sup>, Julien Jeauffre<sup>5</sup>, Ahmed M. Abd-El-Haliem<sup>2</sup> , Jordan Marie-Magdelaine<sup>5</sup>, Tatiana Thouroude<sup>5</sup>, Raymonde Baltenweck<sup>6</sup>, Carine Tisé<sup>3</sup> , Fabrice Foucher<sup>5</sup> , Michel Haring<sup>2</sup>, Philippe Huguéney<sup>6</sup> , Robert C. Schuurink<sup>2</sup>  and Sylvie Baudino<sup>1\*</sup> 

<sup>1</sup>Univ Lyon, UJM-Saint-Etienne, CNRS, BVpam FRE 3727, Saint-Etienne F-42023, France,

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

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

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

<sup>5</sup>IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, Beaucouzé 49071, France, and

<sup>6</sup>Université 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* × *wichurana*, *RwNUDX1-2*. In order to study the function of the *RwNUDX1-2* protein, we analyzed the volatile profiles of an F<sub>1</sub> 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 (C<sub>10</sub> 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- $\alpha$  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*, RwnNUDX1-2. Unlike RhNUDX1, which is involved in monoterpene biosynthesis, RwnNUDX1-2 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**

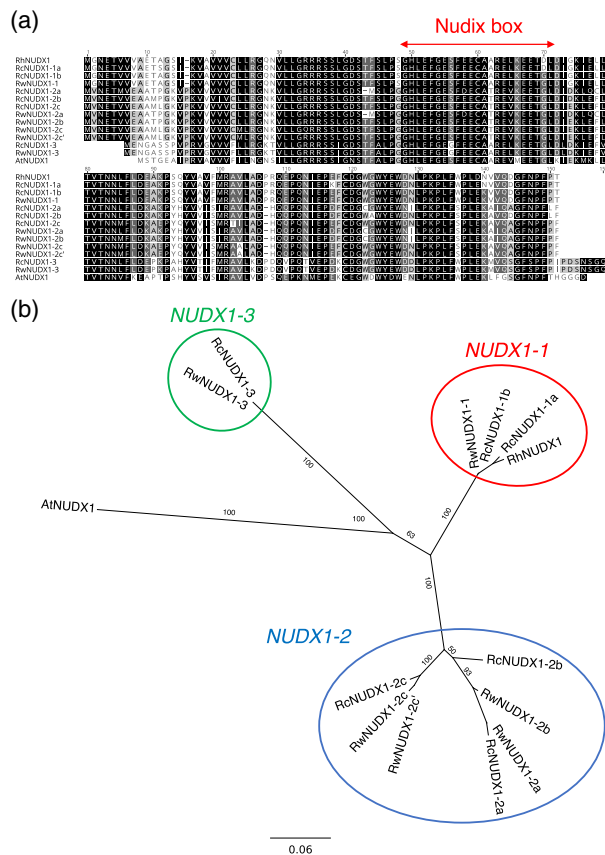
In order to further investigate the functions of Nudix hydrolases 1 in roses, we generated RNA-seq reads from four different rose cultivars: *R. chinensis* cv. 'Old Blush' (OB), *R. x wichurana* (Rw), and two individuals (OW9035 and OW9047) obtained from a crossing between OB and Rw. Next, we pooled 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 Meilland' (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- $\beta$ -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*<sub>1</sub> 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 (*OW*9007, *OW*9011, *OW*9013, *OW*9021, *OW*9024, *OW*9047, *OW*9074, *OW*9099, *OW*9149, and *OW*9204), while *NUDX1-2* was expressed in *Rw* and in 10 hybrids of the progeny (*OW*9007, *OW*9013, *OW*9018, *OW*9021, *OW*9024, *OW*9035, *OW*9037, *OW*9082, *OW*9099, and *OW*9149) (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 *OW*9099 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 Meilland’ (*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 Gx5Ex7REUxEEExGU (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.

**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 <sup>a</sup>	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

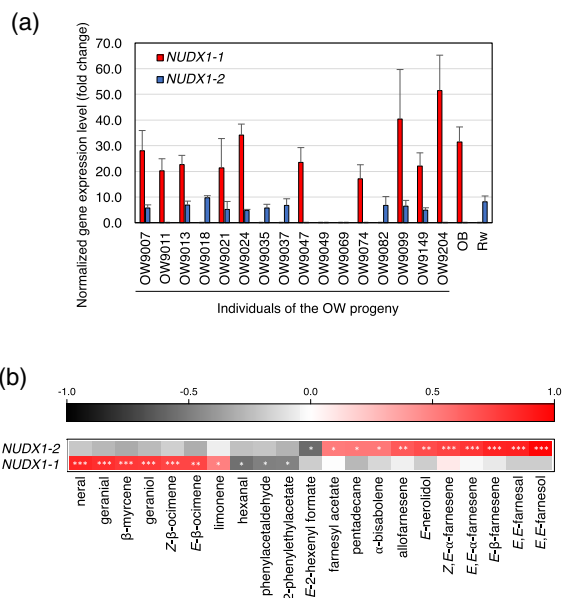
TMB, 1,3,5-trimethoxybenzene; 2PE, 2-phenylethanol; 2PEA, 2-phenylethyl acetate; nd, not detected.

<sup>a</sup>Values 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 ( $P \leq 0.01$ ), allofarnesene ( $P \leq 0.01$ ), 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

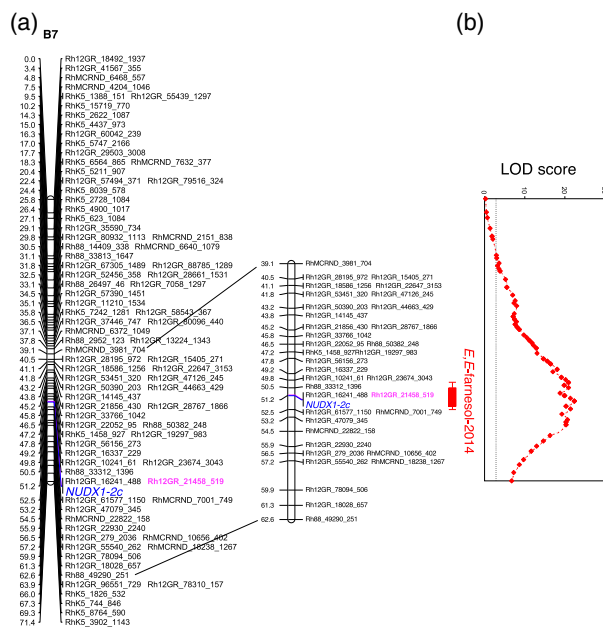
We detected a major QTL for geraniol biosynthesis in the OW progeny on the female linkage group 2, which co-localized with *RhNUDX1* (Magnard *et al.*, 2015). In order to further support the hypothesis that *RwNUDX1-2* was



**Figure 2.** (a) Gene expression levels of *NUDX1-1* and *NUDX1-2* in 16 F<sub>1</sub> 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<sup>®</sup> 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*.

**Table 2** Kinetic parameters of NUDX1 proteins: RcNUDX1-1a, RcNUDX1-1b, RhNUDX1, RwnNUDX1-2c, and AtNUDX1, with several potential substrates

Protein	Substrates	Activities <sup>a</sup>	$K_M$ (M)	$k_{cat}$ (sec <sup>-1</sup> )	$k_{cat}/K_M$ (M <sup>-1</sup> sec <sup>-1</sup> )
RcNUDX1-1a	GPP	+	$1.96 \times 10^{-6}$ ( $1.36 \times 10^{-7}$ )	0.20 (0.09)	$1.02 \times 10^5$
	FPP	+	$4.00 \times 10^{-6}$ ( $1.01 \times 10^{-6}$ )	0.73 (0.48)	$1.82 \times 10^5$
	IPP	+	n.c. <sup>b</sup>	n.c.	n.c.
	dGTP	-	n.d. <sup>c</sup>	n.d.	n.d.
	8-oxo-dGTP	-	n.d.	n.d.	n.d.
RcNUDX1-1b	GPP	+	$2.61 \times 10^{-7}$ ( $1.10 \times 10^{-7}$ )	0.30 (0.09)	$1.15 \times 10^6$
	FPP	+	$5.55 \times 10^{-7}$ ( $6.86 \times 10^{-8}$ )	1.37 (0.81)	$2.47 \times 10^6$
	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	+	$1.13 \times 10^{-6}$ ( $4.13 \times 10^{-7}$ )	0.29 (0.11)	$2.57 \times 10^5$
	FPP	+	$1.54 \times 10^{-6}$ ( $1.88 \times 10^{-7}$ )	1.21 (0.74)	$7.86 \times 10^5$
	IPP	+	n.c.	n.c.	n.c.
	dGTP	-	n.d.	n.d.	n.d.
	8-oxo-dGTP	-	n.d.	n.d.	n.d.
RwnNUDX1-2	GPP	+	$8.78 \times 10^{-6}$ ( $9.01 \times 10^{-7}$ )	0.21 (0.09)	$2.39 \times 10^4$
	FPP	+	$4.95 \times 10^{-7}$ ( $7.90 \times 10^{-8}$ )	1.67 (1.51)	$3.37 \times 10^6$
	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	+	$1.38 \times 10^{-7}$ ( $4.86 \times 10^{-8}$ )	0.26 (0.10)	$1.88 \times 10^6$
	FPP	+	$4.84 \times 10^{-7}$ ( $1.05 \times 10^{-7}$ )	1.91 (0.45)	$3.95 \times 10^6$
	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.

<sup>a</sup>Activities indicate the interaction between the protein and the substrate, + means interaction was detected and - means no interaction was detected.

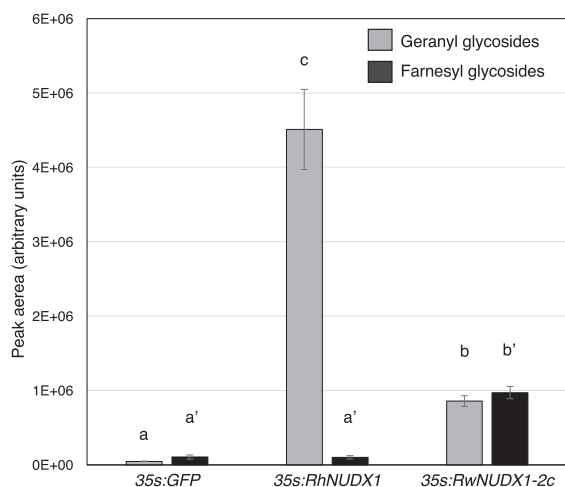
<sup>b</sup>n.c., not calculated, which means that there was a detectable activity but data could not be used for calculation.

<sup>c</sup>n.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, RwnNUDX1-2c, 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 $\alpha$ , 0.19 Å) (Figure 5a). The X-loop (amino acids from position 86 to 90), which connects  $\beta$ 5 and  $\beta$ 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).

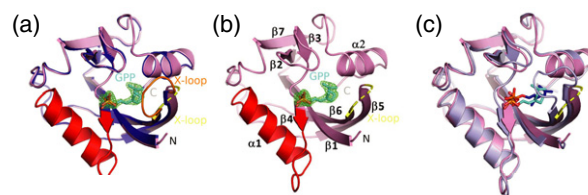


**Figure 4.** Patterns of terpenol glycoside accumulation following transient expression of *NUDX1* genes in *N. benthamiana*. Leaves of *N. benthamiana* were transformed with a *NUDX1* construct (*35S:RwNUDX1-2c* or *35:RhNUDX1*) or with the *35S:GFP* control. For each construct, eight independent biological replicates were used to quantify relative amounts of terpenol glycosides. Relative amounts are given as mean peak areas corresponding to the  $[C_{10}H_{17}]^+$  ion ( $m/z$  137.1325) for geranyl glycosides and to the  $[C_{15}H_{24}]^+$  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 $\alpha$  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 Mg $^{2+}$  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.

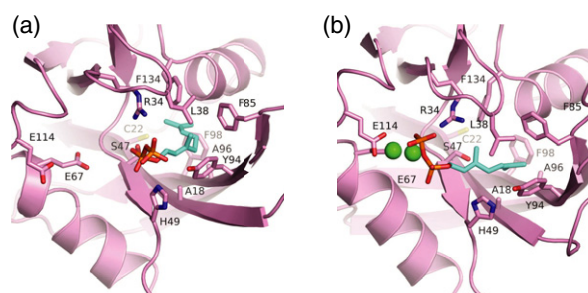
#### Modeling of rose NUDX1 proteins

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  $\sigma$  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  $\beta$ 4-loop- $\alpha$ 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 Mg $^{2+}$  ions in the active site, stabilizing the interaction with their substrates. These structures were used to position Mg $^{2+}$  ions and ligands in the active sites of NUDX1 proteins. Molecular dynamics (MD) trajectories were calculated for 20 ns and were analyzed for C $\alpha$  RMSD, protein residue fluctuation (RMSD), ligand position RMSD, the distance between Mg $^{2+}$  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. Mg $^{2+}$  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 Mg<sup>2+</sup> coordination sites, confirming that in the crystal structures of RhNUDX1/GPP and AtNUDX1/GPP (5GP0), GPP is too far from the Mg<sup>2+</sup> coordination sites for hydrolysis. Thus, MD provides a model for the RhNUDX1/Mg<sup>2+</sup>/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 NUDX1. 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 Rwn, whose petals emit significant amounts of *E,E*-farnesol and other sesquiterpenes. To this end, we characterized F<sub>1</sub> individuals in a cross between OB and Rwn, 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 Rwn, 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*- $\alpha$ -farnesene and *E*- $\beta$ -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 $\alpha$  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

(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 F<sub>1</sub> 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<sup>®</sup> 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<sup>®</sup> III reverse transcriptase, oligo-dT, and RNasOUT<sup>™</sup> 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<sup>™</sup> Real-Time system equipped with C1000<sup>™</sup> Thermal Cycler (Bio-Rad, California, USA). The qRT-PCR reaction mixture consisted of 10 µl of SsoAdvanced<sup>™</sup> SYBR<sup>®</sup> 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<sub>q</sub>) values were the mean of four values from two biological and two technical replicates. These C<sub>q</sub> values were automatically determined by the CFX96<sup>™</sup> 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<sub>t</sub> 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<sup>®</sup> 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  $\mu$ l volume with 10 ng genomic DNA, 1  $\times$  Q5<sup>®</sup> buffer, 0.200 mM dNTPs, 0.5  $\mu$ M of each primer, and 0.02 U of Q5<sup>®</sup> 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<sup>®</sup> technique, where the NUDX1 protein was fused with the protein NusA (Busso *et al.*, 2005). The final constructs were transformed into the bacterial strain Novagen<sup>®</sup> *Rosetta* (Merck, Darmstadt, Germany). Native proteins (~16 kDa) were obtained by using TALON<sup>®</sup> 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  $\mu$ M. The proteins and the substrates were incubated in reaction buffer (50 mM Tris-HCl [pH 8.8], 5 mM MgCl<sub>2</sub>, 14 mM  $\beta$ -mercaptoethanol, 10% glycerol, v/v) with a final volume of 100  $\mu$ 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  $\mu$ l of EDTA (100 mM, pH 8) and 90  $\mu$ l of ethanol/0.5% NH<sub>4</sub>OH 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  $\mu$ M in reaction buffer (100  $\mu$ l final volume) for 1 h at 30°C with shaking at  $7 \times 100 \text{ min}^{-1}$ . 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<sup>-1</sup> MS salt, 10 mM MES, and 0.2 mM acetosyringone) to obtain a final OD<sub>600</sub> 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  $\mu$ 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: [C<sub>10</sub>H<sub>17</sub>]<sup>+</sup> ( $m/z$  137.1325) for geranyl glycosides and [C<sub>15</sub>H<sub>24</sub>]<sup>+</sup> ( $m/z$  205.1952) for farnesyl glycosides. The major geranyl glycosides were putatively identified as hexosyl-geraniol (C<sub>16</sub>H<sub>28</sub>O<sub>6</sub>), malonyl-hexosyl-geraniol (C<sub>19</sub>H<sub>30</sub>O<sub>9</sub>), and pentosyl-hexosyl-geraniol (C<sub>21</sub>H<sub>36</sub>O<sub>10</sub>). Similarly, the main farnesyl glycosides were putatively identified as hexosyl-farnesol (C<sub>21</sub>H<sub>36</sub>O<sub>6</sub>) and malonyl-hexosyl-farnesol (C<sub>24</sub>H<sub>38</sub>O<sub>9</sub>). 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<sup>-1</sup> 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<sup>-1</sup>. The separation was performed with the following program: 40°C for 3 min, and subsequently a gradient of 3°C min<sup>-1</sup> 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<sup>®</sup> technique in the pGWA vector (Busso *et al.*, 2005), which adds a His<sub>6</sub> 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<sup>-1</sup> chloramphenicol and 100 µg ml<sup>-1</sup> ampicillin. The bacteria were then transferred to 1 L Terrific Broth supplemented with 0.8% glycerol, 35 µg ml<sup>-1</sup> chloramphenicol, and 100 µg ml<sup>-1</sup> 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 His<sub>6</sub> tag. RhNUDX1 was further purified using a Superdex 200 gel filtration column and concentrated in Amicon<sup>®</sup> (Millipore, Guyancourt, France) to 2.5 mg ml<sup>-1</sup>. 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 (Sali 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 Mg<sup>2+</sup> ions and ligand, when absent from the structure, were added and positioned according to AtNUDX1–Mg<sup>2+</sup>–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 Sali, 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). Mg<sup>2+</sup> ions were added in the active site, as they are important for binding and catalysis. The charge of the Mg<sup>2+</sup> 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 Nicolè (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., Hugueney, 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 Ca<sup>2+</sup> homeostasis, and its juvenile hormone esters 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., Likic, 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-Martín, 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., Rocchia, 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., Mayoral, 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  $\alpha$ -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.
- Rocchia, 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*)- $\beta$ -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.

- 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.